Classification of Rice Variants using Machine Learning

S Sharanya Bharghavi (Reg no: 11903126)

School of Computer Science and Engineering,

Lovely Professional University Phagwara, Punjab, India

ssharanyab@gmail.com

Abstract: Rice, is one of the world's most extensively grown and consumed cereal crops and is also one of country's most important forms of nourishment as it has high nutritional value and low cost. Rice goes through a cleansing procedure, colour sorting, and categorization on its way from the farm to our table. There are two major varieties of rice- Osmancik, and Cammeo. In order to distinguish between the two rice species, in this research, a modernized vision framework was created to recognize the two rice species. For classification purposes, KNN and SVM models were used.

Keywords: Rice classification, machine learning, SVM, KNN

1. Introduction:

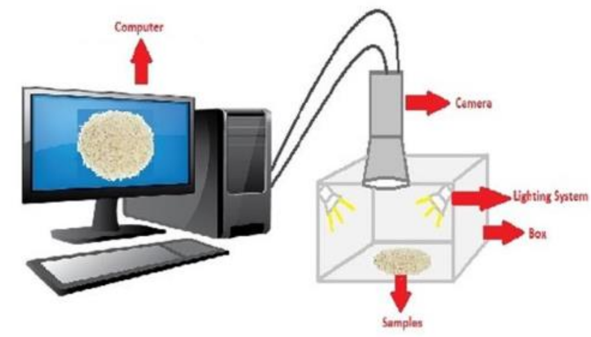
When it comes to grain production around the world, rice is the most significant product, followed by wheat and corn. Rice has a high glucose and starch content. Rice is an important part of human nutrition around the world as it is both healthy and cost-effective. [1]. In our country, many quality standards for rice production are made available. These include physical appearance, cooking characteristics, scent, taste, and smell, as well as efficiency and properties. It is the physical appearance that comes to mind from the criteria that stand out in the rice types that are sold packed on market shelves from the perspective of the end consumer [2]. After manufacturing, technical methods are required type determination, and separation of various quality aspects are inefficient and time-consuming, particularly for those with high production volumes. As a result, current research on cereal items employing computer vision systems and image processing techniques reveal that the products are scrutinised in terms of a variety of physical attributes, including colour, texture, quality, and size.

2. Understanding the dataset

In this study, I have used a dataset which contained images of rice that were obtained and later processed for classification purposes classification. A sum of 3810 rice grain pictures was taken for the two species. The generated photos are first converted to grayscale images, then binary images, and finally noise is removed from the image. The resulting pictures were then subjected to several morphological feature inference techniques in the following step. The rice categorization process was carried out utilising two different machine learning approaches in the modelling phase.

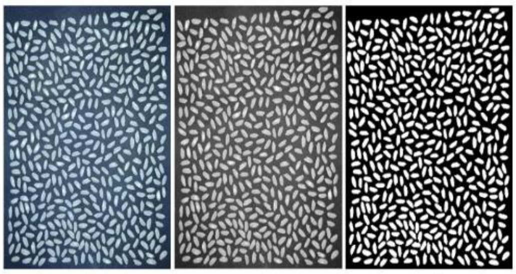
2.1 Image Acquisition:

For obtaining the images in the dataset, the following method was used.



The system's image-taking camera is mounted on a box with a lighting mechanism. The box is constructed in such a way that it does not receive light from the outside, preventing shadow formation on the image generated in this manner. For ease of processing, the box floor has been set to black. The rice samples were taken and transferred to a computer system, where they were recorded.[7]

2.2 Image Processing



In order to prepare for morphological feature inference, the images captured by the camera were transformed to grayscale and binary images, and then noise was removed. [7]

2.3 Feature Extraction:

Feature extraction's most effective morphological features and explanations:

S. no	Name	Explanation
1.	Area	The number of pixels within the rice grain's boundaries is returned.
2.	Perimeter	The circumference is calculated by computing the distance between pixels along the rice grain boundaries
3.	Major Axis Length	The main axis distance, which is the longest line that can be drawn on the rice grain, yields.
4.	Minor Axis Length	The shortest line that can be drawn on a rice grain, also known as the tiny axis distance, yields.
5.	Eccentricity	It determines the roundness of an ellipse with the same moments as a rice grain.
6.	Convex Area	The pixel count of the smallest convex shell of the rice grain's region is returned.
7.	Extent	The ratio of the rice grain's region to the bounding box pixels is returned.

