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

Fruit Recognition Using Diffusion Maps & Machine Learning

[1]John J, [2]Midhun M J, [3]Logeshwaran P, [4]Dr. R. Nagendran
[1,2,3]Student, [4]Professor
[1,2,3,4]Computer Science & Engineering,
[1,2,3,4]Sri Ramakrishna Institute of Technology, Coimbatore, India

Abstract: Fruit Recognition is a key task in numerous applications, including agricultural production, food quality control, and inventory management. Addressing the growing need for automated and efficient classification, this project presents a comprehensive approach for automatic fruit recognition using diffusion maps and machine learning techniques. The proposed methods focuses on capturing high-quality images of fruits, which are then pre-processes to enhance image consistency by normalizing lighting conditions and improving overall image quality.

1. INTRODUCTION

Fruit recognition is essentially is essentially very important in agriculture and food quality control. However, the traditional method of sorting is quite a time-consuming task and tends to make mistakes. Modern progress in machine improved in efficiency as well as accuracy by automated means based on visual features like shape, and texture. The project will eventually develop an advance fruit recognition system, which applies diffusion maps for reducing the dimensionality and utilizing machine learning algorithms for classification purposes. It captures fruit images and pre-processes them for better quality to ensure normalized lighting conditions and extracts the features of which define every fruit type- shape, and texture patterns.

1.1. GENERAL INFORMATAION

The Significant features of this system are derived from the machine learning algorithm, CNN that is very effective in managing complex high-dimensional data and variations resulting from lighting and fruit ripeness. The optimization of computational efficiency results in classification through diffusion maps without any loss of accuracy. The Advanced feature extraction techniques combined with machine learning significantly enhance the performance of classification. This project is structured into a clear methodology where images undergo pre-processing to normalize conditions; thereafter, distinguishing reduction before classification; and classification is used at the end.

1.2. PROBLEM STATEMENT

The project tires to develop a robust fruit recognition system with the help of machine learning algorithm and dimensionality reduction using diffusion maps. The classification will be based on visual characteristics in terms of texture and shape, using algorithm like Conventional Neural Network (CNNs), the system will compare the different techniques and find the best one concerning the accuracy of fruit identification. Thus, it is able to handle the challenges of appearance variability and environment noise, hence making the results more reliable. Applications of the system area agriculture, automating the process of sorting and grading, and retail, inventory management and quality control, in order to ultimately improve efficiency and accuracy.

1.3. TECHNOLOGY

1. Python
2. Streamlit

2. LITERATURE SURVEY

1. **Hossam M. Zawbaa et al., (2014)** Proposed a Novel Method for Automatic Fruit Image Recognition System Based on Shape and Color**.** This paper introduces a system to identify and classify different fruit types based on key visual features, specifically shape and color. The system is structured into three main stages: pre-processing, feature extraction, and classification, each playing a critical role in ensuring accurate recognition. In the pre-processing phase, the fruit images are resized, a technique that reduces the color index and makes the images easier to handle for the next stages of processing. This step ensures that the system can manage the images more efficiently without losing important details necessary for classification. During the feature

extraction phase, advanced techniques like Scale Invariant Feature Transform (SIFT) are used to create a feature vector for each fruit image.

2. **Manali R et al., (2016)** Proposed a Novel Method for Color, Size, Volume, Shape and Texture Feature Extraction Techniques for Fruits. This paper will focus on the various techniques in feature extraction used in the grading and sorting systems of fruits, including important features: color, size, volume, shape, and texture. These features are very essential in improving post-harvest processes to ensure that fruits meet the quality standards of the market. The drawback of sorting by hand is laborious, time-consuming, costly, and subject to human errors and inconsistency. The paper points out several automated techniques that will be more efficient, cost-effective, and consistent in overcoming these. In the case of color detection, direct color mapping techniques are presented, which help in identifying the color of the fruit with much accuracy for grading purposes.

