Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

A Comparative Analysis of Color Channel-Based Feature Extraction using Machine Learning versus Deep Learning for Food Recognition

Yuita Arum Sari*[1], Dwi Cahya Astria Nugraha[2], Sigit Adinugroho[3]

[1,2,3]Faculty of Computer Science, Brawijaya University, Indonesia

Email: [1]yuita@ub.ac.id

Received : Jun 30, 2025; Revised : Aug 14, 2025; Accepted : Aug 15, 2025; Published : Aug 24, 2025

Abstract

Automated Dietary Assessment Accurate food recognition is a big challenge in computer vision which is critical for developing Automated Dietary assessment and health monitoring systems. The key question it answered was whether traditional machine learning with feature engineering by hand can beat modern deep learning approaches? In this Context, this study serves as a comparative analysis of these two paradigms. The baseline method worked by extracting texture (LBP,GLCM) and color information from different channels of five colors spaces (RGB, HSV, LAB, YUV,YCbCr) followed by feeding these features into multiple classifiers such as Nearest Neighbor(NN), Decision Tree and Naïve Bayes. These were then compared to deep learning models (MobileNet_v2, ResNet18, ResNet50, EfficientNet_B0). The best traditional one can reach an accuracy of 93.33%, using texture features extracted from the UV channel and classified with a NN. Nevertheless, the deep learning models consistently presented higher performance and MobileNet_v2 reached up to 94.9% accuracy without requiring manual feature selection. In this paper, we show that end-to-end deep learning models are more powerful and error robust for food recognition. These results highlight their promise for constructing more effective and scalable real-world applications with less need for intricate, domain-specific feature engineering.

Keywords : *Color Channel Based Feature Extraction, Color Space, Deep Learning, Machine Learning, Food recognition, Texture Extraction*

1. INTRODUCTION

Food recognition in healthcare has great potential, especially for improving medical services and personalized health information systems [1], [2], [3], [4], [5]. Research related to diabetes, obesity or hypertension are chronic diseases that can be easier in clinical nutrition monitoring by identifying food types and adjusting portions appropriately [6]. Hospital electronic medical record systems, for example, can improve tracking of patient nutritional intake, reducing errors in human data entry [7], [8]. In addition, food recognition systems can also be used to speed up food recording by integrating deep learning-driven food identification models. Machine learning models can achieve excellent accuracy in classifying food images [9] [10] [11] [12], as demonstrated by studies such as those conducted by Feng et al. [13].

At present, the proliferation of internet and mobile usage has become a pervasive phenomenon, accompanied by the emergence of novel habits among certain demographics. In contemporary society, individuals frequently employ photographic documentation to disseminate information, eschewing solely textual means in their efforts to portray the dynamic interactions that characterize their daily lives. For instance, the practice of documenting one's experiences through social media has become increasingly prevalent. This phenomenon involves the act of taking a photograph prior to partaking in a meal, which has the potential to disseminate information pertaining to the image captured. The advent of the internet has profoundly transformed the landscape of image storage and dissemination. The ease

 Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

with which images can now be accessed, stored, and shared has had a significant impact on society. As indicated by the available literature, the images in question have been demonstrated to represent latent textual descriptions that can be retrieved automatically [5].

Some researches employ this circumstances to develop technologies that can recognize and analyze food images automatically to help some patients according to their health, such as obesity [14]. Obesity becomes main concern since it can inherit some diseases, for instance heart disease, diabetes, and some cancers [15]. One way to tackle obesity issue is by controlling food intake using food logging for dietary program [16] [17]. Preceding activity, people tend to write down on a book of their daily meals and physical exercise in order to maintain their weight loss. In digitization era, Smartphone application is one devices that can facilitate food logging easily and monitoring dietary interactively through food images taken by camera that is embedded in smartphones [18]. Automatic dietary management is necessary to store daily meal habits to prevent health risk and the common dare for this. Food images needs to be inserted in wearable sensors when applying it in different surroundings of problems to further analysis [19].