3. Developing Model

3.1 Support Vector Machine (SVM)

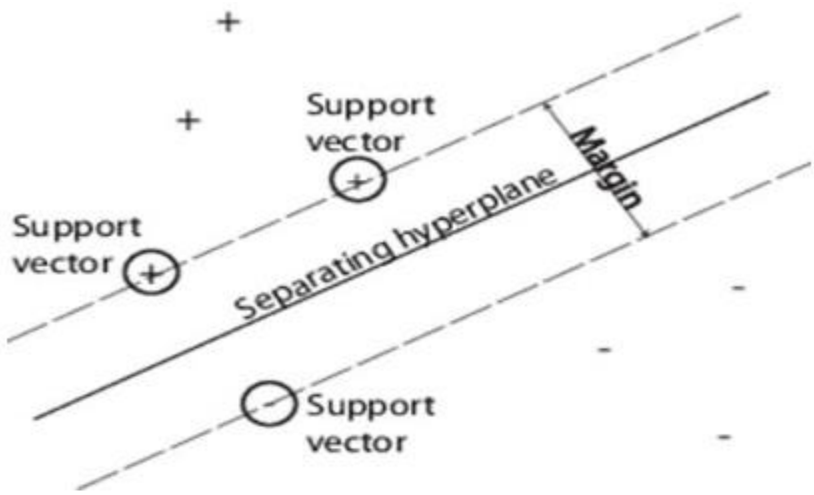


Figure 1: SVM

SVM is a core-based method for classification and regression that creates a hyper plane. SVM can categorise data as linear in two-dimensional space, planar in three-dimensional space, and hyperplane in multidimensional space using separation methods [3,4]. The classification procedure is carried out using SVM by determining the best hyper plane that divides the data. The hyper plane with the biggest margin between the two classes is the best for an SVM [5].

SVMs have characteristics that are comparable to those of other machine learning methods. It resembles neural networks in appearance but more closely resembles the k-NN algorithm in execution. SVM identifies its neighbours based on sample data fed to the algorithm and expects that new data predictions are produced [6].

3.2 K-Nearest Neighbours (k-NN)

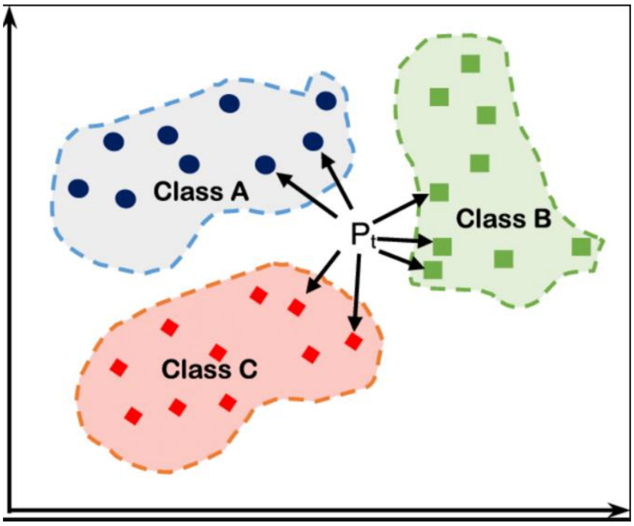


Figure 2: KNN

K-Nearest Neighbours (k-NN) is a machine learning method for large-scale training. Each point is conceptually plotted in a wide-dimensional space using the k-NN algorithm, with each axis in space corresponding to a separate variable. Because there is a limited amount of data to test, the test data is processed one by one alongside the rest of the data. In terms of all the characteristics measured, the test data will have a lot of neighbours that are similar to it. As a result, the k piece data that is closest to the test data is chosen. As a consequence, the selected data is examined to see which class contains the most data, and the data tested is assigned to that class. The algorithm's flexibility makes it easier to manage data sets that aren't explained by linear or nonlinear relationships due to their complexity. The k-NN algorithm's most significant drawback is that, unlike other data mining algorithms, it does not indicate which variables are important in new predictions.

4. Implementation:

Each grain in the dataset has been assigned seven morphological characteristics.

