Research on varietal classification and germination evaluation system for rice seed using hand-held devices

S. Durai and C. Mahesh

Department of Computer Science and Engineering, School of Computing, Vel Tech Rangarajan Dr Sagunthala R & D Institute of Science and Technology, Chennai, India; Department of Information Technology, Vel Tech Rangarajan Dr Sagunthala R & D Institute of Science and Technology, Chennai, India

ABSTRACT

Rice Seed varietal classification and germination evaluation system is developed to identify the variety and evaluate the germination of rice seeds using Digital Image processing system. For economic and ease of usage, we have used mobile phones to take digital images. The objective of our research is, it will be easily used by farmers. Our research is done on four major rice varieties, which commonly cultivated by Tamilnadu farmers, namely (1) Andhra Ponni (2) Atchaya Ponni (3) KOS50 and (4) IR 20 was collected from Tamilnadu Agricultural University Tiruchirappalli, Tamilnadu, India. We have extracted 24 features: 3 colour features, 13 morphological features and 8 textural features. Created data set tested with all possible classification algorithms, out of which Ensemble classification algorithm gives 91.6% accuracy for Variety Identification and SVM gives 63% of accuracy for germination prediction. According to the germination percentage, a support vector machine (SVM) was utilised to categorise the seeds specimens into 3 groups: healthy, old, and deceased. The categorisation prediction accuracy has always been significant. We have created the data set for successful identification of varieties and germination prediction for the above-mentioned varieties; it is publicly available for usage.

Introduction

Rice is one of the important food sources in India, especially in south India 90% of people have rice as the major food source. The demand of rice is increasing day by day, the farmers are cultivating rice plants, but the production result is not as they expect. Production results depend on many factors such as insufficient water, weather change, insects attacks, mixed varieties, less germinated plants, diseases and fertilisers used etc. Recent research proved that the seeds chosen for forming will decide the production. Seed is a significant living item and should be developed, handled effectively and collected to get the greatest yield, great profitability in the agrarian market. The good quality seeds always lead to better production. The following three factors decide the quality of the seed:

- Varietal purity
- Seed viability
- Moisture content

The seed’s varietal integrity can be clearly distinguished from that of its breeder germ, and it is biologically clean. The integrity of the grains as seed variation is unknown. Quality criteria, such as the elimination of off-types, diseased seedlings, weeds’ plant species, and other field crops, are followed out throughout the seed supply chain. The viability of a seed entry is a measurement about how many seedlings are alive and capable of developing into self-replicating plants under the correct circumstances. The proportion of the volume of water in a specimen to the weight of the aggregate sample, represented as a proportion, is called moisture content (Gao et al. 2020a).

Varietal purity identified by other varieties mixed, red rice seeds, germination capacity, weed seeds, moisture content and unwanted materials such as soil, dust and stones

Purity % = (Weight (or number)of factor taken)/ (Weight (or number)of samples taken)

Factor example:

Weed % = (Weed Weight)/(sample seeds weight taken)

To put it another way, 0.0044 g is 0.21 time the mass of a rice grain, and a grain of rice weighs 4.8 times the
value. Despite large variances in variety, a single slice of ordinary short-, mid-term, and lengthy rice contains approximately 0.021 g after heating on aggregate (Gao et al. 2020b).

Seed viability depends on germination capacity, moisture level and temperature of seeds kept in storage. When storing seeds for a long time, they should be dehydrated until the water content decreases below 11%, and the preservation temperature range should be kept below 20 degrees Celsius and 50% humidity. The seeds must also be placed in a waterproof container. For selecting good seed, we need to depend on unadulterated seeds from certified sellers, get from experienced formers or prepare our own great seeds. The amount of water level in the seed decides the moisture content of the rice seed. If the moisture content is less than 14%, we can say it is a good seed.

For accessing the quality of the seed, we need to depend on certified buyers and experienced formers. Formers made the decision with their previous experience and what they saw and felt about the seeds. Most of the time, this will not give the assurance of germination success and no proof in their prediction. But researchers proved the germination result depends upon the features such as cracks in the seeds, insect-affected seeds, fungal attacks, other varieties mixed, damaged seeds, chemical composition of seeds. Practically if we want to check the result of germination it will take 15 days. Ultimately, we need a simple, time-efficient, economic, non-destructive, accurate and automated method to ensure the germination evaluation of seeds.

For selecting the good germination seeds, formers are following a special and simple method called Gravity. In this method, all the seeds for testing must be put on the water that contains salt or ethanol. The floating seeds on the surface of water will not be germinated, so they will remove them from the selected seeds. Seeds that lacked viable embryo or nutrition reserves can rise because they are less compact than ‘good’ seedlings which fall in liquid. Alternatively, they could hover because they include air bubbles that don’t usually impair seed vigour or survivability. The coffee median filter seems to be the only reliable way to test sprouting. Only seeds that sink in water will be chosen for further germination. If we want to seed right away in the field, this method is ideal. However, it is not suggested for testing and later use. This process is time-consuming, and selected seeds may be thrown away due to the presence of ethanol or salt in the water used in the gravity method. If we are going with our own seed production, the gravity method is not a big problem, but most of the formers depend on third-party sellers. Gravity method uses a mixture of the seed size and thermal properties.

Here, the flotation concept is used. A seed water is pumped onto the slanted perforation seat’s low end of the scale. The seeds travel at varying rates around the deck due to a seat’s alternating motion. After testing the seeds with this method then buying is not feasible for them and moreover in India formers are following an alternate crop cultivating method, so storing of their cultivated seeds for the longer time is again a big task.

