

# Crop Classification based on Soil Information

Krish L Sharma

*School Of Computer Science and Engineering  
Vellore Institute Of Technology, Bhopal  
Bhopal, India  
krish.sharma2020@vitbhopal.ac.in*

Yashaswi Patel

*School Of Computer Science and Engineering  
Vellore Institute Of Technology, Bhopal  
Bhopal, India  
yashaswi.patel2020@vitbhopal.ac.in*

Ankur Jain

*School of Computing Science and Engineering  
Vellore Institute of Technology, Bhopal  
Bhopal, India  
ankur.jain@vitbhopal.ac.in*

**Abstract**—This document is a model and instructions for  $\text{\LaTeX}$ . This and the `IEEEtran.cls` file define the components of your paper [title, text, heads, etc.]. **\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

**Index Terms**—component, formatting, style, styling, insert

## I. INTRODUCTION

This document is a model and instructions for  $\text{\LaTeX}$ . Please observe the conference page limits.

## II. LITERATURE REVIEW

In [1] paper, it was observed that C. Chen and H. McNairn optimised the neural network design process by using the hierarchical procedure. The hierarchical procedure helps in finding the best network structure and the optimal network parameters. These questions of the neural network are very important and difficult to find as it is a trial and error-based finding which is impractical to try all the combinations. Hence, the hierarchical procedure is being used. The procedure is as follows:

- 1) Keep the learning rate and the momentum as constant and try all the design structures and training iterations.
- 2) After finding the best design structure, move forward on trying all the learning rate and momentum combinations.

The study has been done on 7 different models with 3 one-layered models and 4 two-layered models. Out of the 7 models, 2 models were found optimal which were: **6-18-3** and **6-12-6-3**. Among the two models, **6-12-6-3** was selected as the best model because it had the highest accuracy and was more stable amongst the models for classification. Moreover, the training had a dynamic range of training time which can be easily optimised. The conclusion being that 2 hidden layer architecture was more stable and giving high accuracy if they have the following rules for the hidden layers:

- 1) The size of the first hidden layer should be greater than that of the input layer and the second hidden layer.
- 2) The second hidden layer should be less than that of the first hidden layer but not close to the output layer size.

Moreover, [2] has tried all the available neural networks with a range of numbers [1,10] of hidden layers based on classification. According to the study, it has been concluded that all the neural networks for classification that have been used are made of 2 hidden layers and with sigmoid function. The paper has trained all types of the neural networks with 70% training data, 15% validation data and 15% testing data. Trying to achieve more accuracy they tried to change the number of hidden layers with a range of 10 hidden layers. The study has trained all with combinations of number of hidden layers and finally came to a conclusion that 10 layered neural network has shown promising accuracy and optimality but expensive with respect to computational due to multiple layers.

This [3] study aimed classification by using three different classifications. These models were logistic regression, artificial neural networks, and a stacking model. It was observed that a total of 34 features were obtained out of which 30 important features were selected. The authors evaluated the performance of the proposed approach using various metrics, including accuracy, precision, recall, and F1-score. The stacking model combined the outputs of the logistic regression and ANN models. The results showed that the Stacking Model outperformed the other two algorithms.



Fig. 1. [3] Stacking Model Flowchart

For classifying different plants and crops, which include barley, pomegranate, cherries, grapes, pistachio, red pepper, tomato, wheat, cotton, apricot, sunflower, tangerine, corn, beans and apple, this [4] study relied heavily on a pre-trained Convolutional Neural Network model. For the classification

task, the authors employed convolution neural network, and support vector machine techniques. Instead of connecting each node in the current layer to the nodes in the previous layer, each node in this case was only connected to a local region of the input volume. The CNN model was observed to learn and optimise the filters in each layer via the back-propagation mechanism. The classification performance of a CNN-based approach was compared to that of an SVM classifier with different kernels and features. SVM classifier was tested with RBF and polynomial kernels, as well as various types of features, such as four-connected LBP, GIST with 12 orientations per scale, and 4 blocks methods. The accuracy rates fluctuated between 69.81% to 97.47%. It was discovered that features learned from the CNN framework outperform state-of-the-art solutions that make use of carefully selected hand-crafted features. Even with different kernels and features, the SVM classifier falls short of the performance of our CNN-based approach.

