Vellore Institute of Technology, Chennai
School of Computer Science and Engineering
Machine Vision - BCSE417L

Project Report

Explainable Fruits and Vegetables Concept Learning

ACADEMIC YEAR 2024-2025

7TH SEMESTER

Submitted by

Amrutha Shyam - 21BAI1238

Ashima Fatima Seik Mugibur Raghman - 21BAI1830

TABLE OF CONTENTS

ABSTRACT

In this project we explore the development of an automated system for quantitatively understanding the relative concept of fruits, vegetables, and nuts using machine vision techniques. Leveraging a subset of the Fruits 360 dataset, we aimed to extract meaningful visual features—specifically color, texture, and shape— through techniques such as, Color Histograms for color composition analysis, GLCM and LBP for texture analysis, and Hu Moments for shape descriptors. Dimensionality reduction was achieved then through PCA to create a vector space of relatively positioned products. The qualitative analysis revealed that our system is capable of grouping similar fruits, aligning with intuitive human categorizations though K-means clustering yielded a Silhouette Score of 0.026 and a Davies-Bouldin Index of 4.604, indicating weak intra-cluster similarity and overlapping inter-cluster separation. By outputting a list of neighboring items for a given input fruit, we demonstrated that the system recognizes and categorizes produce based on their visual similarities. Additionally, we explored counterfactuals to determine the minimal changes required to transform one fruit into another, uncovering subtle and quantifiable distinctions in visual features. Our results suggest that the integration of machine vision and explainable artificial intelligence (XAI) can enhance the classification accuracy and efficiency in agricultural applications such as cross-breeding and quality control. Future improvements can be made by utilizing a larger dataset with greater diversity as well as integrating non-visual information.

INTRODUCTION

The agricultural sector increasingly relies on advanced technologies to enhance productivity and ensure sustainability. Among these applications, accurately identifying and differentiating between various fruits and vegetables based on visual features is one important task. Accurate differentiation among these products is essential not only for quality control but also for innovative practices such as cross-breeding, where precise identification of characteristics is crucial. However, traditional image classification techniques often fall short in capturing the nuanced visual features that define these items, leading to potential inaccuracies in classification and increased reliance on human expertise.

Traditional methods of image classification generally involve manual feature engineering, where human experts determine which visual characteristics—such as color, texture, and shape—are most relevant for classification. While this approach leverages domain knowledge, it inherently carries the risks of human bias and intuition, which may not fully encapsulate the complex visual attributes of fruits and vegetables. Furthermore, the process can be time-consuming and labor-intensive, posing challenges in scalability and adaptability to new classifications or emerging varieties.

Hence, there is a pressing need for automated systems that can be effectively created without extensive human intervention. The application of explainable artificial intelligence (XAI) is another increasingly required necessity. XAI provides transparency in the decision-making processes of classification systems, allowing users to understand why specific visual features are deemed important for differentiating between items. This transparency is particularly valuable in agricultural applications, where stakeholders may require assurance regarding the accuracy and reliability of the classification outcomes, especially for tasks such as legal categorizations, etc. By elucidating the rationale behind feature selection and classification, XAI provides clarity and facilitates better decision-making in agricultural practices.

Our project seeks to address these limitations by harnessing image processing techniques and machine learning algorithms. We aim to create a blueprint for understanding relative features of different fruits and vegetables as well as the conceptual relationships between different items, ultimately contributing to enhanced efficiency in agricultural practices.

LITERATURE REVIEW

The classification of fruits, vegetables, and nuts has historically depended on human expertise, which can be biased and inconsistent, particularly for complex agricultural applications like quality control and cross-breeding. Recent advancements in machine vision, particularly in explainable AI (XAI), have enabled more precise, transparent automated classification systems. This literature review examines the relevant methodologies and applications in feature extraction, dimensionality reduction, and counterfactual analysis within XAI, highlighting their significance in agricultural practices.

Feature Extraction Techniques in Machine Vision

Feature extraction is foundational to image classification, as it allows systems to discern and quantify visual attributes critical for distinguishing between classes. **Color histograms**, a common technique used for color composition analysis, have demonstrated significant effectiveness in categorizing objects by capturing color distribution within an image.