In the automated machine vision system for seed quality and germination prediction system the physical properties of seed are going to be extracted by their digital images. Once we are using high-quality capturing devices the accuracy results in better prediction results. Digital image processing systems are utilised in many research areas such as food processing, satellite image processing, medical image processing, soil testing, and many agricultural applications and already proved they gave best accurate results.

Many researchers tried on non-destructive testing methods, which are used in estimating seed quality, selecting good seeds, grading the seeds and germination test. For estimating the varietal purity, seed viability and moisture content of seed, IIRI (http://www.knowledgebank.irri.org/) developed an assessment kit Figure 1; it contains certain components, and each component is used for a particular test. The kit is too costly and for handling the kit formers need to be trained. The following components are available in the kit.

- Graduated Cylinder (Flask)
- Scale
- Calliper
- Infrared Thermometer
- IRRI Moisture Tester
- Indented Sheet Grader
- Magnifier
- Rice Milling Chart

**Vision systems**

The artificial intelligence in the machine vision system is digital image processing, which simulates human vision. The subsequent four components are common in machine vision systems: (1) Illumination system, (2) capturing device, (3) lens used, (4) computing system to process the captured images. The machine vision system is functioning with the physical properties of captured images. Physical properties are such as area, minor, major axis, colour features, shape, and size; this technique was already implemented and proved with good accuracy in some of crops in agriculture species.
We’ve listed next the various capturing devices utilised in the sector of digital image processing system.

**Spectroscopy**

Spectroscopy system investigates input images with the assistance of electromagnetic wave emitted by the spectrum. This method uses spectroscopic techniques such as mid-infrared, near-infrared, Raman spectroscopy, Fourier transform-infrared and fluorescence. Raman method is a non-chemical examination technique that can provide details about chemical composition, phases and polymorph, crystalline nature, and cellular structures. It is caused by the interaction of sunlight with an object’s organic molecules. This spectroscopic system helps to analyse and evaluate the standard of seeding to make sure the viability of seeds, chemical composition, cracked seeds, damaged by insects and other causes.

*Figure 1. IIRI rice seed quality assessment kit.*
**Hyper spectral imaging**

This hyper spectral system contains both machine vision system and spectroscopy features. The purpose of the feature extraction method is to extract meaningful information from the raw data and express it in a lower-dimensional environment. A greater proportion of multispectral images in a hyperspectral set of data may allow distinction between the more precise classes feasible, but far too many spectral bands employed in categorisation may reduce prediction performance due to the dimensionality. By using this technique, we will obtain both spectral and spatial features of the taken images. It consists of detectors, wavelength dispersion devices and lightweight sources. We are able to capture a large number of samples and may be analysed.

**Thermal imaging**

In this technique to extract features and further analysis, the hidden radiation pattern is transformed into visible images. Using this technique, the surface temperature of captured object in the image is transformed into 2D values with high resolution. Like other vision systems this system does not require illumination devices. This will be used to detect the type of insects, variety of diseases, crop water stress, viability and yield estimation. Based on weather and environmental conditions this system performed well.

**Others**

X-ray imaging system uses electromagnetic waves; we can detect internal voids, defects in the object and insect damage. Consecutive numbers of electronic and chemical sensors are used in the electronic nose system, to recognise complex and simple odours. For kernel classification and Pathogen detection, this system is used. The electronic nose is made up of a series of electrochemical devices that are linked to a pattern-recognition program that detects smells as they pass by. Varied scents elicit different approaches in the detectors, which result in a signal pattern that is unique to that fragrance (Manogaran et al. 2021).

To capture the image samples they used a charge coupled cadget (CCD) camera with 640X480 pixels, viewing field was $12 \times 9$ and 0.19 mm/pixel spatial measurement (Liu et al. 2005). CCD detectors are the most common Electron microscopy detectors. CCD detectors are made up of a two-dimensional grid of crystalline silicon detection and the circuitry is designed to understand the information from each one. Between the samples table and the focal point to reduce the impact, they use a dark light chamber. All samples taken for the testing were accepted as good seeds by experienced formers (OuYang et al. 2010). They set up the capturing hardware with a container that dropped out the seeds and distributed to the belt which moves for taking images. CCD camera is used to capture the images on the moving belt. Captured images are stored in a PC through an attached camera set-up. In the PC the stored images are further pre-processed, they use surface colour as dim green and back-ground as dark.

To capture image samples they use a CMOS image (NIKON D300S) camera with 640×480 pixels (Hong, Hai, Le Thi Lan, et al. 2015). In the flat bed set-up they use white as background colour, rice seeds are distributed in the surface area of $10 \times 16$ cm. Each image contains 30–60 seeds; they segment the individual seeds and process further (Chaugule and Mali 2014). They use a Sony 18.9 Megapixels digital camera for capturing the samples and used dark material sheet as background. They use the seeds collected from the Pune Seed Testing Laboratory, India. The seed varieties were K6, R2, R4, and R24 (Khunkhett and Remsungnen 2014). They use MP287 canon manufactured 1200X2400 dpi scanner with Compaq V3000 Dual core processor notebook.