Food image classification have been conducted to increase the accuracy and efficiency of automatic food consumption reporting systems[8] . Research in the domain of food image recognition is fraught with challenges, primarily due to the sheer diversity of foodstuffs and the inherent deformability of many of them. This category encompasses non-rigid objects, including porridge, pudding, and various types of soft food. The presence of color, shape, size, and texture in a given food is indicative of its similarity to other foods. The integration of these characteristics is regarded as a fundamental aspect of the classification of food images. [14].

Food recognition contains both color and texture which color represents valuable source of information [20] which can be seen indirectly and texture indicates the contour of food images to define clearly the unique feature of food [21]. Color features independent supports its higher ability to discriminate the naturally vivid foods that have distinct chromatic differences [22] [23] but in the case to communicate surface patterns or structural information texture features may be sufficient. However, using only color or texture may not always be sufficient to represent so food item, and therefore the accuracy of recognition will increased by considering both features, when two foods have similar color but different texture (bug), or one has same texture with another different color [24], [25]. Colour and texture extraction are depending on the process of segmentation [26]. So that, segmentation phase is considered to take the main area of food image [27]. The rest properties that is not indicated as food is ignored. Segmentation involved color space transformation [28] help to recognize food image in segmentation phase since the color of food can be distinguished easily [29].

The segmented image is then taken for its features and fed into the classifier. This paper compares various classifiers such as NN Euclidean, NN Manhattan, 3-NN Euclidean, 3-NN Manhattan, Naive Bayes (NB), J48 (DT), and SVM Polynomial Kernel. In addition to using machine learning, the deep learning approach is also compared to gain broader insights into how to recognize food types. While previous studies have successfully applied texture features like GLCM [30] or employed deep learning models for classification, a systematic analysis of how individual color channels (e.g., R, G, B, H, S, V) influence the performance of different feature extractors (LBP vs. GLCM) remains under-explored. Furthermore, a direct benchmark of this granular, traditional approach against a suite of modern deep learning models on a consistent dataset has not been thoroughly investigated. This gap leaves an open question regarding the trade-offs between complex, manual feature engineering and automated, end-to-end learning in the context of food recognition [31] [32] [33] [34].

Several studies have been proposed including methods that combine machine learning techniques with features from artificial neural networks (CNN) to ensure accurate food recognition despite varying image quality [35]. In addition, there is also research that creates models that can overcome the problem

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

of imbalanced food data with special techniques so that food features can be analyzed better [36]. In this research, a comparative study was conducted on the difference in the effect of colour channel-based feature extraction with traditional machine learning and deep learning processes.

Consequently, this research aims to carry out a systematic assessment of traditional machine learning approaches against state-of-the-art deep learning models for food recognition using detailed color channel and texture feature analysis. Our aim is to determine which method(s) are most reliable and consistent, while illustrating the interplay between interpretability and performance.

2. RESEARCH METHODOLOGY

2.1. Dataset

To observe the function of color channel-based segmentation in each color space, a dataset was created. A total of 19 food categories were documented through photographic evidence, meticulously captured by hand using a smartphone camera. The photographic process did not entail the use of any specific technique, such as precise angles or rotations, in the capture of the food images. The photographic documentation was carried out using an iPhone7 smartphone. The foods were meticulously arranged on a white plate, positioned precisely in the center. The employment of a white background has been demonstrated to facilitate the segmentation phase, thereby enabling the distinction of the primary object. In the segmentation step, the use of a white backdrop facilitates the identification of the primary item. This method is most effective in conditions of unrestricted lighting during the photographic capture process. In this scenario, the lighting conditions remain constant during image capture due to the arbitrary nature of the illumination changes.

A total of 376 food images were identified, which were subsequently categorized into 19 distinct classes, as illustrated in Figure 1. The images in question contain a single, unvarying image of food. There is no specific rotation when capturing foods. Nevertheless, the majority of the images were captured at a perpendicular angle, with an approximate distance of 30 centimeters from the uppermost point. Nevertheless, an imbalance in the number of images per class was observed, as illustrated in Figure 1. The results of this case study suggest that an imbalance may become a troublesome component because of its ability to affect the training model and classification techniques in general. However, this study showed that the best categorization results may be obtained by using complete raw data.