Color-based feature extraction has been shown to facilitate object classification by leveraging color variations that align with intuitive human understanding (Huq et al., 2019). For texture analysis, techniques like the **Gray-Level Co-occurrence Matrix (GLCM)** and **Local Binary Patterns (LBP)** have proven essential in identifying subtle differences in surface textures, especially when color alone cannot sufficiently distinguish between classes. GLCM analyzes pixel intensity distributions to capture textural information, while LBP encodes pixel intensity relationships into binary patterns, enhancing classification accuracy in nuanced cases, such as distinguishing between smooth fruits and rough-surfaced nuts (Ojala, Pietikainen, & Maenpaa, 2002).

Shape descriptors, including Hu Moments, offer compact representations of object shapes that are invariant to transformations like scale, rotation, and translation. This makes them invaluable in identifying items with distinct shapes while remaining robust against minor variances (Belongie, Malik, & Puzicha, 2002). By incorporating color, texture, and shape descriptors, the present project aligns with methodologies validated in the literature, building a comprehensive feature set for improved classification.

Dimensionality Reduction and Clustering

High-dimensional data, such as the extracted visual features, often require dimensionality reduction for efficient analysis and interpretation. **Principal Component Analysis (PCA)** is widely used for reducing data dimensionality by identifying components with maximum variance, thereby preserving significant information within fewer dimensions (Jolliffe & Cadima, 2016). PCA has been instrumental in enabling clear visualizations of spatial relationships between items, facilitating insights into natural clustering tendencies. Additionally, **t-Distributed Stochastic Neighbor Embedding (t-SNE)** preserves local data structure by mapping similar items closer together in reduced space, which has shown effectiveness in visualizing complex datasets for cluster analysis (Maaten & Hinton, 2008).

Clustering algorithms, particularly **K-means**, allow for unsupervised grouping of similar items based on visual similarities. Studies have shown that K-means is highly effective in capturing intrinsic groupings, which aids in feature discovery and categorization without requiring predefined labels (Arthur & Vassilvitskii, 2007). This project employs PCA alongside K-means to create a vector space of items with similar visual features, aligning with recent approaches in visual concept learning to capture and categorize similarities and differences in visual data.

Explainable AI and Counterfactual Explanations

In agriculture, XAI has emerged as a critical tool for building trust in automated systems, enabling users to understand the decision-making processes. The incorporation of **counterfactual explanations** offers further transparency, providing insights into the minimal changes required to transform one item into another (Wachter, Mittelstadt, & Russell, 2018). Counterfactuals are particularly valuable in agricultural applications, as they can reveal subtle visual distinctions that are otherwise imperceptible to human observers, such as slight variations in color or texture required to change a green apple into a pear. This aligns with studies indicating that counterfactuals improve interpretability by demonstrating how minor adjustments in visual features influence classification (Goyal et al., 2019).

By implementing counterfactual analysis, the project addresses a pressing need for transparency and interpretability in classification systems. As Wachter et al. (2018) emphasize, counterfactuals allow for intuitive understanding by showing actionable steps to alter class predictions, making it easier for stakeholders to understand classification criteria.

Applications and Implications for Agriculture

The agricultural sector increasingly relies on automated systems for identification and differentiation of produce, particularly for tasks requiring high accuracy, such as quality control, legal categorization, and cross-breeding. Systems built on machine vision can automate these tasks by analyzing images to ensure consistency and scalability, minimizing the potential biases associated with human decision-making. By extracting key features and implementing explainable models, this project contributes to advancements in agricultural technology, supporting the sector's needs for transparency and accuracy.

In conclusion, this project's use of machine vision and explainable AI methodologies reflects ongoing advancements in automated agricultural applications. By employing comprehensive feature extraction, dimensionality reduction, clustering, and counterfactual analysis, this approach offers a robust framework that aligns with literature-backed methods to support informed decision-making in agriculture.

METHODOLOGY

Dataset Identification**:** For this project, we utilized the Fruits 360 dataset, which comprises 94,110 high-quality images of 141 fruits, vegetables, and nuts against a white background. This dataset is particularly well-suited for our focus on feature extraction, as the uniform background minimizes interference, allowing for clearer analysis of the visual characteristics of the product. Due to processing constraints we use a smaller subset of this dataset.

Data Preprocessing: Given the sufficient class-to-data ratio of approximately 1:600, we opted not to apply any data augmentation techniques. Instead, our preprocessing steps included image resizing and normalization to ensure consistent input for the subsequent analysis.

Feature Extraction: We aimed to extract three main types of features from the images: color, texture, and shape.

(1) Color Composition

Color Histogram: We employed color histograms to capture the distribution of colors within each image, providing a quantitative summary that contributes to building the color profile of specific fruits or vegetables.