Due to the germination time of the seeds, determining the viability of the seeds is a time-consuming and challenging operation. Seed producers conduct seed germination tests; on the consumer or buyer side, we must blindly accept the producers’ certification. Even if customers want to check, there are no non-destructive traditional methods are available and will take more than ten days. The seed germination evaluation will be calculated by

$$GR\% = \frac{NGS}{NS} \times 100$$

GR is the germination rate; NGS is the number of germinated seeds; NS is the number of seeds used for testing. Germination evaluation is still a complex and time-consuming task because a huge number of seeds have to be analysed; germination verification is also complex because germinated seeds may be overlapped and the separation will be done manually. Transpiration, or when a seedling draws in moisture from the earth, is the first step in seed germination. This causes the plant’s roots to expand, allowing it to absorb additional water. Seed germination is critical for spontaneous plant growth and agricultural production for human consumption. It is critical to understand that seedlings deposited in a GenBank database will grow into plants. The survivability of seedlings at the commencement of
preservation will also define the accession’s life span, based on environmental circumstances. In this chapter, we are going to summarise automated non-destructive methods, various classification methods, features extracted in each method and result accuracy. Colour and computerised object detection, Visual NIR spectroscopic, NMRI, wireless sensors, ultrasonic, x-ray scanning, and biosensing are non-destructive performance assessment methods have shown significant promise for meat. Although several non-destructive testing procedures can uncover rejection faults in welding, sensor fusion tensile inspection is the most economic and most effective.

Figure 2 shows the parts of the rice seed (Fang and Yibin 2004). To examine incompletely closed glumes in the rice seeds, they developed a machine vision inspection system. They proved incompletely closed glumes seeds are having less moisture content compared to normal seeds. Incomplete closed glumes seeds are deformed; sometimes, normal and good seeds always have well-closed glumes. They classify: (1) Good Seed, (2) fine fissure seed – incompletely closed glume seeds, these seeds must be treated separately along with good seeds and (3) unclosed seeds – the germination percentage of these seeds is low.

They used Jiayou, Zhongyou207, Jinyou402, Shanyou10 and Ilyou seed varieties to test their proposed germination estimation rate. Hough transform with Digital image processing techniques was applied and the accuracy is Good seeds – 96%, Fine fissure seeds – 92% and unclosed glumes seeds – 87%.

Nguyen et al. (2018) developed an automatic identification tool to estimate the germination rate of rice seeds in real-time application. Deep learning and digital image processing techniques are used to assess the germination rate of seeds. They have implemented this for separating germinated seeds from the germination paper. Images are taken from the germinated paper and image processing techniques are applied. The prepared data set is publicly available for research; they used different varieties. Usually to evaluate the germination rate the seeds will be separated from the germination sheet and placed on a dark ceramic tray. This method is a time-consuming and complex task because seeds are overlapped with each other. For seed segmentation and seed localisation, they applied fully convolution networks – U-Net was used. For seed classification Res-Net-101 [deep residual networks with 101 layers] were used. CNN converts an image into a vector, which is commonly utilised in classification issues. In U-Net, however, an image is turned to a vector, which is then translated back to a picture using the same mappings. This reduces distortions by keeping the picture’s original building. ResNet employs a skip connection, which means that an originating input is also appended to the convolutional frame’s result.

To evaluate the standard of rice seeds they use a digital image processing system for capturing, processing seeds and multilayer feed-forward neural network techniques (Shah et al. 2013). Features extracted are major axis Length, minor axis Length, eccentricity and the area of each seed. To test it they used kamod rice, usually it is lengthy. To produce good quality rice, the lengthy seeds were selected. For the selection of good seeds the tiny seeds must be separated from the samples. The number of small seeds and other varieties mixed in the samples decides the standard of seeds. They obtain 100 percent accuracy in this variety.

Hong, Hai, Hoang, et al. (2015) implemented a method to identify the Breeder seed. They considered four standards of seeds commonly used in agricultural markets: 1. Breeder Seed 2. Foundation seed 3. Registered seed and 4. Commercial seeds. All seeds are graded by considering the percentage of excellent seed, percentage of contamination, percentage of germination, percentage of mixed species, percentage of red rice seeds, and percentage moisture content. They used Khao Dawk Mali 105 seeds to achieve their recommended strategy. Area, Diameter, Average histogram R, G, and B were recovered from Khao Dawk Mali 105, and they employed a Degree of linking superficial pattern with 98 per cent accuracy.

To evaluate germination rate they used three colour features, seven morphological and eight textural features of rice seeds (http://www.knowledgebank.irri.org/). Figure 3 shows the proposed method of their work. There are three colour features (Average Red, Green, and Blue), as well as seven morphological aspects (Major axis, Minor axis, Orientation, Eccentricity, Area, Roundness, Aspect ratio, eight textural features, Contrast horizontal, vertical, Correlation horizontal, vertical, Entropy horizontal, vertical, Homogeneity horizontal,
vertical). They used cp-111 rice seed and applied artificial neural network-18-1-2 for evaluation. Their results are 7.66% false accepted, and 5.42% false rejected.

**Data set preparation**

In image processing, data set preparation is the important and core step. Figure 3 shows the proposed method structure of our research work. Both the input and output are intensity images; hence pre-processing is a frequent moniker for actions with pictures at the simplest level of complexity. Pre-processing is used to enhance visual information by suppressing undesired aberrations or enhancing certain visual qualities that are relevant for further processing. The accuracy of the result always depends on the data sets used to train the machine. The data set (Hong, Hai, Hoang, et al. 2015) preparation contains the following important steps.

1. Hardware requirements
2. Samples
3. Hardware set-up
4. Naming the images
5. Pre-processing
6. Features’ Extraction
7. Decision on the conclusion attribute

The above steps are common for machine vision systems. Some hardware is common and must for digital image processing systems [Flat bed, camera, storage, lighting, camera holder etc.]; depending upon the cost and accuracy the hardware configuration may be varied. Samples are the data we are going to capture and process as an image. Samples must be carefully selected, choosing the samples following many factors, will be discussed further in the next sections. Samples must be collected from any certified research centres or certified buyers. They must be mixed carefully for classification. Each sample must be named and noted for future reference. Figure 4 shows the structure of data set preparation involved in our research work.