Figure 1. Example dataset with 20 labels

Figure 2 shows the distribution of the number of images in each class in the food dataset. There are 20 food classes represented by numerical labels. Each vertical bar represents the number of images in each class. From the graph, it can be seen that the data distribution is not completely balanced. Some classes, such as classes 001, 002, 015, 018, and 029 have the highest number of images, which is about 25 images per class. In contrast, classes such as 004, 011, and 032 have fewer images, below 15.

Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

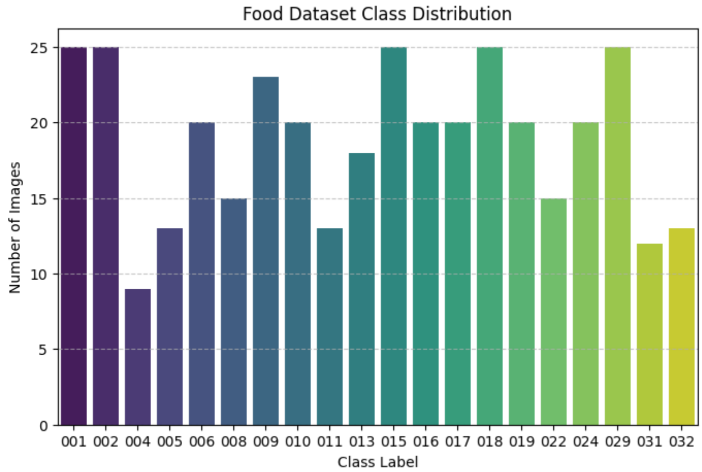

Figure 2. Food dataset class distribution

2.2. Food Recognition using Machine Learning and Deep Learning

2.2.1. Traditional Machine Learning

Figure 3, in general, the input is generally taken using a smartphone camera and then the segmentation process is carried out. In segmentation using color channel based is done to find out its effect on the color space (see Figure 4). Then from the preprocessing, the features are extracted using GLCM and LBP color and texture statistical features, then classification is carried out. Texture information about food boundaries and surface structure is extracted with features such as the Gray Level Co-occurrence Matrix (GLCM) [13]. Whether this is because two foods have the same colour, or even if they share different tones in their own right, each option of texture can help highlight your selections. Specifically, in our case we extract GLCM features with the pixel distance=1 of 4 angular directions (0°, 45°,90° and 135°); contrast, correlation, energy and homogeneity. We use 8 neighboring points with a radius of 1 for LBP to capture fine-grained local texture patterns. Texture basedIt depends on accurate segmentation methods and several texture-based approaches are vulnerable to changes in illumination conditions. Like porridge or pudding, which are not rigid foods that can change with texture.

Figure 4 describes the processing flow of a food image (sunny side up egg) using color-based segmentation and filtering techniques, specifically with the transformation to Lab color space (L*a*b*). The following is a step-by-step explanation of the process shown: Original image is the original image of food (eggs), then resizing is done to equalize the size of the image to make it easier to process. For the color transformation process, the image is converted from RGB to L*a*b* color space. This color space separates lightness (L*) and color information (a*, b*), making it more effective for color analysis and segmentation. This is also applied to other color spaces. Then for per channel processing in Fugure 4 for Per Channel Processing (L*, a*, b*)several steps are carried out,such as histogram equalization, average filtering, gaussian filtering, segmentation. Histogram Equalization: Increases the contrast of the channel to bring out the details. Average Filtering: Reduces noise by flattening the surrounding pixel values. Gaussian Filtering: Smoothing the image with a Gaussian filter, smoother than average. Segmentation: Separates the main object from the background based on the channel pixel values. The segmentation results from the three channels are combined to find the contours and edges of the object, resulting in an output image that shows only the important parts of the optimally selected food object.

One of the food image classification processes is carried out using the K-Nearest Neighbor (KNN) method, a distance-based classification algorithm that determines the label of test data based on the K nearest neighbors of the training data. KNN works on the principle that similar objects tend to be located close to each other in feature space. This study uses Euclidean and Manhattan Distance.

