JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Rice seeds classification using RF classifier

S. VARSHA REDDY(M.Tech) & DR. SREEDHAR BHUKYA(Professor)

Department of computer science and engineering

Sreenidhi institute of science and technology, Hyderabad.

Email Id: varshareddy102@gmail.com

Abstract: Nowadays, rice has become very important food in our country. Rice is a staple and cheap food in our country. 3/4th of the population depends on the rice. Most of the people are suffering from starvation due to insufficient supply of rice. This is primarily due to rice damage, which is primarily caused by the introduction of weeds and off-types into the crop. This effects the price and quality of the product. Inspection of rice is a very difficult task for the farmers as it ensures quality of the product. Recent approaches have been used some of the techniques in order to ensure its quality. Some of approaches utilize appearance-based features extracted from RGB and some approaches utilize spectral features from hyper spectral imaging system. It takes labor- intensive and time-consuming. In this project, a method is done to automatically screen and classify the rice seeds by the combination of spatial features and spectral features extracted from RGB and HSI system.

Keywords: Authenticate, credential security, public cloud availability, key is generated are some of the terms used in this paper.

1. Introduction

The main task for the farmers is to ensure that all seeds in a batch belong to one variety. Due to the rice damage the rice is affecting in large quantities. The significant impact on rice exporting countries like Thailand, Vietnam, and China. Damage to rice mainly affects the price and quality of the product. The certification process is done by Rice seed authenticating centres. In order to ensure its quality, a sample of rice seeds are taken from a batch to perform manual screening to accept or reject grains based on the appearance-based features such as shape, length, width and color. This is a laborious and timeconsuming process. Machine learning is the process of creating computer systems that can learn and adapt without being told what to do, utilizing algorithms and statistical models. Many algorithms have been developed throughout the years to extract knowledge from massive sets of data. There are various challenging approaches, including 3 as classification, association, and clustering, to name a few. Here we will focus on classification methodology. This set is usually referred to as a training set, because, in general, it is used to train classification technique to perform its classification.

1.1 OBJECTIVE:

- The system that combines spatial features and spectral features to classify rice seeds.
- The combined features give discrimination ability classification performance.
- “The classification of rice seeds” predicts the rice seeds using RF classifier algorithm.
- This project is implemented by using PYTHON technology.

1.2 PROBLEM STATEMENT

Recent approaches have made manual screening to classify the rice seeds. Due to which it causes time consuming and laborious.

2. LITERATURE SURVEY

Analysis and identification of Rice adulteration using Terahertz Spectroscopy and Pattern Recognition Algorithms.

Author: Chao li, bin li , dapping ye.

• Description: • Rice adulteration is a major issue for both suppliers and consumers because there are hundreds of different rice species and rice is a staple food for half of the world's population.

• Using terahertz spectroscopy and pattern recognition algorithms, we will analyses and identify rice adulteration to determine whether high quality rice is combining with low quality rice at five levels of mixing different quantities.

• Terahertz transmission mode, with spectral data reduced using principal component analysis.

• Following that, the absorption spectra were combined with partial least squares discriminate analysis (PLS-DA), support vector machine (SVM), and a back propagation neural network (BPNN) after various pretreatments.

• The results show that an SVM model using absorption spectra with a 1st derivative has a prediction set accuracy of up to 97.33 percent.

• This finding demonstrates that terahertz spectroscopy in combination with chemo metric approaches can be a useful tool for detecting levels of rice adulteration.

Rice blast detection using internet of things and artificial intelligence technologies.

Author: Wen-Liang Chen, Yi-Bing Lin, Finagling Ng, Chun-You Liu, Yun-Wei Lin.

• Description: • Rice is considered as a staple food for millions of people in the world today. Rice crop failures potentially cause starvation.

• A major reason for rice crop failure is the attack of rice blast that is one of the most serious plant diseases.