**Materials used and hardware set-up**

Capturing the image we have used VIVO Z1 Pro mobile with 16Mega Pixel camera, for storage and processing image HP Laptop with corei5 8th Gen processor, 4GB NVIDIA GeForce GTX, 1 TB HDD system used. The main objective of our research is to reduce the cost of data set preparation and to be easy to use. The mobile phone we have used gives decent images with good details. Image processing technique provides unravelled performance and great deal of versatility. The adoption of bespoke processors that perform similar tasks can drastically improve the performance. Finally, because the circuits are tailored for the purpose, energy consumption can be decreased (Deepa et al. 2020).

For the Flatbed scanner, we have used black background, black colour sheets are available in shops at less cost. A flatbed scanner is a type of imaging device that scans materials on a smooth surface. The scanners can capture all information on a sheet and do not require the material to be moved. In the black sheet, we have drawn a square with 10 cm side and in the centre of the square mark a dot for place our sample. And in the top of the flat bed we have pasted one white colour small sheet for naming the sample. Fixing the mobile we have used wooden materials, and the distance between flatbed and camera is 30 cm, which gives a good detailed image. We have tested the same with different distances such as 40, 20 and 50 cm. But 30 cm distance set-up gives the results as expected.

In capturing the image, a lot of options now come with mobile such as photo, AI BEAUTY, NIGHT, PANO, PRO and DOC mode. We cannot use other than PRO mode.
Figure 4. Data set preparation.
mode, because mobile phones have built-in software to adjust the image quality, such as auto contrast enhancement, brightness, colour change etc. But image processing always depends on those features only. For the same sample if you are taking multiple photos with other modes each and every image is giving different feature values. But for our system, it must be constant, for that we can use only the PRO mode, where we can fix the capturing properties; once fixed the same values must be maintained for all sample capturing. In our preparation, we have set EV:−1.7 and ISO:400. In the room lighting, we have used a Philips Stellar Bright 20-Watt Round LED Bulb and the room temperature is constantly maintained with 88 degree.

**Samples ‘preparation**

For preparing the datasets we have used four major rice seeds cultivated from Tamilnadu Agricultural university Trichirappalli, Tamilnadu, India, namely, Andhra Ponni, Atchaya Ponni, KO50 and IR20. After a few years of drought, the monsoon season has provided a plenty of rain. This will help to enhance agriculture and GDP at a time when global growth is slowing. Poor rainfall and a dry period in the regions are predicted to impact the crop, which is anticipated to be 28% lower than typical. We have collected 100 g in each variety; all are certified seeds and 100% germinating viability. But for our research we need some non-germinating seeds also, so for each variety, we have collected the non-germinating seeds also from the university. The seeds are not germinated for the following reasons: red rice seeds, physical damage, fungal attack and discoloration. The growth step of a seed is known as germination. Breaking the seed coat takes a lot of energy, and as the seed grows, the amount of energy required grows as well. Non-germinating seeds, on the other hand, are quiescent and do not use much oxygen. In sequence for the seeds to survive, it must breathe (Balamurugan et al. 2020).

For each variety we have taken 20 samples, 10 are good and 10 are non-germinating seeds. To name the seed the following short names are used for the entire dataset preparation shown in Table 1.

**Table 1. Variety name and identifier.**

| Variety name | Identifier name (Good sample) | Identifier name (Bad sample) |
|--------------|-------------------------------|------------------------------|
| Andhra Ponni | GAP                           | BAP                          |
| Atchaya Ponni| GATP                          | BATP                         |
| KO50         | GK50                          | BKO50                        |
| IR20         | GIR20                         | BIR20                        |

**Figure 5.** Parts of a rice seed.

are placed one by one and the names will be written on white sheet such as GATP1, GATP2 for all sample examples shown in Figure 6 captured samples of BAP number 4 and Figure 7 captured the sample of GKO50 number 15.

**Figure 6.** Capture image of sample bad Andhra Ponni.
Germination test

Once the sample is captured, it must be kept inside a white colour cotton cloth for the germination test. In the white cloth write the name with a permanent marker. In the germination test, the samples will be submerged in water for 12–24 h (Fang and Yi-bin 2004). A germination test measures what proportion of seedlings in a single seed are alive. While sprouting times vary significantly between kinds, seedlings should absorb water in a couple of days and form a stem and the first plant in four. The seed is said to have grown at this stage. As per the suggestions from the experienced farmer, we have submerged the sample with cotton for 23 h (from 6 pm to the next day 5.00 pm). After 23 h we opened each sample cotton package, a small shoot appeared on most of the samples.

For planting the samples we have used paper cups for each sample named with the same terms used to identify the samples. All the paper cups are filled with good soil for planting the seeds suggested by farmers. The same quality and quantity of soil is filled in all cups and the cups are now placed in sunlight. Every day morning the cups are filled with water. After 10 days the seeds are successfully germinated and the result of each sample is recorded, for the machine learning the germination status is the resultant attribute. Figure 8 shows the step-by-step method of germination verification samples.

Extracting features

Extracting the features you can use any image processing tool, we have used matlab for extracting the features. Before processing the image, the taken one must be cropped. The crop will be done on the square we have drawn with same dimension for all images and save the file name as the identified term, but we are not supposed to change the colour properties. For our research work we have extracted totally 24 features shown in Table 2. The prepared data set is freely available in https://github.com/duralrichy/Riceseed.

Colour features

The colour features are (1) average red colour (R), (2) average green colour (G) and (3) average blue colour (B). The mean of each colour in the entire rice seed image is used in all of the average RGB characteristics.