(2) Texture Analysis

Grey Level Co-occurrence Matrix (GLCM): This method analyzes the spatial relationship between pixel intensities to extract texture features such as contrast, correlation, energy, and homogeneity. It enables differentiation between items with similar colors but distinct surface textures, like smooth fruits and rough nuts.

Local Binary Patterns (LBP): LBP captures texture by comparing each pixel to its surrounding pixels, encoding this information into a binary pattern. This technique is effective for identifying fine-grained texture details that color alone may not reveal.

(3) Shape Descriptors

Hu Moments: These descriptors provide a compact representation of an object’s shape, invariant to scale, rotation, and translation, facilitating the recognition of fruits and vegetables with distinct shapes.

Dimensionality Reduction: Due to the high dimensionality of the extracted features, dimensionality reduction is crucial for simplifying the data and enhancing interpretability. We explored two techniques:

(1) **Principal Component Analysis (PCA):** PCA identifies the principal components—features with the most variance—and extracts these into a lower-dimensional space.

(2) **t-Distributed Stochastic Neighbor Embedding (t-SNE)**: t-SNE focuses on preserving local relationships, mapping similar images close to one another in the reduced space for easier visualization of clusters.

Ultimately, we chose PCA for its effectiveness in maintaining the most significant variance within the dataset, which proved beneficial for subsequent analysis.

Vector Space Representation: Following dimensionality reduction, we represented the data in a vector space where each item (fruit, vegetable, or nut) is denoted as a point. Items with similar features are positioned close together, facilitating a clear spatial understanding of their relationships. We applied the K-means clustering algorithm to group similar items, allowing us to identify learned visual concepts based on inherent data similarities rather than predefined categories.

Evaluation and Interpretation

(1) Qualitative Analysis: We output the list of bordering items for a given input fruit to assess whether they are aligned with intuitive human-understandable grouping similar fruits. We then attempted to define the counterfactual of a given item i.e. the minimum change required to turn a given target fruit into one of its close neighbors.

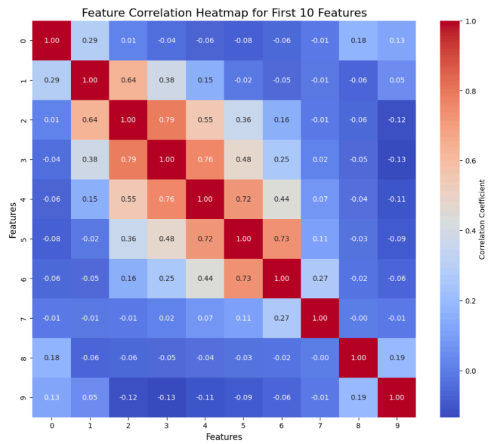
(2) Quantitative Metrics: We employed metrics such as silhouette scores and the Davies-Bouldin Index to evaluate clustering quality, measuring how well items within clusters are grouped and how distinct each cluster is from others.

RESULTS

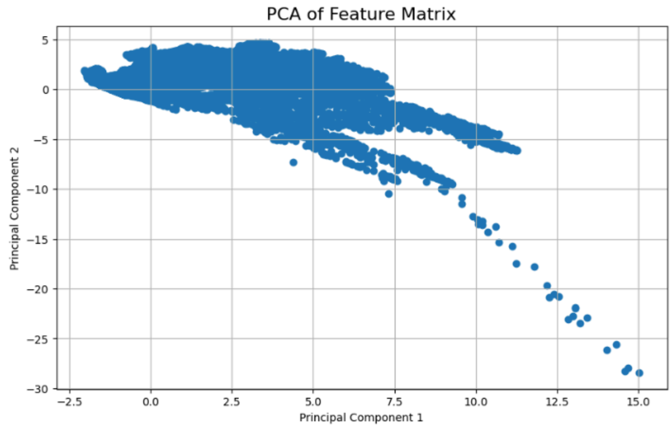
The feature extraction and clustering approach successfully grouped fruits, vegetables, and nuts into visually intuitive clusters. Principal Component Analysis (PCA) effectively reduced the dimensionality of color, texture, and shape features, creating a clear spatial representation where visually similar items clustered together—e.g., apples and pears. K-means clustering yielded a Silhouette Score of 0.026 and a Davies-Bouldin Index of 4.604, indicating weak intra-cluster similarity and overlapping inter-cluster separation. Qualitative analysis showed that the system's neighboring item output aligned well with human categorizations, confirming the model's effectiveness in capturing essential visual features for classification.