3. **Yamini M.A et al., (2017)** Proposed a Novel Method for Fruit Maturity Classification Via Image Analysis. This study focuses on using image processing techniques to classify the maturity of fruits, assessing their quality in line with USDA standards that categorize maturity into three stages: green, pink, and red. The researchers used MATLAB to analyze images of fruits taken from different angles, which helped assess key attributes like color and physical properties, such as firmness and rupture force, that are crucial for determining ripeness. The proposed image processing method was highly effective, achieving an impressive 89% accuracy when compared to traditional manual grading methods.

4. **Xiaoyang Liu et al., (2019)** Proposed a Novel Method for A Detection Method for Fruits on Color and Shape Features. This paper introducing a new method for detecting apples in orchard images, focusing on improving accuracy even under varying lighting conditions and in challenging environments. The method leverages color and shape features to better identify apples in real-world settings. It uses Simple Linear Iterative Clustering (SLIC) to break down the images into smaller sections called super-pixel blocks, with a particular focus on color to locate potential fruit regions. This allows the system to pinpoint areas where apples might be located more efficiently.

5. **Wenbo Wang et al., (2024)** proposed a Novel Method for Fruit Classification based on Diffusion Maps and Machine Learning. The paper presents an automated approach to fruit classification. Starting from image pre-processing using techniques like Gaussian filtering to smooth the image and convert it into grayscale for easier handling of data, these steps help enhance quality so that the next treatments can make an improved analysis. It extracts key features from these images, such as texture, shape, and wavelet transforms, in order to find patterns that distinguish different fruits. Then, it applies diffusion maps for dimensionality reduction of the data to remove redundant information, easing the classification process.

3. EXISTING METHODOLOGY

All the fruit recognition systems developed so far have gone one step ahead in automating fruit classification with the help of image processing and machine learning techniques. These analyse images of fruits for main features like shape and texture that distinguish one type of fruit from the others. Training is performed using a large dataset of images of various fruits, hence testing the model's ability to find unique characteristics that define each fruit type. Among the feature extraction methods, the most used for capturing visual features of fruits are normally color histograms and analyses of texture. It is then classified using machine learning algorithms, probably through Support Vector Machines, K-Nearest Neighbours, or Decision Trees. While these systems exhibit excellent performance under controlled conditions, they are yet to be optimized to serve effectively in on-site settings such as farms or food processing plants.

3.1. DISADVANTAGES

Fruit recognition has a lot of promise, yet there are some quite serious drawbacks to using these systems in more real and practical applications. Probably among the greatest of these is environmental sensitivity. These systems usually do quite well in controlled environments where conditions can be relatively stable in factors such as lighting and background. In the real world, lighting may change, the backgrounds may be cluttered, and even the appearance of the fruit may change with ripeness or damage. For example, a fruit, which would be colorful and fresh under ideal circumstances, might appear absolutely different under poorer lighting conditions or if it's bruised, which would cause it to be more challenging for the system to classify correctly.

4. PROPOSED METHODOLOGY

The System will start acquiring a comprehensive dataset images, which will form the foundation for training and testing the model. Following this, image pre-processing steps will be carried out, including resizing the image to a consistent size, reducing noise to improve image quality, and segmenting the images to isolate the fruits from the background. These pre-processing steps are essential to enhance the clarity of the visual features, making it easier for the system to identify key attributes.

Next, important visual features such as texture and shape will be extracted from the images. These features serve as the primary basis for distinguishing between different fruit types. To manage the complexity of the data, dimensionality reduction using diffusion maps will be applied. This technique will simplify the data while maintaining the crucial information necessary for accurate classification.

4.1. ADVANTAGES

1. **High Accuracy with Feature Extraction**: By leveraging difference maps along with machine learning, the model can accurately distinguish between subtle variations in fruits' shape, texture, and color, leading to precise recognition even for visually similar fruits.
2. **Automated and Efficient Sorting**: This technology can be used in agricultural or retail industries to automate the sorting process, ensuring faster and more consistent classification of fruits based on quality, ripeness, or type.
3. **Adaptability to New Fruit Types**: The model can be retrained or fine-tuned to recognize new or exotic fruits, making the system scalable and adaptable to diverse datasets without significant redesign.
4. **Cost-Effective Quality Control**: Using machine learning for fruit recognition reduces the need for manual inspection, lowering operational costs for businesses while ensuring high standards of quality control in food production and processing environments.
5. **Real-Time Recognition**: Machine learning models, once trained, can quickly recognize fruits in real-time, making the system highly efficient for applications such as automated checkout systems, fruit packaging lines, or smart kitchens.