Morphological features

The thirteen morphological features describe image regions based on an ellipse and consist of

(a) Area: By summing the areas of each pixel in an image, the area of all of the pixels in the image is estimated. The area of a single pixel is calculated by looking at its two-by-two neighbours.
(b) Major Axis: Returns as a scalar the length (in pixels) of the major axis of the ellipse with the same normalised second central moments as the region.
(c) Minor Axis: Returns as a scalar the length (in pixels) of the minor axis of the ellipse with the same normalised second central moments as the region.
(d) Orientation: As a scalar, the angle between the x-axis and the major axis of the ellipse with the same second-moments as the region.
(e) Eccentricity: Returns the eccentricity of an ellipse with the same second-moments as the region as a scalar. The eccentricity of an ellipse is the ratio of the distance between its foci and the length of its major axis.
(f) Euler Number: A scalar representing the number of items in the region minus the number of holes in those objects.
(g) Equiv Diameter: Formula: sqrt(4*Area/pi).
(h) Perimeter: The distance around the region’s boundary was returned as a scalar. The perimeter of a region is calculated by measuring the distance between each contiguous pair of pixels along the region’s border.
(i) Convex Area: Number of pixels in ‘ConvexImage’, returned as a scalar.
(j) Extent: A scalar representing the ratio of pixels in the region to pixels in the whole bounding box. Calculated by dividing the area of the bounding box by the area of the region.
(k) Solidity: Proportion of the pixels in the convex hull that are also in the region, returned as a scalar. It is computed as Area/ConvexArea.
(l) Circularity: Object circularity is returned as a struct with the field Circularity. The circularity value for

Figure 7. Captured image of sample good KOSO.
The value of circularity is calculated as follows:

$$\text{Circularity} = \frac{4 \times \text{Area} \times \pi}{\text{Perimeter}^2}$$

**Figure 8.** Germination evaluation steps.

Each object in the input image is stored in the struct.

(m) Aspect Ratio AR: $\text{AR} = \frac{\text{MJAX}}{\text{MNAX}}$

MJAX – Major axis length
MNAX – Minor axis length
The grey-level co-occurrence matrix (GLCM) is a statistical approach of assessing texture that takes into account the spatial interaction of pixels. The GLCM during this research applies four texture features: (1) contrast, (2) correlation, (3) entropy and (4) homogeneity. The system employs eight textural features using GLCM in two directions: horizontally and vertically. The eight texture features are as follows: (1) contrast horizontal, (2) correlation horizontal, (3) entropy horizontal, (4) homogeneity horizontal, (5) contrast vertical, (6) correlation vertical, (7) entropy vertical and (8) homogeneity vertical. The textural values of the GLCM are defined as follows: \( P_{i,j} \) is the entry in a normalised grey-tone spatial-dependence matrix, and \( N \) is the number of unique grey levels inside the quantised image. The Level Co-occurrence Matrix (GLCM) approach is a technique for obtaining probabilistic shape information of the second sample. The third and higher-level textures examine the connections between three or even more pixel, and have been employed in a wide range of applications (Table 3–5).

a) The contrast texture feature assesses the GLCM’s local variances. The texture of contrast can be estimated using:

\[
\sum_{i,j=0}^{N-1} P_{i,j} (i - j)
\]

b) The correlation texture feature determines how linearly the grey levels of neighbouring pixels are related. The texture of correlation can be calculated using:

\[
\sum_{i,j=0}^{N-1} P_{i,j} (i - \mu_i) (j - \mu_j) / \lambda_i \lambda_j
\]

where \( \mu_i, \mu_j, \lambda_i \) and \( \lambda_j \) are the mean and standard deviations of \( P_{i,j} \).

c) The entropy texture feature characterises the texture of the input image using a statistical measure of unpredictability. The texture’s entropy can be estimated by using:

\[
\sum_{i,j=0}^{N-1} P_{i,j} (i - \ln P_{i,j})
\]

d) The homogeneity texture feature assesses how close the elements in the GLCM are distributed to the GLCM diagonal. The texture of homogeneity

### Table 2. Extracted features.

| Colour features | Morphological features | Textural features |
|-----------------|------------------------|------------------|
| Average Red colour | Major axis length | Contrast (Horizontal & Vertical) |
| Average Green colour | Minor axis length | Correlation (Horizontal & Vertical) |
| Average Blue colour | Orientation | Entropy (Horizontal & Vertical) |
| | | Homogeneity (Horizontal & Vertical) |
| | | Area |
| | | Euler number |
| | | Equivalent Diameter |
| | | Perimeter |
| | | Convex area |
| | | Extent |
| | | Solidity |
| | | Circularity |
| | | Aspect Ratio |

### Table 3. Sample extracted colour feature values.

| Variety name | R colour | G colour | B colour |
|--------------|----------|----------|----------|
| AP           | 64.7758  | 64.621   | 64.3939  |
| AP           | 65.7698  | 65.5277  | 65.1802  |
| AP           | 66.0183  | 65.9375  | 65.7044  |
| ATP          | 66.4123  | 66.2308  | 65.961   |
| ATP          | 66.9279  | 65.8902  | 65.8999  |
| ATP          | 66.0688  | 65.4546  | 65.4118  |
| IR20         | 71.9369  | 70.4771  | 70.6304  |
| IR20         | 70.9271  | 69.5236  | 69.5177  |
| IR20         | 71.3181  | 70.3642  | 70.4938  |
| K50          | 72.2553  | 71.4895  | 71.6597  |
| K50          | 71.373   | 70.864   | 70.6765  |
| K50          | 71.2457  | 70.8528  | 70.4273  |