• The fungus causing rice blast is Magnaportheoryzae (or Mangalore grisea), which results in lesions on leaves, stems, peduncles, panicles, seeds, and roots. • Rice blast is one of the most serious plant diseases. Monitoring the farm for rice blast is time consuming and labor intensive.

• Existing AI and IOT studies detected the plant diseases by using image and non-image hyper spectral data, which requires manual 11 • Operations to take the photos and extract the data.

• We develop the Rice Talk project that utilizes non-image IOT devices to detect rice blast.

• Unlike the image-based plant disease detection approaches, our agriculture sensors generate nonimage data that can be automatically trained and analyzed by the AI mechanism in real time

. • We also propose a spore germination mechanism as a new feature extraction model for agriculture.

• In the current implementation, the accuracy of the Rice Talk prediction on rice blast is **89. 4**%..

Classification of Paddy Rice Using a Stacked Generalization Approach and the Spectral

• Description: • Paddy rice is a key stable food, accounting for around 20% of the world's food supply.

• Paddy rice has a significant ecological impact on the region.

• Using coarse spatial resolution pictures [Moderate Resolution Imaging Spectral Radiometer, (MODIS)], this work presents a layered generalization and spectral mixture strategy to map paddy rice

. • The time series MODIS enhanced vegetation index images, phonological variables, land surface water index, elevation, and slope images are all used in this method to create the best feature combination, which is then used to map paddy rice using the stacking algorithm.

• The stacking model outperforms single classifiers in terms of paddy rice categorization 13 accuracy. Furthermore, the MODIS-derived rice map created using the layered generalization methodology and the spectral mixture method shows a high correlation coefficient (R2 =0.9975) with government statistics.

• The results show that the suggested method may be used to map huge areas of paddy rice utilizing coarse spatial resolution photos.

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

In order to classification of the rice seeds, they used to select a sample of rice seeds from a batch to perform manual screening to classify the rice seeds. Some of the approaches utilize appearance-based features like shape, length, width extracted from RGB images. Some utilize spectral features extracted from Hyper spectral imaging system.

.

3.1.1 Disadvantages of Existing System

Manual approach
Less accurate
Less efficient
Labor-consuming
Time-consuming.

3.2 Proposed System

In this project, a method is used to classify the rice seeds by the combination of spatial features and spectral features extracted from RGB and HSI system using RF classifier. This project is done by using 96 variety of rice seeds with 90% of accuracy

3.2.1 Benefitsfrom The Work

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

. • It shows that the combination of features gives discrimination ability and classification performance

3.3 Proposed System Design

In this project work, I used these modules and eachmodulehas own functions, such as:

1. Dataset Description and Processing
2. Spatial Feature Extraction
3. Spectral Feature Extraction
4. Dimensionality Reduction
5. Rice Seed Classifier

DATASET DESCRIPTION AND PROCESSING

In this project 96 variety of rice seeds are been used. These 96 varieties of rice seeds are divided into 2 batches each containing 48 individual rice seed samples. Each set 48 seeds were sended through RGB and HSI camera to extract spatial and spectral features. Each of the 96 variety of rice seeds were captured in 2 hyper spectral data cubes and 2 high resolution RGB images. This process is to extract spatial features and spectral features. The rice seeds which are captured in RGB and HIS camera are calibrated and normalized for obtaining lens and planar effects using rotation and transformation matrix. 22 Next the RGB data is segmented, segmentation is done on high spatial resolution RGB images to ensure that the complete kernel is captured. The spatial features are obtained from the binary masks for each rice seed in the RGB segmented data. The masks obtained from the process of RGB segmentation are transformed to HSI space to segment rice seeds in the HSI data. The segmented seeds of the HSI data are used to extract spectral features

SPATIAL FEATURE EXTRACTION

Spatial features are extracted from the binary masks of each and every rice seed in the RGB segmented data. The spatial features are selected due to their effectiveness among the species.