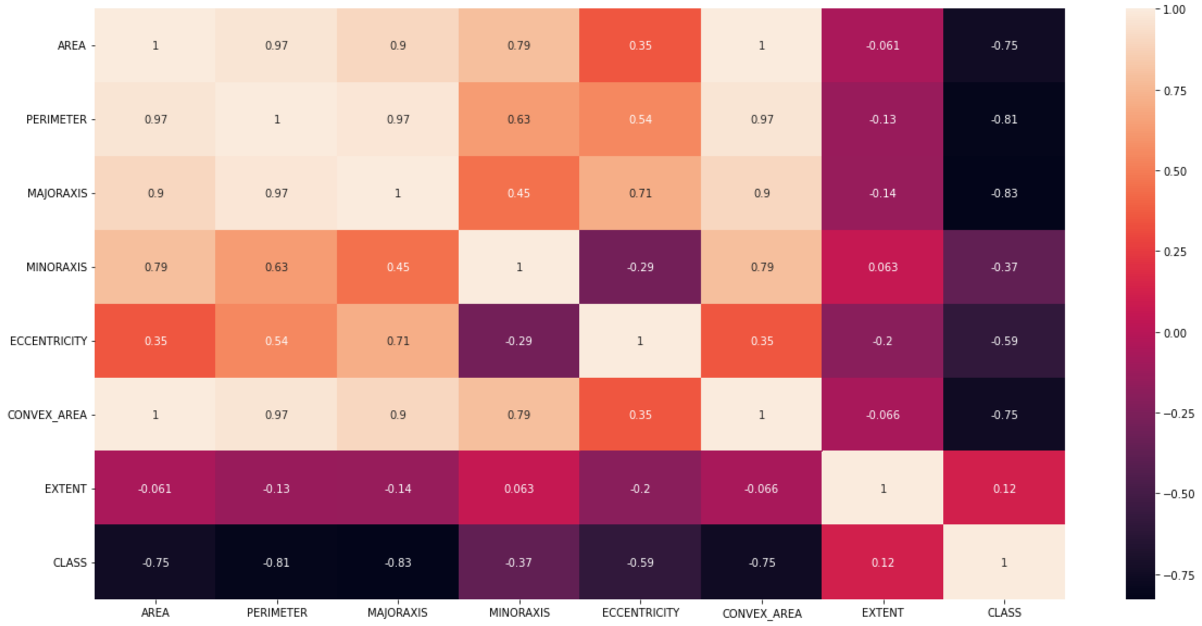


Figure 3: Correlation

In the dataset correlated labels convex area, perimeter, major axis, minor axis was dropped as they were corelated with area.

KNN and SVM models were created. On all models utilised, the k value for cross validation is set to 42.

Code is implemented in python using sklearn (repo link: [ssharanyab/Rice-Classification (github.com)](https://github.com/ssharanyab/Rice-Classification))

5. Results:

The Confusion Matrix values for all the methods utilised are listed here:

Algorithm	Confusion Matrix	
SVM	313	41
	20	388
KNN	311	43
	22	386

Metrics such as Accuracy, Precision, F1-Score, recall are generated for classifying performance metrics utilising the complexity matrix for each model. The following table shows the results:

	SVM	KNN
Accuracy	0.92	0.91
Precision	0.94,0.90	0.93,0.90

Recall	0.88, 0.95	0.88,0.95
F1-Score	0.91, 0.93	0.91,0.92
Support	354, 408	408

As per the reports I derived, KNN algorithm predicts with an accuracy if 0.91 while support vector machine gives 0.92.

References

[1] Sapaloglu, A., The structure of marketing channels and rice marketing margins in the rice production-consumption chain: the example of the province of Edirne.2015, Namik Kemal University

[2] Mutafcilar, Z.C., Molecular characterization of registered paddy varieties in Turkey.2019, Trakya University Graduate School Of Science.

[3] Abhang, P., B. Gawali, and S. Mehrotra, Technical aspects of brain rhythms and speech parameters.Introduction to EEG-and Speech-Based Emotion Recognition, 2016: p. 51-79

[4] Kaynar, O., et al. Emotion Analysis with machine learning methods.in International Artificial Intelligence and Data Processing Symposium (IDAP'16), September. 2016

[5] Mathworks. Support Vector Machines for Binary Classification. 2019 [cited 2019 12 September 2019]; Available from: https://ch.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html#bsr5b42

[6] Miner, L., et al., Practical predictive analytics and decisioning systems for medicine: Informatics accuracy and cost-effectiveness for healthcare administration and delivery including medical research. 2014: Academic Press

[7] Classification of Rice Varieties Using Artificial Intelligence MethodsIlkay CINAR*1, Murat KOKLU2