### Table 4. Sample extracted morphological feature values.

| Variety name | Area | Major Axis | Minor Axis | Eccentricity | Orientation | Euler Number | Equivalent Diameter | Perimeter | Convex area | Extent | Solidity |
|--------------|------|------------|------------|--------------|-------------|--------------|---------------------|-----------|-------------|--------|----------|
| AP           | 4865 | 135.71     | 47.5574    | 0.9366       | –4.1637     | 7004         | –139                | 83.2404   | 0.7723      | 0.7038 | 684.966  |
| AP           | 5306 | 125.633    | 56.8811    | 0.8916       | –10.263     | 6642         | –128                | 82.1937   | 0.7989      | 0.6477 | 680.347  |
| AP           | 4120 | 127.935    | 43.9062    | 0.9393       | –4.1637     | 5364         | –93                 | 72.4275   | 0.7681      | 0.6311 | 743.822  |
| ATP          | 5287 | 156.893    | 46.5944    | 0.9549       | 0.7484      | 7167         | –105                | 82.0464   | 0.7377      | 0.5901 | 796.989  |
| ATP          | 3712 | 118.754    | 41.4245    | 0.9372       | –5.3048     | 4744         | –72                 | 68.7478   | 0.7256      | 0.5044 | 658.458  |
| ATP          | 4442 | 144.203    | 40.6323    | 0.9595       | 11.0023     | 5396         | –98                 | 75.2046   | 0.8232      | 0.6427 | 662.685  |
| IR20         | 6260 | 152.145    | 54.4644    | 0.9337       | –1.8532     | 7666         | –128                | 89.2725   | 0.8166      | 0.6987 | 843.67   |
| IR20         | 6250 | 147.017    | 55.7076    | 0.9254       | 11.4082     | 7613         | –137                | 89.2602   | 0.821       | 0.6425 | 846.797  |
| IR20         | 6249 | 148.43     | 55.1287    | 0.9285       | 9.052       | 7690         | –125                | 89.1991   | 0.8126      | 0.6103 | 758.803  |
| K50          | 8272 | 167.682    | 66.0304    | 0.9192       | 82.1282     | 1004         | –181                | 102.627   | 0.8234      | 0.6839 | 898.032  |
| K50          | 5888 | 151.801    | 54.798     | 0.9326       | 7.9129      | 8357         | 151.162             | 86.5843   | 0.7046      | 0.46   | 801.544  |
| K50          | 5578 | 143.819    | 50.616     | 0.9366       | 1.599       | 6859         | –90                 | 94.2741   | 0.8132      | 0.6553 | 696.028  |
Input:
Training data \( \{ x_i, y_i \} \) where \( i = 1 \) to \( n \), new data \( x \), subspace distribution \( D \), criterion \( C \), integers \( B_l \) and \( B_2 \), type of base classifier \( T \)

Output:
predicted label \( C(x) \), the selected proportion of each feature \( F_j \)

1. Independently generate random subspaces \( S_j \sim D \), \( 1 \leq j \leq B_1 \), \( 1 \leq k \leq B_2 \)
2. for \( j = 1 \) to \( B_1 \) do
   a. select the optimal subspace \( S_f \) from \( \{ S_j \} \) where \( k = 1 \) to \( B_2 \) with respect to \( C \) and \( T \)
3. end
4. Construct the ensemble decision function \( V(x) = \frac{1}{B_1} \sum_{j=1}^{B_1} C(x|S_j) \)
5. Set the threshold \( a \) according to (2)
6. Output the predicted label \( C(x) = 1(V(x) > a) \), the selected proportion of each feature \( F = \{ f_1, ... , f_p \} \) where
   \[ F = B_f^{-1} \sum_{j=1}^{B_1} \frac{1}{B_1} 1(V_j > a) \]

Varietal classification

In the previous section, we have explained about the data set’s preparation. In this section, we are going to describe automatic varietal classification using deep learning. Random Subspace Ensemble (RaSE) is a classifier based on deep learning used for training the samples. They combine multiple weak learners in the RaSE approach, in which each loss function is a base classification learned in an ideally chosen region from a pool of random subsets. For the experimentation we have used four varieties, which are commonly cultivated by Tamilnadu farmers, namely, 1) Andhra Ponni 2) Atchaya Ponni 3) KO50 and 4) IR 20 were collected from Tamilnadu agricultural university Tiruchirappalli, Tamilnadu, India. We have applied different possible classification learner’s algorithms which are available with matlab and the accuracy percentage is displayed in Table 6.

Table 5. Sample extracted textural features values.

| Variety name | Contrast H | contrast V | Correlation H | Correlation V | Entropy H | Entropy V | Homogeneity H | Homogeneity V |
|--------------|------------|------------|---------------|--------------|-----------|-----------|--------------|--------------|
| AP           | 0.153526   | 0.144061   | 0.720753      | 0.737899     | 0.372137  | 0.376675  | 0.925274     | 0.928687     |
| AP           | 0.163351   | 0.152702   | 0.692437      | 0.712454     | 0.37462   | 0.380166  | 0.920029     | 0.924302     |
| AP           | 0.161794   | 0.154367   | 0.697502      | 0.711316     | 0.381491  | 0.383621  | 0.921711     | 0.923521     |
| ATP          | 0.019818   | 0.015484   | 0.753491      | 0.805848     | 0.959725  | 0.960185  | 0.992796     | 0.993054     |
| ATP          | 0.032187   | 0.027531   | 0.690287      | 0.733364     | 0.928885  | 0.929568  | 0.986676     | 0.987098     |
| IR20         | 0.026555   | 0.021814   | 0.712279      | 0.761954     | 0.941843  | 0.94225   | 0.989526     | 0.989888     |
| IR20         | 0.032635   | 0.021814   | 0.712279      | 0.761954     | 0.941843  | 0.94225   | 0.989526     | 0.989888     |
| IR20         | 0.021528   | 0.018454   | 0.75261       | 0.787607     | 0.950099  | 0.949761  | 0.991547     | 0.991592     |
| IR20         | 0.073169   | 0.068975   | 0.660063      | 0.679436     | 0.801967  | 0.801846  | 0.966126     | 0.966386     |
| IR20         | 0.023408   | 0.020791   | 0.654059      | 0.761954     | 0.941843  | 0.94225   | 0.989526     | 0.989888     |
| AP           | 0.015222   | 0.010499   | 0.753491      | 0.805848     | 0.959725  | 0.960185  | 0.992796     | 0.993054     |
| AP           | 0.012353   | 0.015484   | 0.690287      | 0.733364     | 0.928885  | 0.929568  | 0.986676     | 0.987098     |
| AP           | 0.019818   | 0.015484   | 0.690287      | 0.733364     | 0.928885  | 0.929568  | 0.986676     | 0.987098     |
| AP           | 0.073169   | 0.068975   | 0.660063      | 0.679436     | 0.801967  | 0.801846  | 0.966126     | 0.966386     |
| AP           | 0.021528   | 0.018454   | 0.75261       | 0.787607     | 0.950099  | 0.949761  | 0.991547     | 0.991592     |