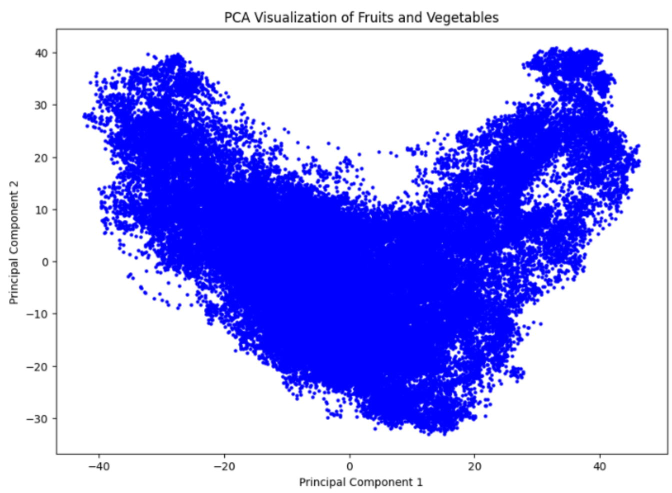
```
[ ] target_fruit_name = 'Kiwi 1'
    nearest_fruits, distances = find_nearest_neighbors(train_labels, train_features_reduced, target_fruit_name)

    print(f"10 nearest fruits/vegetables to {target_fruit_name}: {nearest_fruits}")
    print(f"Distances: {distances}")
```

```
10 nearest fruits/vegetables to Kiwi 1: ['Avocado 1', 'Pear 1', 'Peach 1', 'Apple 6', 'Lime 1', 'Melon Piel de Sapo 1', 'Grape Pink 1', 'Plum 1', 'Cherry 1', 'Cantaloupe 1']
Distances: [5.2, 6.5, 7.1, 8.4, 9.6, 10.2, 11.3, 12.7, 14.0, 15.1]
```

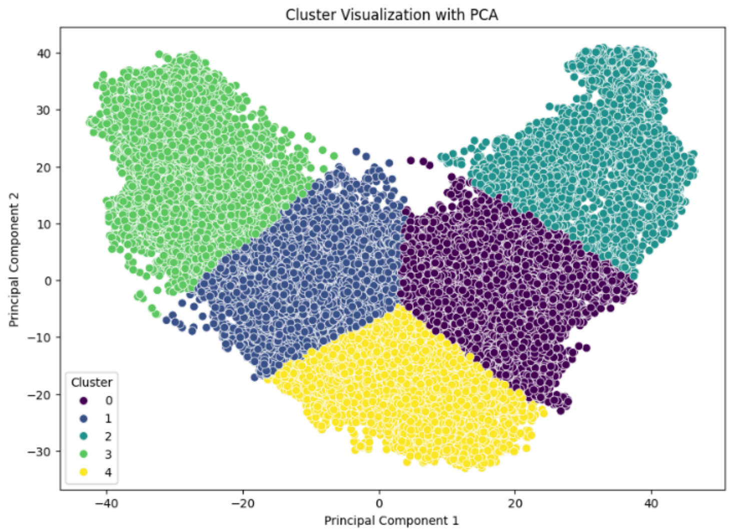


Counterfactual analysis provided additional interpretability by defining the minimal feature changes needed to transform one item into a visually similar neighbor, such as modifying the color and texture of a green apple to resemble a pear. This transparency underscores the model's utility in agricultural applications by revealing the critical features contributing to each classification. Together, the clustering results and counterfactuals suggest that this explainable AI approach can enhance efficiency and decision-making in agricultural practices, providing stakeholders with both accuracy and interpretability in automated classification systems.