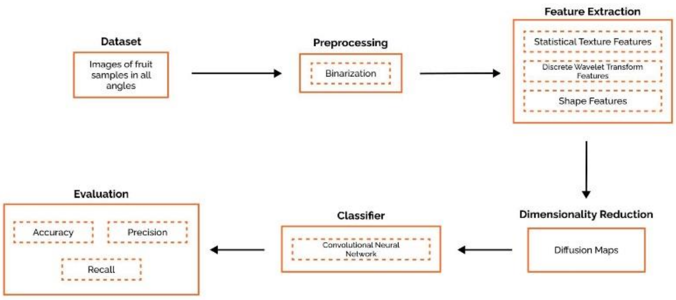


Fig 4.1. Block Diagram

5. RESULTS

SL No.	Fruit Name	Image
1.	Apple	
2.	Banana	
3.	Avocado	
4.	Cherry	
5.	Kiwi	
6.	Mango	
7.	Orange	
8.	Pineapple	
9.	Strawberry	
10.	Watermelon	

Table 5.1 Classes of Fruits with Sample Images

5.1. BINARIZATION

Binarization is the process of simplifying data by dividing it into two distinct categories, typically represented as 0 and 1, or black and white. In image processing, it's like converting a grayscale image, which has many shades between black and white, into a purely black-and-white image. This is done by setting a threshold: pixels brighter than the threshold are turned white, and those darker are turned black.

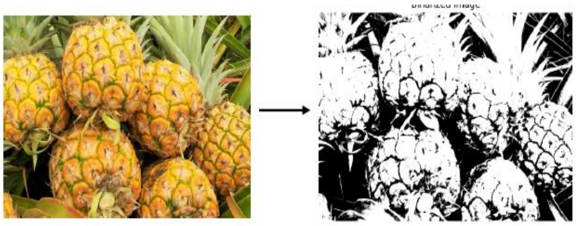


Fig 5.1.1. Binarization

5.2 EXTRACTED FEATURES

The various characteristics are analyzed to classify fruits accurately. Shape features, such as area, perimeter, and aspect ratio, help to determine the overall geometry of the fruit, with the area measuring its size, the perimeter indicating the boundary length, and the aspect ratio comparing the width to the height of the enclosing bounding box. Texture features, including contrast,

```
Shape Features:
  Area: 915507.0
  Perimeter: 32160.675999760628
  Aspect Ratio: 1.4403829416884246
  Bounding Box Width: 1655
  Bounding Box Height: 1149

Texture Features:
  Contrast: 192111514
  Energy: 1080832770
  Homogeneity: 548690.7334711701

Wavelet Features:
  LL Mean: 283.35000630119725
  LH Mean: 0.26393404746901916
  HL Mean: 0.6631590002100399
  HH Mean: -0.0034299516908212514
  LL Std: 111.40904260189978
  LH Std: 9.101049367140687
  HL Std: 9.861697560946729
  HH Std: 4.5245602493836055
```

energy, and homogeneity, provide insight into the fruit's surface patterns, where contrast measures brightness variation, energy indicates texture smoothness, and homogeneity reflects the uniformity of the texture. Additionally, wavelet features, obtained through wavelet transformations, break the image into frequency components at different scales, with LL capturing general brightness and smoothness, while LH, HL, and HH represent finer horizontal, vertical, and diagonal details that further characterize the fruit's texture. Together, these features form a comprehensive profile for fruit recognition.