Table 5 shows the list of classification learner algorithms applied for the data set prepared and accuracy percentage calculated. Out of these algorithms, Ensemble (Subspace Discriminant) classification learning algorithm gives high accuracy with 91.6%. The Random Subspace Ensemble Classification algorithm is exported for testing the samples. Figure 9–11 shows the prediction model, confusion matrix and True-Positive Rates (TPR) and False-Negative Rates (FNR) of Discriminant prediction. The True-Positive Rate (TPR) is the percentage of true forecasts in positive class forecasts. The proportion of statistical significance fails to reject a null hypothesis, when the null hypothesis is accepted it is referred to as the false negative rate.

The model is trained and it is tested with test data. In the testing it produced 83% accuracy. I testing we used only less of data. In the test data, we have given Andhra Ponni (5 numbers), Atchaya Ponni (4 numbers), IR20 (4 numbers) and KO50 (4 numbers). The trained model result is for Andhra Ponni, 1 data is misclassified as

Table 6. Classification algorithms applied and accuracy detected.

| S.No | Classification algorithm     | Accuracy percentage |
|------|------------------------------|---------------------|
| 1    | Tree                         | 61.4                |
| 2    | Linear Discriminant          | 86.7                |
| 3    | Naive Bayes (Gaussian)        | 69.9                |
| 4    | Naive Bayes (Kernel)          | 74.7                |
| 5    | SVM (Linear)                 | 74.7                |
| 6    | SVM (Quadratic)              | 77.1                |
| 7    | SVM (Cubic)                  | 78.3                |
| 8    | SVM (Medium Gaussian)        | 77.1                |
| 9    | SVM (Coarse Gaussian)        | 62.7                |
| 10   | KNN (Fine KNN)               | 79.8                |
| 11   | KNN (Medium KNN)             | 71.1                |
| 12   | KNN (Cosine KNN)             | 75.9                |
| 13   | KNN (Cubic KNN)              | 75.9                |
| 14   | KNN (Weighted KNN)           | 73.5                |
| 15   | Ensemble (Bagged Trees)      | 72.3                |
| 16   | Ensemble (Subspace Discriminant) | 91.6               |
| 17   | Ensemble (RUS Boosted Trees)  | 69.9                |
**Figure 9.** Ensemble subspace discriminant prediction model.

**Figure 10.** Confusion matrix of discriminant prediction model for a number of observations.
IR20, 1 data is misclassified as KO50 and KO50, 1 data is misclassified as IR20. So in the testing 14 data varieties successfully identified out of 17.

**Germination prediction**

For the germination prediction, with extracted features the germination results are also added as one feature in the data set as shown in Table 7. To train the model, we have applied the possible classification learning models available with matlab. The accuracy percentage of each algorithm is displayed in Table 7.

Table 7 shows the accuracy percentage of each algorithm for germination prediction for the four sample varieties we have taken. Out of these algorithms, Logistic Regression classification learning algorithm gives high accuracy with 67.5%. To describe data and understand the relationship between the dependent variable and one or even more conventional, continuous, intermediate, or ratio-level outcome variables, logistic regression is utilised. The supervised training classifiers logistic regression model predicts the likelihood of a response variable. There are only two classes since the target or predictor variables have a dualistic structure. It's one of the most fundamental machine learning algorithms, and it can help with spam filtering, diabetes prognosis, cancer detection, and other classification jobs. So, the model is exported for testing the samples data set.

**Figure 11.** True-positive rates (TPR) and false-negative rates (FNR) of discriminant prediction.

**Figure 12–14** shows the prediction model, confusion matrix and True-Positive Rates (TPR) and False-Negative Rates (FNR) of Logistic Regression for the given training data.

The model is trained and it is tested with test data. In the testing it produced 76% accuracy. Testing we used only less data. In the test data, we have given the same data set what we have used for testing the varieties, except variety name, whose feature is not needed for germination prediction. The trained model result is 4, of them misclassified and 13 classified.

**Table 7.** Classification algorithms and accuracy for germination prediction.

| S. No | Classification algorithm | Accuracy percentage |
|-------|--------------------------|---------------------|
| 1     | Tree                     | 57.8                |
| 2     | Linear Discriminant      | 61.4                |
| 3     | Quadratic Discriminant   | 61.4                |
| 4     | Logistic Regression      | 67.5                |
| 5     | Naive Bayes (Gaussian)   | 61.4                |
| 6     | Naive Bayes (Kernel)     | 59.0                |
| 7     | SVM (Linear)             | 57.8                |
| 8     | SVM (Quadratic)          | 61.4                |
| 9     | SVM (Cubic)              | 59.0                |
| 10    | SVM (Fine Gaussian)      | 53.0                |
| 11    | SVM (Medium)             | 59.0                |
| 12    | SVM (Coarse)             | 55.4                |
| 13    | KNN (Fine KNN)           | 61.4                |
| 14    | KNN (Cosine)             | 60.2                |
| 15    | Ensemble (Bagged Trees)  | 61.4                |
| 16    | Ensemble (Subspace Discriminant) | 61.4            |
| 17    | Ensemble (Subspace KNN)  | 55.4                |
correctly. Three non-germinated seeds are classified as germinated and 1 germinated seed is classified as non-germinated.