The purpose of this [5] study was to classify wheat crops using multi-temporal images. The study looked at two approaches for the effective classification of multi-temporal images: Multi Label Classification(MLC) with different strategies:

- 1) Sequential MLC
- 2) PCA-based MLC
- 3) Iterative-based MLC
- 4) ANN

In terms of feasibility of implementation, completeness of classification, and labelling reliability, Iterative-based MLC was found to be superior to the other two MLC strategies. ANN demonstrated the ability to perform classification in just a single stretch, whereas traditional MLC produced significantly less reliable results. Despite the fact that Iterative-based MLC resulted in a significant improvement in classification over MLC.

### III. OBJECTIVE

The following are the objective of our study on classification of the crops based on the soil nutrients:

- 1) **Building an accurate model:** The first objective is to build a deep learning-based neural network model that accurately predicts the type of crop that can be grown on a given piece of land based on the soil information extracted from the land. The model should be able to learn the complex relationships between the soil characteristics and the crop types and generalize well to new unseen data.
- 2) **Improving crop yield:** The second objective is to improve crop yield by providing accurate information about the types of crops that can be grown on a given piece of land. By accurately predicting the crop type, farmers can make informed decision about which crops to plant on their land, leading to better crop yields and potentially increased profitability. This can also have environmental benefits by reducing the used of fertilizers

and other inputs that may be unnecessary for certain crops.

### IV. PROBLEM FORMULATION

Let  $X$  be a set of  $n$  samples, where each sample  $x_i$  represents soil information extracted from a piece of land. Each sample  $x_i$  contains  $m$  feature  $\{x_{i1}, x_{i2}, \dots, x_{im}\}$ , where each feature represents a soil characteristics such as pH level, nutrient content, etc.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
5	69	37	42	23.058049	83.370118	7.073454	251.055000	rice

Fig. 2. Dataset

Let  $Y$  be a set of  $n$  labels, where each label  $y_i$  represents the type of crop that can be grown on the corresponding land. The label  $y_i$  belongs to a set of  $k$  possible classes  $\{y_1, y_2, \dots, y_k\}$ , where each class represents a specific crop as seen in Fig. 2.

```
crop_dataset['label'].unique()
array(['rice', 'maize', 'chickpea', 'kidneybeans', 'pigeonpeas',
       'mothbeans', 'mungbean', 'blackgram', 'lentil', 'pomegranate',
       'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple',
       'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'],
      dtype=object)
```

Fig. 3. Crops for Classification

The objective is to build a deep learning-based neural network model that can learn the mapping between the soil information  $X$  and the crop label  $Y$ , and predict the correct crop label for new samples of soil information.

Let  $f(X, \theta)$  be a neural network model with parameters  $\theta$ . The model takes as input a sample  $x_i$  from  $X$  and outputs a probability distribution over the  $k$  possible classes. The predicted label for  $x_i$  is the class with the highest probability in the output distribution.

The model is trained by minimizing a loss function  $L(Y, f(X, \theta))$ , which measures the difference between the predicted probability distribution and the true label  $y_i$ . A common choice for the loss function is the categorical cross-entropy loss:

$$L(Y, f(X, \theta)) = - \sum_i^k y_i \log(f(x_i, \theta))$$

where  $f(x_i, \theta)$  is the predicted probability distribution for  $x_i$  and  $y_i$  is the true label for  $x_i$  encoded as a one-hot encoder.

The parameters  $\theta$  are updated using gradient descent or a variant of it, such as Adam. The model is trained using a training set of labelled soil samples and evaluated on a

separate validation set to monitor its performance and prevent overfitting. Once the model is trained and validated, it can be used to predict the crop label for new unseen soil samples.