P-ISSN: 2723-3863
E-ISSN: 2723-3871
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

$$\boldsymbol{d}_{\text{Euclidean}}(\boldsymbol{x},\boldsymbol{y}) = \sqrt{\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{y}_i)^2} \qquad (1)$$

Where, $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$ **is** feature vector of the test food image**,** $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)$ **is** feature vector of the training food image**,** $\boldsymbol{n}$ **is** number of feature dimensions. **While** $\boldsymbol{x}_i, \boldsymbol{y}_i$ **is the** $\boldsymbol{i}$**-th** feature value of the respective images

$$\boldsymbol{d}_{\text{Manhattan}}(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{n}|\boldsymbol{x}_i - \boldsymbol{y}_i| \qquad (2)$$

Where x and y is feature vectors of the test and training food images, n is number of feature dimensions., and $|x_i - y_i|$is absolute difference of the $i-th$ feature value.

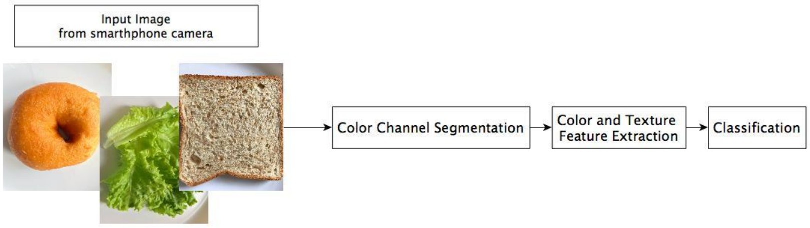

Figure 3. General phase of traditional machine learning

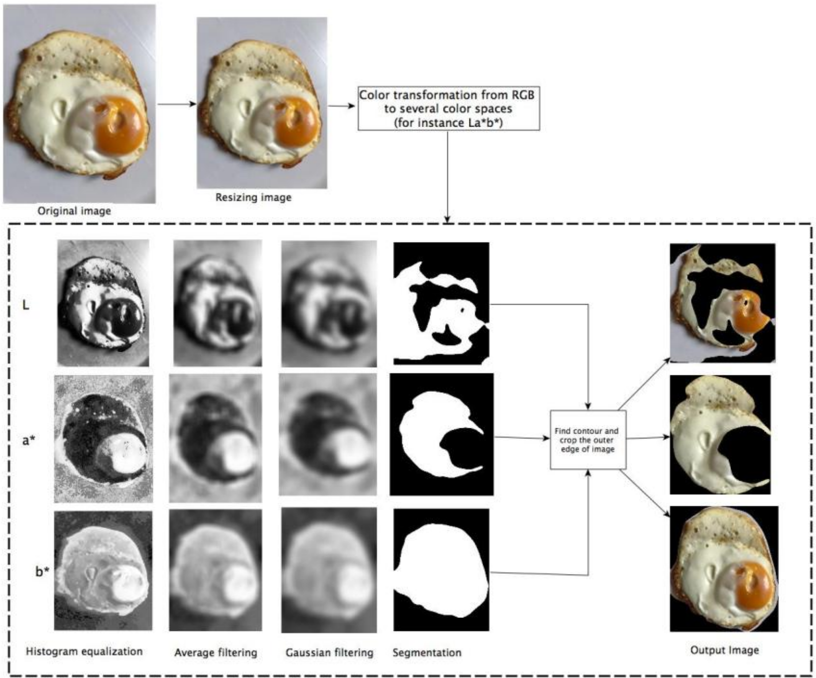

Figure 4. color channel based food image segmentation. From this figure, it can be shown that the b* channel in LAB plays a better role in food image segmentation.