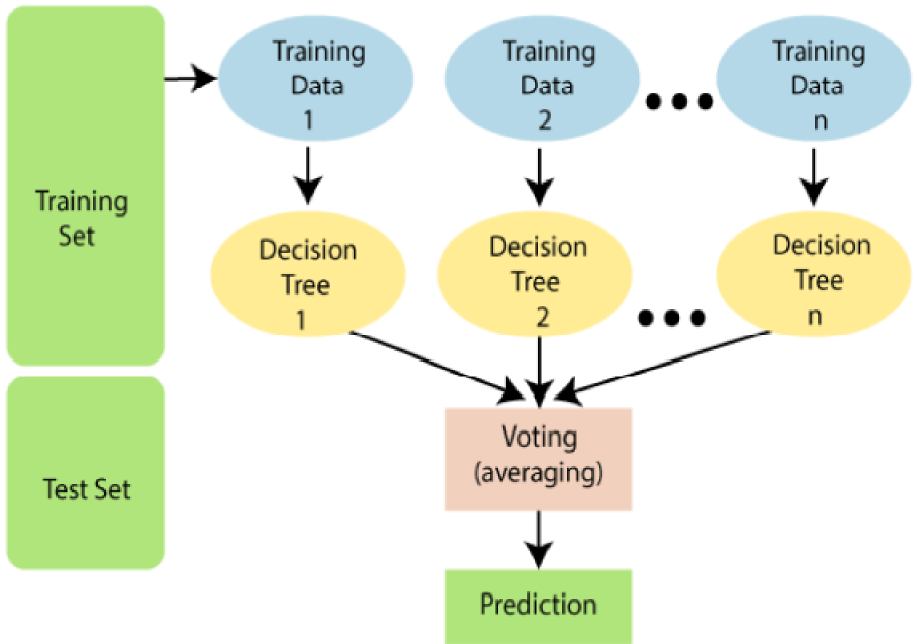
- It contains of six attributes they are
- F1: area
- F2, f3: major axis length and minor axis length
- F4: f3/f2
- F5: perimeter/area
- F6: Eccentricity: foci distance/major axis length.

SPECTRAL FEATURE EXTRACTION

The segmented seeds of the HSI data are used to extract spectral features. After normalizing the spectral data, the masks obtained from the rice seeds segmentation in RGB are sanded to HSI using transformation matrix for the HSI seed segmentation. From the HSI seed segmentation spectral features are extracted. Extraction of spectral features consists of two main phases they are data correction and feature extraction. yλ (x), x ∈ X, λ ∈ 3 where lambda stands for wave length belonging to^, x represents pixel in X, where X in 2-D coordinate by row m and column n.

DIMENSIONALITY REDUCTION

The collected is of high dimensionality. LDA is a statistical tool which is used to reduce dimensionality of data. LDA is an algorithm uses labels provided in the data set used to select features which are of high degree of variation among the 90 species of the dataset.



RICE SEED CLASSIFIER

The classification of rice seeds detect whether the species belong to a batch or not. 46 Now we will classify the rice seeds from spatial and spectral features which have been extracted.

1.spatial features only

2. spectral features only 3. combination of spatial and spectral features

4. Combination of LDA components extracted spectral data fused with spatial data. For the classification of rice seeds, we use RF classifier algorithm, which will give better classification results. Random forests contain many decision trees. Trees are allowed to grow to maximum extent after

selection of the best split from subset of features. Random forests make predictions on new data and are able to process large datasets in a fast, efficient way while achieving high classification accuracy.