## V. METHODOLOGY

The methodology to solve the problem of classifying crops based on soil information extracted from land using a deep learning-based neural network model can be broken down into the following steps:

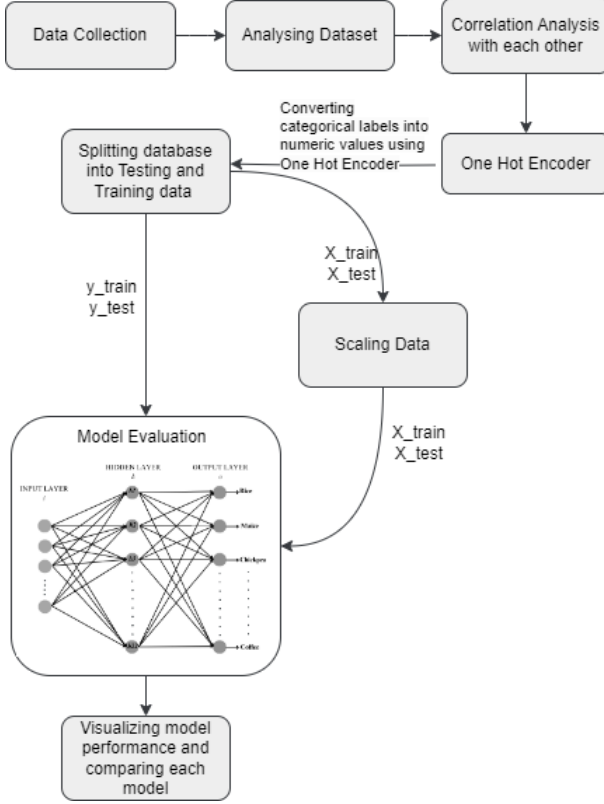


Fig. 4. Flowchart of the model

### A. Data collection and preprocessing

The first step is to collect soil data and corresponding crop labels for a variety of lands. This can be done by taking soil samples from different locations and testing them in a laboratory to extract various soil characteristics such as pH level, nutrient content, etc.

```
crop_dataset.columns
Index(['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall', 'label'], dtype='object')
```

Fig. 5. Different soil characteristics

The crop labels can be obtained by surveying the lands or by consulting with local farmers. Once the data is collected, it needs to be preprocessed, which may involve cleaning, normalizing, and splitting it into training, validation, and testing sets.

### B. Model selection and design

The next step is to select an appropriate deep learning-based neural network model and design its architecture. The model architecture may involve selecting the number and type of layers, the number of neurons in each layer, and the activation functions to be used. The model architecture can be optimized through hyperparameters (node size of each layer, number of epochs, learning rate, etc.) tuning to improve its performance. In this case we have used 8 Models with a combination of sigmoid, tanh and relu activation function in the output layer and the hidden layer. The size of hidden layer is 22 which is same as that of the output layer. Moreover, the hyperparameters such as number of epochs and metrics are taken 200 and accuracy, respectively.

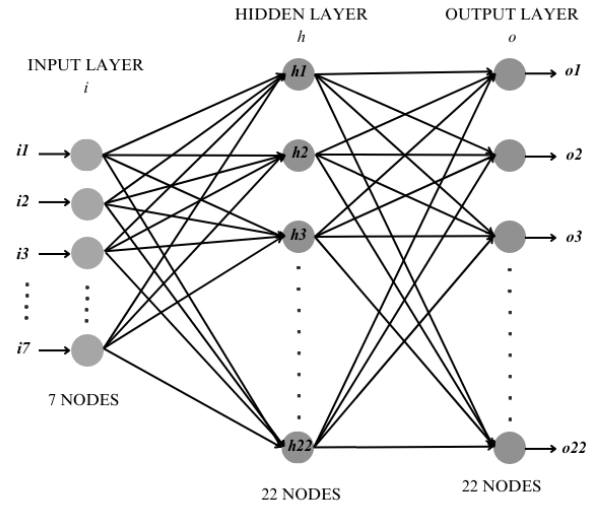


Fig. 6. Neural Network Architecture

### C. Training and validation

The model is then trained using the training set and validated using the validation set. During training, the model learns the complex relationships between the soil characteristics and the crop types. The model's performance is monitored during training and validation using metrics such as accuracy, precision, recall, and F1 score. The training can be stopped when the model's performance on the validation set starts to deteriorate, indicating overfitting. The training data is 80% and the testing data is 20% of the data.

### D. Model evaluation

Once the model is trained, it is evaluated using the testing set. This gives an estimate of the model's performance on unseen data. The evaluation metrics can be compared to those obtained during training and validation to check for any significant drops in performance.