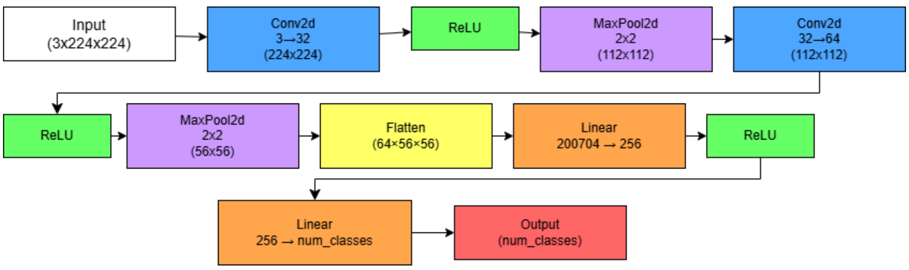

Figure 5. simple architecture of CNN for food recognition

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 4, August 2025, Page. 2484-2499
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

2.2.2. Deep Learning

Figure 5 depicts a simple Convolutional Neural Network (CNN) architecture where the Input Layer receives a color (RGB) image with a size of 3 x 224 x 224 pixels depicted as a white block on the leftmost side of the diagram. The Feature Extraction Layers consist of two convolution blocks, each followed by ReLU and MaxPooling which consists of Conv2d Layer 1 (Light Blue) which includes using 32 filters of size 3×3 with padding=1. The output is a 32 x 224 x 224 feature map. Then followed by ReLU Activation (Green) which is an activation function to add non-linearity. Then the next process is MaxPooling Layer 1 (Purple) by reducing the resolution to 32 x 112 x 112. Next is Conv2d Layer 2 (Light Blue) by using 64 filters of size 3×3 with padding = 1 with Output: feature map 64 x 112 x 112. Then using ReLU Activation (Green) and MaxPooling Layer 2 (Purple) which produces a 64 x 56 x 56 feature map. Then to convert the 3D tensor into a 1D vector of 200,704 elements (64 × 56 × 56), the Flatten Layer (Yellow) is performed which marks the transition from spatial data to vectors. Then the last is Fully Connected Layers (Classifier which is marked with orange blocks, namely Linear Layer 1, namely Input: 200,704 neurons, in which output: 256 neurons, then apply ReLU Activation (Green). In Linear Layer 2, the output is the number of classes (num_classes) according to the dataset. Finally, the Output Layer (Red) produces a prediction score for each class.

Pre-trained weight from the ImageNet dataset was employed to train these deep learning models using transfer learning. The final layer for classification was changed to 19 target classes for our dataset. The model was trained using the Adam optimizer for 50 epochs with a learning rate of 0.001 and batch size of 32. Within each epoch, setting the model to training mode and transferring both input images and labels to GPU(or CPU) device and running through forward pass where features were extracted. The cross-entropy loss was then backpropagated to update the model parameters using gradient descent. This would also keep a track of loss and accuracy among batches and how well the models were learning over time.

2.2.3. Evaluation Matrices

Accuracy is the most intuitive and the most common evaluation metric we can use in classification problems, that is used to know how well our model has predicted. The idea is that this metric gives a more general view of the model correctly predicting something.

The accuracy is: the number of correct founded objects are divided by all objects in the data set. It indicates how accurate the model predicts, only to the positive class and even in negative classes. We assess accuracy as stated in Equation (1) by comparing these correctly populated squares to the total number of predictions made at all. In mathematical term, accuracy is the ratio of total number of True Positives (TP) and True Negatives (TN) over total instances i.e. TP + TN + False Positives(FP)+ False Negatives(FN).

$$\text{ACCURACY} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \times \mathbf{100\%} \qquad (3)$$

3. RESULT

3.1. Traditional Machine Learning

The feature extraction results in 36 colour channel features, 26 LBP features, and 60 GLCM features. Figure 6 shows the results of how accurate each feature is. The Euclidean NN and Manhattan NN are usually the most accurate in most color spaces. They often have accuracy values above 91%. Both methods are reliable and the best choice for this data. The SVM Polynomial Kernel method also performed well, with accuracy ranging from 90% to 92%. However, the 3-NN with Manhattan distance and 3-NN with Euclidean distance methods performed worse, with an average accuracy of 86% to 88%. This suggests that using more neighbors does not necessarily improve performance. However, the J48

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 4, August 2025, Page. 2484-2499
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

(DT) and Naive Bayes (NB) methods were consistently less accurate, with accuracy ranging from 78% to 84%. The two best ways to sort this dataset are Manhattan NN and Euclidian NN. Based on this result, we elaborate on experiment in detail of feature type accuracy in each color channel using NN with Euclidean distance as stated in Table 1, where the 'All Feature' column shows the results when all feature types (Color, LBP, and GLCM) are combined into one feature vector.