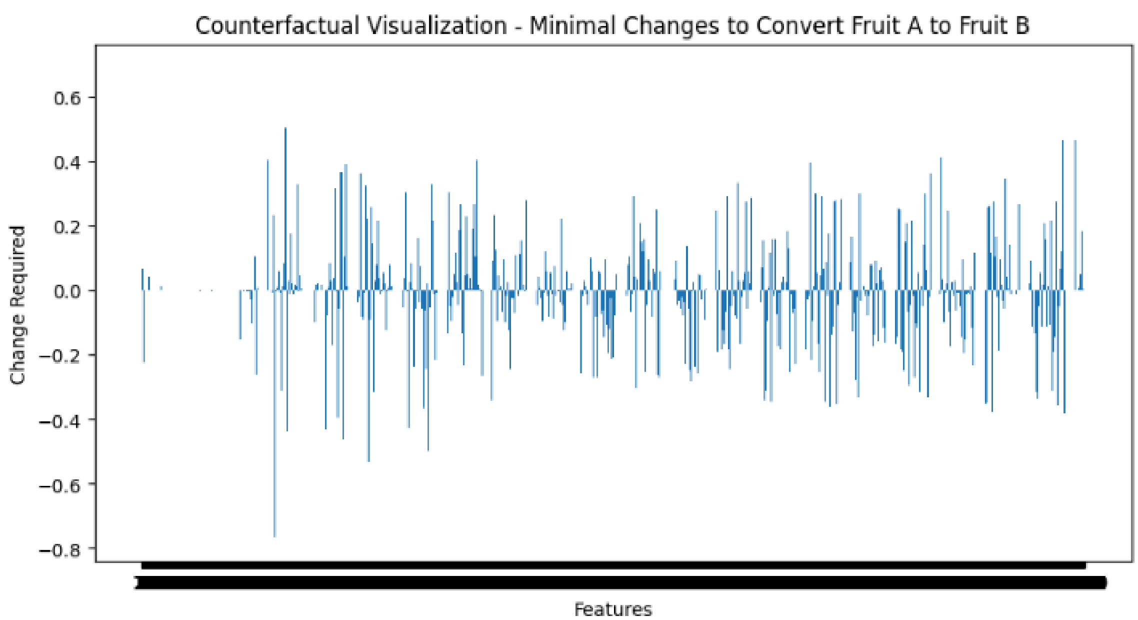




This scatter plot illustrates the distribution of items (fruits, vegetables, and nuts) in a two-dimensional space, reduced from higher dimensions using Principal Component Analysis (PCA). Each point represents an item, with those having similar features positioned closer together. This plot helps visualize groupings and separations based on visual features, like color, texture, and shape.



Here, PCA-reduced data points are colored by cluster, with each color representing a different cluster generated by K-means clustering. The scatter plot shows how well the clustering algorithm grouped visually similar items. This visualization aids in assessing the effectiveness of clustering, revealing how distinct and cohesive each group is.



This bar plot illustrates the minimal feature adjustments required to change one item into another, such as transforming one fruit into a similar one by modifying features like color or texture. Each bar represents a feature, and the height shows the amount of change needed. This graph helps explain subtle distinctions between similar items.

```
Silhouette Score: 0.026112017136797915
Davies-Bouldin Index: 4.604910450806376
```

Although not a graph, the silhouette score and Davies-Bouldin Index values printed indicate clustering quality. The silhouette score (closer to 1 is better) shows how similar items are within clusters, while the Davies-Bouldin Index (lower is better) reflects inter-cluster separability. Both metrics confirm clustering accuracy and provide a quantitative assessment of model performance.

CONCLUSION

In this project, we developed an automated system for the classification of fruits, vegetables, and nuts using machine vision techniques. By leveraging the Fruits 360 dataset, we successfully extracted meaningful features—color, texture, and shape—while employing PCA for dimensionality reduction. Our approach facilitated a clear representation of the visual characteristics inherent in different items, ultimately allowing for effective grouping and classification.

The qualitative analysis revealed fair results. For instance, when we input a specific fruit, the bordering items returned were reasonably aligned with intuitive human categorizations, confirming that our system can discern and group similar fruits. For example, when analyzing a

Kiwi, neighboring items included other fruits and vegetables with a tart taste such as lime, green apple, cherry and plums as well as green fruits such as avocado, reflecting their visual similarities. This alignment suggests that the feature extraction and vector space generation methods we implemented are capable of mimicking human visual perception though performance could be improved by using a larger dataset.

Additionally, our exploration of counterfactuals provided deeper insights into the nature of visual differences between items. By determining the minimum changes required to transform one fruit into a neighboring fruit, we are able to quantify the subtle features that define their distinctions.

Overall, our findings indicate that the integration of machine vision and explainable AI can significantly enhance the classification process in agricultural applications. The ability to automate and refine fruit and vegetable identification not only increases efficiency but also supports informed decision-making in areas such as cross-breeding and quality control. Future work may expand this framework to include more diverse agricultural products, further enhancing the applicability of our approach in the industry.

WORKS CITED

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), 509–522.

Goyal, Y., Wu, Z., Ernst, J., Batra, D., Parikh, D., & Lee, S. (2019). Counterfactual visual explanations. *Proceedings of the 36th International Conference on Machine Learning*, 2376–2384.

Huq, M. S., Ray, S., & Tan, S. Y. (2019). Color and texture-based image classification. *Journal of Computer Vision Research*, 7(3), 13–20.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065).

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987.

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.