	N	P	K	temperature	humidity	ph	rainfall
1994	106	46	20	23.438217	78.633888	6.200672	81.150721
372	11	61	21	18.623288	23.024103	5.532101	135.337803
576	24	45	19	26.858519	48.824639	5.952385	34.742646
211	43	66	79	19.462340	15.225390	7.976608	74.585651
520	11	44	17	26.340433	55.591604	8.016211	35.105120
1211	27	145	205	9.467960	82.293355	5.800243	66.027652
234	31	79	75	18.820225	16.107479	8.204862	89.731194

Fig. 7. Training Data

	N	P	K	temperature	humidity	ph	rainfall
1320	105	14	50	26.214884	87.688398	6.419052	59.655908
1367	91	12	46	24.644585	85.499382	6.343943	48.312190
1291	14	121	203	9.724458	83.747656	6.158689	74.464111
264	52	73	79	17.257695	18.749440	7.840339	94.002872
728	32	66	17	34.946616	65.267740	7.162358	70.141514
70	61	52	41	24.976695	83.891805	6.880431	204.800185
565	11	41	19	26.859113	41.814208	5.131779	44.138271

Fig. 8. Testing Data

### E. Model deployment

After the model is trained and evaluated, it can be deployed to predict the crop type for new soil data. The deployment can be done using a web or mobile application that takes in the soil data as input and returns the predicted crop type as output.

### F. Model maintenance and updates

The final step is to maintain and update the model as new data becomes available. The model's performance can be monitored regularly to check for any deterioration in its performance, and updates can be made to the model's architecture or hyperparameters as necessary. Additionally, new data can be added to the training set to improve the model's performance over time.

Moreover, this model can be used on IoT based devices to make it more reachable to the farmers. Also, same can be used to make a recommendation system using Fuzzy Logic system. This will help the farmers to gain good knowledge about their farmlands. We can fuse the weather and climate pattern details to the dataset to upgrade the model into recommendation system.

## VI. RESULT

We have applied in total 8 models and in figure we can see the graph showing the performance of each model. For each model the graph on left shows the accuracy curve and on the right the loss curve.

All the model performance are listed below with their accuracy:

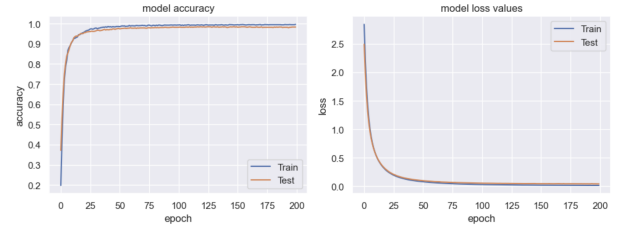


Fig. 9. Model 1 Performance

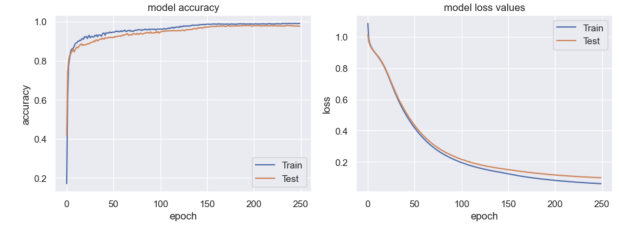


Fig. 10. Model 2 Performance

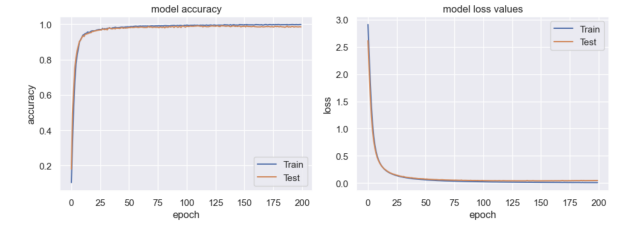


Fig. 11. Model 3 Performance

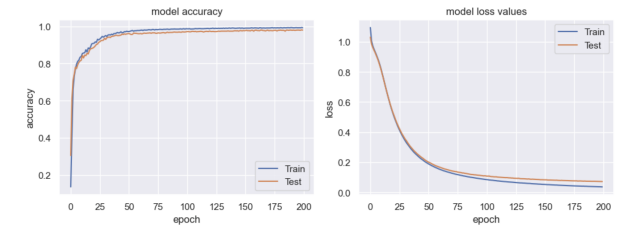


Fig. 12. Model 4 Performance

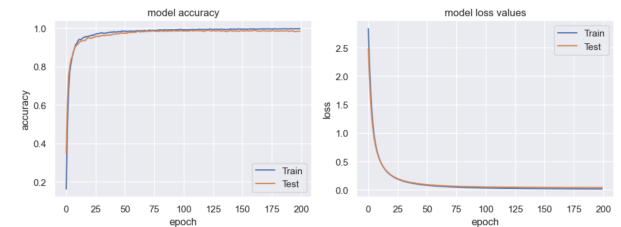


Fig. 13. Model 5 Performance

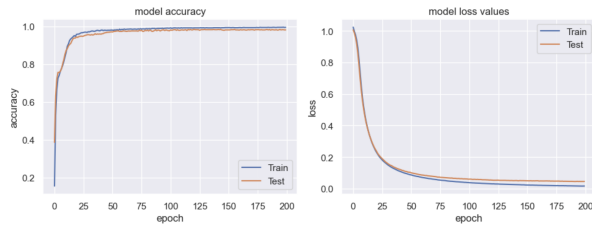


Fig. 14. Model 6 Performance

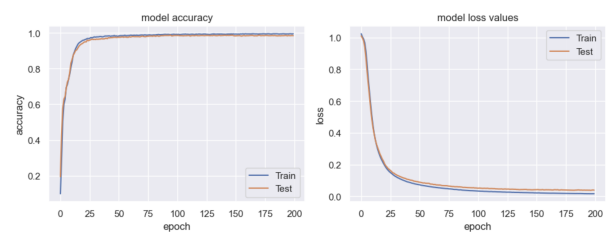


Fig. 16. Model 8 Performance

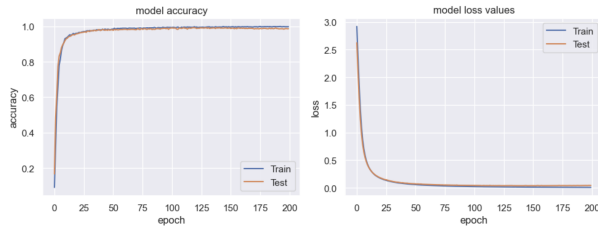


Fig. 15. Model 7 Performance

- 1) Tanh-Sigmoid-CategoricalCrossEntropy: Training Accuracy **99.772%** and Testing Accuracy **98.181%**.
- 2) Tanh-Sigmoid-CategoricalHinge: Training Accuracy **98.712%** and Testing Accuracy **97.272%**
- 3) ReLU-Sigmoid-CategoricalCrossEntropy: Training Accuracy **99.848%** and Testing Accuracy **98.522%**
- 4) ReLU-Sigmoid-CategoricalHinge: Training Accuracy **99.091%** and Testing Accuracy **97.5%**
- 5) Tanh-Softmax-CategoricalCrossEntropy: Training Accuracy **99.621%** and Testing Accuracy **98.522%**
- 6) Tanh-Softmax-CategoricalHinge: Training Accuracy **99.621%** and Testing Accuracy **98.182%**
- 7) ReLU-Softmax-CategoricalCrossEntropy: Training Accuracy **99.848%** and Testing Accuracy **98.977%**
- 8) ReLU-Softmax-CategoricalHinge: Training Accuracy **99.545%** and Testing Accuracy **98.522%**

Out of all the models tried and tested we can see that models {5, 7, 8} are giving very promising result out of which Model 7 is the best model out of all having a very good generalization and accuracy for the given classification problem.