Fig 5.2.1. Extracted Features

5.3. VALIDATION ACCURACY

```
[22] print("Validation Set Accuracy: {} %".format(training_history.history['val_accuracy'][-1]*100))

     Validation Set Accuracy: 91.16809368133545 %
```

Fig 5.3.1. Validation Accuracy

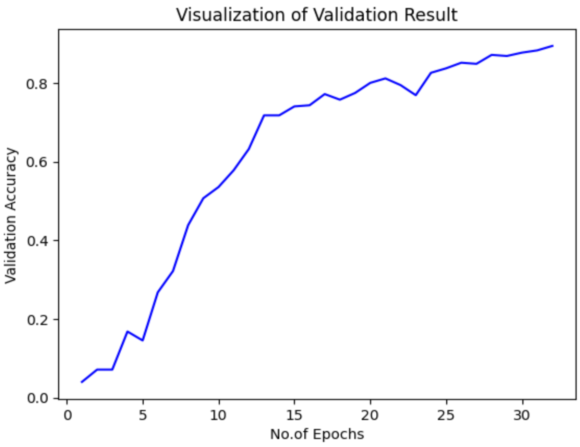
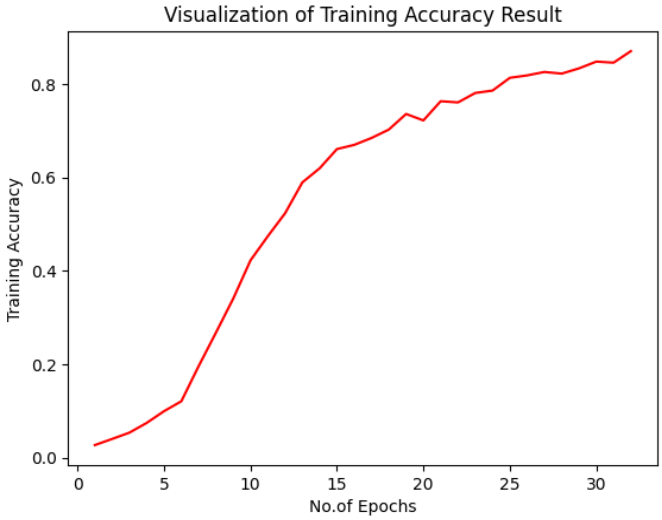


Fig 5.3.2 Validation Result

Fig 5.3.3. Training Accuracy Result



Our model achieved an impressive accuracy of 91.16% on the validation set, demonstrating its strong ability to classify fruit images accurately. This high level of accuracy means the model can reliably distinguish between various fruit classes, even in cases where differences might be subtle. It indicates that the model has learned the unique features of each fruit, such as shape, and texture, and can apply that knowledge to new, unseen images.

With this accuracy, the model shows real potential for practical use in applications like automated fruit sorting, quality control, or inventory management. The 91.16% accuracy not only reflects the model's effectiveness but also highlights its promise for real-world fruit recognition tasks where reliable and efficient classification is essential.

5.4. ACCURACY

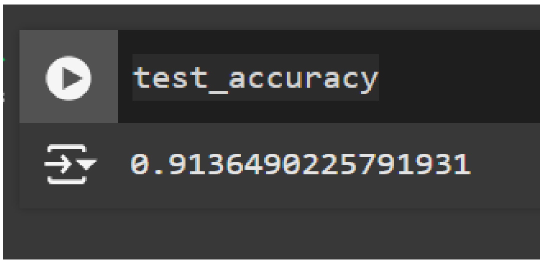


Fig 5.4.1 Test Accuracy

Our model achieved a test set accuracy of 91.36%, which is a strong indicator of its ability to classify fruit images effectively. This high level of accuracy shows that the model is not only performing well on the data it was trained on but also generalizing well to new, unseen data. In other words, it can accurately distinguish between different fruit classes when presented with entirely new images, making it highly reliable for practical applications.