4. ARCHITECTURE

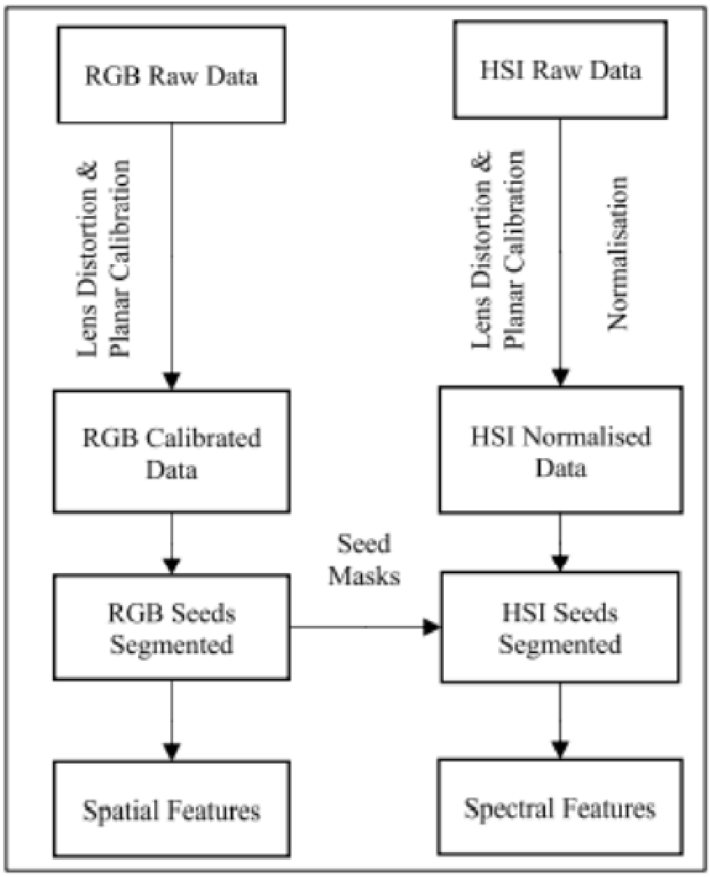


Fig 1: Architecture diagram

5. EXPERIMENTAL RESULTS AND ANALYSIS

The RF classifier performance is evaluated on the basis of three approaches they are average precision P, average recall R, and F1 Score.

P = tp/ tp + fp R = tp/ tp + fn

Where tp is true positive , fp is false positives, tn is true negatives, fn false negatives. F1 is the mean of the precision and recall.

F1score =2 ∗ P ∗ R/ P + R

Classification of rice seeds is done in three scenarios In the first scenario, all the 90 species of rice seeds are trained using RF classifier. In the second scenario, we take 6 subsets from the dataset and perform RF classifier where each and every subset consists of 6 species. 48

In the third scenario, we select 6 subsets from the dataset of 90 species and perform RF where each subset consists of random number of species like 6, 15, 30, 60, 90. This is to see how increasing in the number of species in the dataset affects classifier performance.

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

ANALYSING PERFORMANCE ON 90 SPECIES.

In this we train all the 90 species using RF classifier algorithm. 49 In the above table average precision, average recall and f1 score will decrease in Spatial, precision, recall, f1 score will increase in spectral. LDA is used to reduce the dimensionality of the spectral features. The outputs of the LDA are separately combined with the spatial features extracted from our data set. These data were trained using RF classifier. The classification results obtained using this approach is to differentiate between 90 species.

ANALYZING SELECTED SUBSET OF 6 SPECIES.

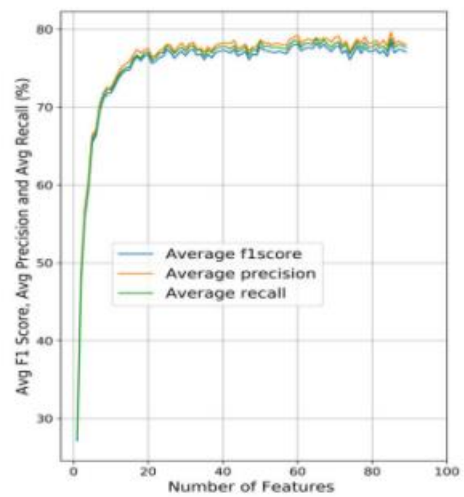
In this scenario we select 6 subsets from our data set and train with RF classifier where each subset consists of 6 species. We use LDA to extract spectral features from the subset and combine with spatial features. These 6 subsets were separately used to train with RF classifier and provide classification results. 50 The classification results while using 6 subsets is increased up to 61%.