## VII. CONCLUSION

## ACKNOWLEDGMENT

## REFERENCES

- [1] C. Chen, H. McNairn, "A neural network integrated approach for rice crop monitoring"
- [2] Abdulhamid Haidar, Haiwei Dong and Nikolaos Mavridis, "Image-Based Date Fruit Classification"
- [3] Murat Koklu, Ramazan Kursun, Yavuz Selim Taspinar and Ilkay Cinar, "Classification of Date Fruits into Genetic Varieties Using Image Analysis"
- [4] Hulya Yalcin, Salar Razavi, "Plant Classification using Convolution Neural Network"
- [5] C. S. Murthy, P. V. Raju and K. V. S. Badrinath, "Classification of wheat crop with multi-temporal images: performance of maximum likelihood and artificial neural networks"

- [6] Nguyen-Thanh Son, Chi-Farn Chen, Cheng-Ru Chen, Piero Toscano, Youg-Sing Cheng, Hong-Yuh Guo, Chien-Hui Syu, "A phenological object-based approach for rice crop classification using time-series Sentinel-1 Synthetic Aperture Radar (SAR) data in Taiwan"
- [7] Alex O. Onojeghuo, George A. Blackburn, Qunming Wang, Peter M. Atkinson, Daniel Kindred, Yuxin Miao, "Mapping paddy rice fields by applying machine learning algorithms to multi-temporal Sentinel-1A and Landsat data"
- [8] Zhiyuan Pei, Songling Zhang, Lin Guo, Heather McNairn, Jiali Shang, Xianfeng Jiao, "Rice identification and change detection using TerraSAR-X data"
- [9] A. Beyaz, R. Ozturk, and U. Turker, "Assessment of mechanical damage on apples with image analysis"
- [10] Oussama Aiadi, Belal Khaldi, Mohammed Lamine Kherfi, Mohamed Lamine Mekhalfi, Abdullah Alharbi, "Date Fruit Sorting Based on Deep Learning and Discriminant Correlation Analysis"
- [11] Jui-Feng Yeh, Kuei-Mei Lin, Chen-Yu Lin, Jen-Chun Kang, "Intelligent Mango Fruit Glade Classification Using AlexNet with GrabCut Algorithm"
- [12] Puteri Khatya Fahira, Zulia Putri Rahmadhani, Petrus Mursanto, Ari Wibisono, Hanif Arief Wisesa, "Classical Machine Learning Classification for Javanese Traditional Food Image"
- [13] C. Arun, Akshatha Prabhu, Mohammed Zeeshan, N. Shobha Rani, "A Study on Various Classifier Techniques Used in Image Processing"
- [14] Ahmed Chiheb Ammari, Lazhar Khriji, Medhat Awadalla, "HW/SW Co-design for Dates Classification on Xilinx Zynq SoC"
- [15] Adnan Ahmed Abi Sen, Nour Mahmoud Bahbouh, Ahmad B. Alkhodre, Ashwaq Mohammed Aldhawi, Fatmah Ahmad Aldham, Manar Ibrahim Aljabri, "A Classification Algorithm for Date Fruits"
- [16] Ghulam Muhammad, M. Shamim Hossain, Abdulsalam Yassine, "Tree-Based Deep Networks for Edge Devices"
- [17] Puteri Khatya Fahira, Ari Wibisono, Hanif Arief Wisesa, Zulia Putri Rahmadhani, Petrus Mursanto, Adi Nurhadiyatna, "Sumatra Traditional Food Image Classification Using Classical Machine Learning"
- [18] Anindita Septiarini, Hamdani Hamdani, Heliza Rahmania Hatta, Anita Ahmad Kasim, "Image-based processing for ripeness classification of oil palm fruit"
- [19] Hamdi Altaheri, Mansour Alsulaiman, Ghulam Muhammad, "Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning"
- [20] Hu-Lin Kuang, Leanne Lai Hang Chan, Hong Yan, "Multi-class fruit detection based on multiple color channels"
- [21] Ghulam Muhammad, "Automatic Date Fruit Classification by Using Local Texture Descriptors and Shape-Size Features"
- [22] A. M. Vyas, B. Talati, and S. Naik, "Colour feature extraction techniques of fruits: a survey"
- [23] R. Pandey, S. Naik, and R. Marfatia, "Image processing and machine learning for automated fruit grading system: a technical review"
- [24] I. Cinar and M. Koklu, "Classification of rice varieties using artificial intelligence methods"
- [25] S. Jana and R. Parekh, "Shape-based fruit recognition and classification"