5.5. IMAGE PREDICTION RESULTS

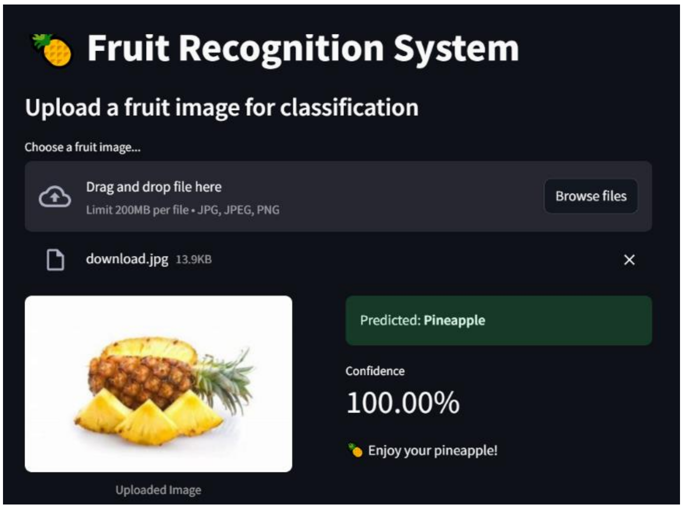


Fig 5.5.1 Prediction Results in Streamlit

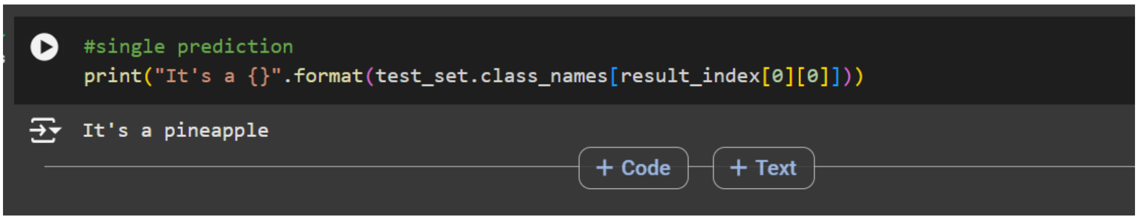


Fig 5.5.2 Prediction Results in Colab

In the final stage of image prediction, our model performs quite well, predicting fruit images with a high degree of accuracy. It successfully identifies most of the fruit classes, showing that it has learned to recognize key features such as shape and texture. However, while the results are promising, there are still a few areas where the model could be improved.

[6] CONCLUSION

In conclusion, the fruit recognition project successfully showcased how powerful machine learning algorithms can be in classifying different types of fruits based on visual features extracted from images. The model's test set accuracy of **91.36%** is a testament to its ability to generalize well, meaning it can accurately predict fruit types even with new, unseen data. The model's validation accuracy of **91.16%** reflects its strong generalization capability, indicating its consistent performance in predicting fruit types on previously unseen validation data. This strong performance demonstrates the potential of machine learning in automating fruit classification tasks. A key aspect of this success was the careful attention to data pre-processing, which played a vital role in preparing the images for analysis. By extracting meaningful features like shape, and texture, the model was able to distinguish between fruit classes with clarity. Moreover, the project explored machine learning algorithm like Convolutional Neural Networks (CNNs) to achieve high accuracy. By analysing the patterns and distributions of the fruit images, we were able to better understand which features were most important for classification.

7. REFERENCE

1. A. Majumdar , A. Bhattacharya, " A Comparative Study in Wavelets, Curvelets and Contourlets as Feature sets for Patterns Recognition", The International Arab journal of Information Technology, Vol 6, No.1, pp.47-51, January 2009.

2. D.Lee, J.Archibald and G.Xiong," Rapid Color Grading for Fruit Quality Evaluation Using Direct Color Mapping", IEEE Transactions On Automation Science And Engineering, Vol 8, No.2,pp.292-302, April 2011.

3. C. Lin , C. Su, H. Haung and K. fan , “ Colour Image Segmentation Using Relative Values of RGB in Various Illumination Circumstances”, International Journal of Computers, Issue 2, Vol-5,pp.252-261, 2011.

4. H. Kachariya,K. Vasaniya, S. Dhameliya, R. Savant, “A Review of Volume Estimation Techniques of Fruit”, International Journal of Engineering Research & Technology (IJERT) , Vol. 4 Issue 01 ,pp.217 221, January-2015.

5. D. Kakadiya, N. Shah, M. Patel, R. Shah, C. Kachariya, K. Sukhwani, “ Shape Extraction Methods for Fruits: Technical Review” , International Journal of Computer Applications(0975 8887) , Vol.111, No.1, pp.43-48, Feb.2015.