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

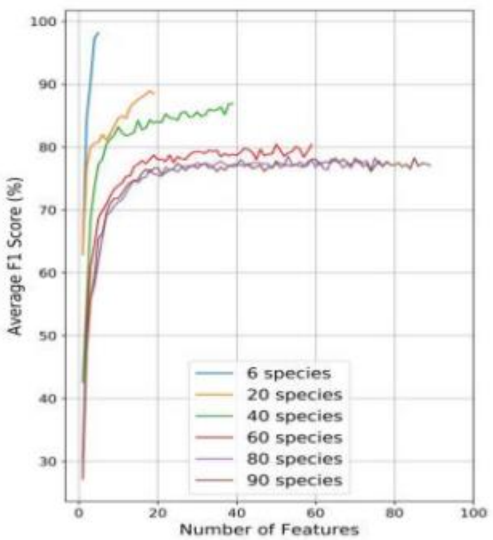
Subsets	Average precision (%)	Average Recall (%)	Average F1 Score (%)	Min–Max F1 score (%)
1	96.03	96.46	96.18	89.66-100
2	96.23	96.31	96.21	89.47-100
3	98.59	98.33	98.42	97.30-100
4	98.55	97.93	98.17	95.45-100
5	96.39	96.73	96.52	92.86-100
6	61.99	61.12	61.29	35.71 - 85.71

ANALYZING THE SUBSETS OF VARIOUS SPECIES

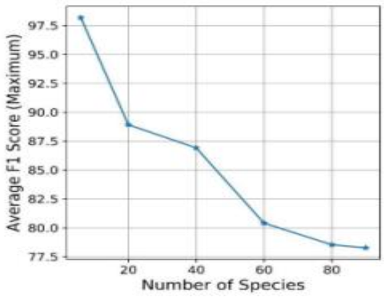
In this scenario we select 6 subsets from our dataset of 90 species. This time we vary number of species to each and every subset like 6, 20, 35, 50, 70. We apply LDA algorithm to extract spectral features from each and every subset and combine with spatial features. The classification is of 90% of accuracy.



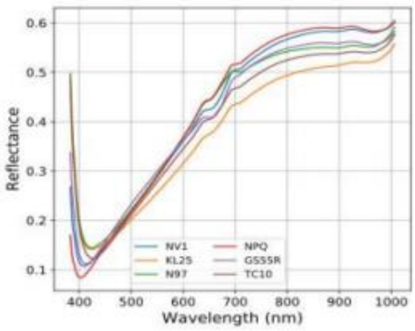
Results using the output of LDA and 90 species



Results of F1 score of six subsets of various species



F1 score against number of species.



Average spectral features of some species

6. CONCLUSION

In this project a method is used to automatically screen and classify the rice seeds. 69 Many numerous rice seeds were selected and spatial and spectral features extracted from the images and constitute the dataset. Experimental results show very good classification results in training with different species..

FUTURE WORK

Future work will focus on the similarity of rice seeds and the classification of rice seeds with more accuracy.

7. REFERENCES

- Analysis and identification of Rice adulteration using Terahertz Spectroscopy 70 and Pattern Recognition Algorithms.Chaoli,bin li ,dapenye 2018.
- 2.Rice blast detection using internet of things and artificial intelligence technologies. WenLiang Chen, Yi-Bing Lin 2019.
- 3.Classification of Paddy Rice Using a Stacked Generalization Approach and the Spectral Mixture Method Based on MODIS Time Series. Meng Zhang , Huaiqing Zhang 2020.
- 4.Modeling and Visualization of Rice Roots Based on Morphological Parameters LE YANG , JUN PENG 2020.