**Conclusion**

In this article, we have proposed an economic and automated method to classify the variety and evaluate the
prediction rate of germination capacity for the four major rice varieties cultivated in Tamilnadu, India. The four varieties, namely, Andhra Ponni, Atchaya Ponni, IR20 and KOS which are collected from Tamilnadu agricultural university, Tiruchirappalli, Tamilnadu, India. In this machine vision implementation system we have used mobile phones for taking the image samples, to reduce the cost for data set preparation and we have extracted a total of 24 features from the images taken. For the varietal classification, we have used Ensemble subspace Discriminant prediction model which produced 91.6% in training and 83% during testing. For germination prediction, we have used the Logistic Regression prediction model which gives 65.7% during training and 76% accuracy during testing. By using our research, formers can use it for selecting seeds for the cultivation. And our model is easy to use and convenient. In the near future, we plan to implement our model as a mobile application for the ease of use by the formers.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Notes on contributors**

*S. Durai* is a Research Scholar and Assistant Professor, received his BE degree in Computer Science and Engineering from Anna University Chennai, India in 2009. He completed Master of Engineering in Software Engineering from Anna University of Technology, Trichy, India in 2011. He is a Research Scholar and Assistant Professor in the Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India. He is having good knowledge in data structures, algorithm analysis, operating systems and Programming Languages. His research area is Image processing in agriculture.

*Dr C. Mahesh* working as a Associate professor and Head of the department in Information technology, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India. He is having more than 12 years in teaching experience. His research area are data mining, Image Processing, Bio-informatics, Machine Learning and Guiding many research scholars. He published more than 20 articles in various Scopus indexed and SCI journals.

**ORCID**

S. Durai http://orcid.org/0000-0002-6140-0791

**References**

Balamurugan S, Muthu BA, Peng SL, Wahab MHA. 2020. Call for special issue papers: Big Data analytics for agricultural disaster management. Big Data. 8(6):544–545. doi:10.1089/big.2020.29038.cfp2.

Chaugule A, Mali SN. 2014. Evaluation of texture and shape features for classification of four paddy varieties. J Eng. 2014. doi:10.1155/2014/617263.

Deepa N, Khan MZ, Prabadevi B, Durai Raj Vincent PM, Maddikunta PK, Gadekallu TR. 2020. Multiclass model for agriculture development using multivariate statistical method. IEEE Access. 8:183749–183758. doi:10.1109/access.2020.3028595.

Fang C, Yi-bin Y. 2004. Machine vision inspection of rice seed based on Hough transform. J Zhejiang Univ Sci A. 5(6):663–667. doi:10.1007/BF02840977.

Gao J, Wang H, Shen H. 2020a. Task failure prediction in cloud data centers using deep learning. IEEE Trans Serv Comput. 1–1. doi:10.1109/tsc.2020.2993728.

Gao J, Wang H, Shen H. 2020b. Machine learning based workload prediction in cloud computing. 2020 29th International
conference on computer communications and networks (ICCCN). doi:10.1109/icccn49398.2020.9209730.

Hong PTT, Hai TTT, Hoang VT, Hai V, Nguyen TT. 2015. Comparative study on vision based rice seed varieties identification. In 2015 seventh International Conference on Knowledge and Systems Engineering (KSE), October. IEEE; p. 377–382. doi:10.1109/KSE.2015.46.

Hong PTT, Hai TTT, Le Thi Lan VTH, Thuy NT. 2015. Identification of seeds of different rice varieties using image processing and computer vision techniques. J Sci Devel. 13(6):1036–1042.

Khunkhett S, Remsungnen T. 2014. Non-destructive identification of pure breeding rice seed using digital image analysis. The 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE), March. IEEE; pp. 1–4. doi:10.1109/JICTEE.2014.6804096.

Liu ZY, Cheng F, Ying YB, Rao XQ. 2005. Identification of rice seed varieties using neural network. J Zhejiang Univ Sci B. 6(11):1095. doi:10.1631%2Fjzus.2005.B1095.

Manogaran G, Alazab M, Muhammad K, Albuquerque VH. 2021. Smart sensing based functional control for reducing uncertainties in agricultural farm data analysis. IEEE Sensors J. 1. doi:10.1109/jsen.2021.3054561.

Nguyen TT, Hoang VN, Le TL, Tran TH, Vu H. 2018. A vision based method for automatic evaluation of germination rate of rice seeds. In 2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR), April. IEEE; p. 1–6. doi:10.1109/MAPR.2018.8337511.

OuYang AG, Gao RJ, Sun XD, Pan YY, Dong XL. 2010. An automatic method for identifying different variety of rice seeds using machine vision technology. 2010 Sixth International Conference on Natural Computation, August. IEEE; Vol. 1, p. 84–88. doi:10.1109/ICNC.2010.5583370.

Shah V, Jain K, Maheshwari CV. 2013. Non-destructive quality analysis of Kamod Oryza sativa ssp Indica (Indian rice) using machine learning technique. 2013 International Conference on Communication Systems and Network Technologies, April. IEEE; pp. 95–99. doi:10.1109/CSNT.2013.29.