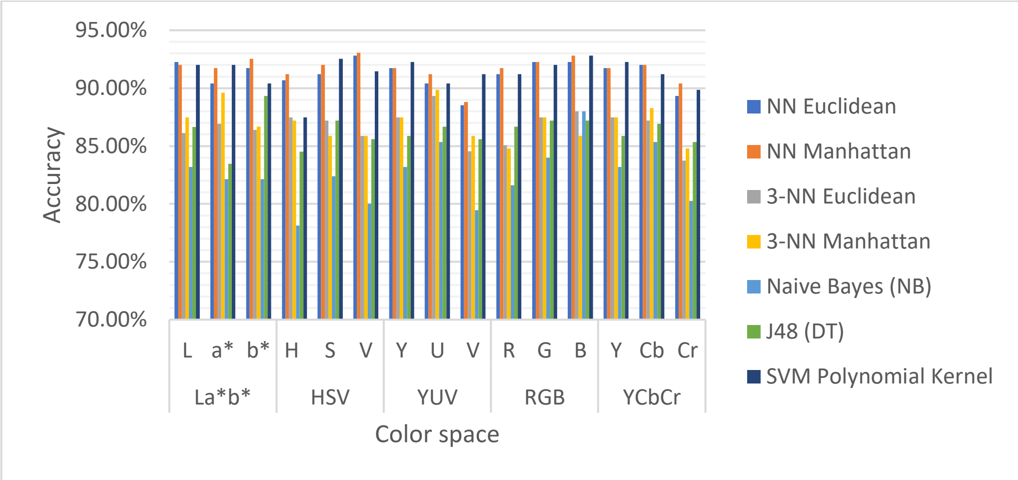

Figure 6. Accuracy in all features

Table 1 presents the classification performance across different color channels and feature types employing NN Euclidean Distance, with the peak accuracy of 93.07% reached on the V (Value) channel using an integration of all features. The B_RGB channel also shows an advantage, especially in texture characteristics, achieving an accuracy of 93.33%. The primary advantage is the dependability of the texture feature, especially in the RGB channel, which often achieves accuracy greater than 92%, showcasing the effectiveness of the textural feature in distinguishing image types. Despite this, the V_YUV and Cr channels show significant shortcomings, attaining the lowest accuracy rates of 89.60% and 89.33% for both overall and specific features, indicating their limited sensitivity to detecting changes in image patterns. Additionally, channels U and V_YUV show varying degrees of feature accuracy, with a significant difference observed between color and texture features. This implies that the color data in these channels is of reduced importance when texture features are not present. This result confirms the importance of selecting the right color channel and using texture features to achieve optimal accuracy, and shows that not all color channels are equal in their contribution to classification performance. The conclusions related to the feature comparison experiment are in Table 2.

Table 2 indicates that the texture feature consistently yields the highest accuracy in over 80% of the channels, particularly in RGB and LAB color spaces. The LBP feature demonstrates consistent performance, particularly in the Y_YUV and HSV color spaces. At the same time, GLCM generally proves less effective than LBP, particularly in chrominance channels like Cb and Cr. The Color feature showed strong performance in the V (HSV) and G_RGB channels, but was less effective in the Cr and U channels, indicating its limitations in gathering distinguishing information.

Table 3 presents an overview of the segmentation outcomes according to color space, channel, optimal feature type, and classification approach that produces the greatest accuracy. Texture feature clearly stands out as the most effective attribute across nearly all color channels, leading to significant accuracy in various color spaces, such as the B_RGB channel achieving the top accuracy of 93.33% with Manhattan NN, and the V (HSV) channel at 93.07% in Manhattan NN when all features are employed simultaneously.

Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

Table 1. Classification accuracy per feature type and color channel in traditional machine learning using NN with Euclidean Distance

Color Channel	All Feature(%)	Color(%)	LBP(%)	GLCM(%)	Texture(%)
L	**92.27**	87.20	91.47	89.87	92.00
A	**92.00**	89.93	91.47	89.87	**92.00**
B	**91.73**	88.53	91.47	90.13	**91.73**
H	89.60	88.53	92.53	89.87	**92.80**
S	90.40	90.13	92.53	88.27	**92.80**
V	**93.07**	90.40	92.53	92.53	92.80
Y_YUV	91.73	87.47	91.20	90.93	**92.00**
U	**91.20**	83.73	90.40	88.53	**91.20**
V_YUV	**89.60**	73.20	88.53	86.67	**89.60**
R_RGB	91.20	87.20	91.73	88.27	**92.00**
G_RGB	92.27	87.47	91.20	92.27	92.53
B_RGB	92.27	90.13	92.80	91.47	**93.33**
Y_CbCr	91.73	87.47	91.20	90.93	**92.00**
Cb	**92.00**	85.87	91.20	85.87	91.47
Cr	89.33	81.33	89.07	87.47	**90.13**

Table 2. Summary of all features

Feature	Summary
Color	Highly contributes in HSV (V), G_RGB, but weak in Cr, U
LBP	Consistently gives high results, especially in Y_YUV and HSV
GLCM	Often does not exceed LBP, sometimes weaker in chroma segment
Texture	Provides best results in >80% of segments, especially in RGB and LAB

Table 3. Color space segmentation summary

Color Space	Channel	Best Feature	Best Method	Top Accuracy (%)
LAB	L	Texture	NN Euclidean	92.27
LAB	A	LBP	NN Euclidean	91.73
LAB	B	Texture	NN Euclidean	91.73
HSV	H	Texture	NN Euclidean	89.6
HSV	S	Texture	NN Manhattan	90.4
HSV	V	All Feature	NN Manhattan	93.07
YUV	Y	Texture	NN Euclidean	92
YUV	U	Texture	NN Euclidean	91.2
YUV	V	Texture	NN Euclidean	89.6
RGB	R	Texture	NN Euclidean	92
RGB	G	Texture	NN Euclidean	92.53
RGB	B	Texture	NN Manhattan	**93.33**
YCbCr	Y	Texture	NN Euclidean	92
YCbCr	Cb	Texture	NN Euclidean	91.47
YCbCr	Cr	Texture	NN Euclidean	90.13

This result is consistent with previous studies which show that texture characteristics are very important for improving classification ability. In addition, the Euclidean NN and Manhattan NN

Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

methods also consistently produced the highest accuracy. Certain colour channels, including H (HSV) and V (YUV), showed reduced accuracy (below 90%). This indicates a lack of informativeness without additional colour features. Therefore, choosing the right colour channels and features has a great influence on obtaining the best accuracy.

3.2. Deep Learning

Table 4. Mean Accuracy of Each Model

Model	Mean Accuracy
MobileNet_v2	**0.949438596**
ResNet18	0.94677193
ResNet50	0.944105263
EfficientNet_B0	0.941438596

Table 4 displays the average accuracy of the four deep learning models assessed: MobileNet v2, ResNet18, ResNet50, and EfficientNet_B0. The findings show that MobileNet v2 has the highest performance, achieving an average accuracy of 94.94%. ResNet18 and ResNet50 ranked second and third, achieving 94.68% and 94.41% respectively. With a mean accuracy of 94.14%, EfficientNet_B0 shows the lowest average accuracy of the four models. This difference in accuracy suggests that simpler models, such as MobileNet v2, can achieve equivalent or even better performance than more complex models. This is due to the ability of MobileNet v2 to reduce overfitting while improving computational efficiency. Deep Learning consistently outperforms with an average accuracy of ~94%–95%, surpassing all traditional machine learning results, including the best ones.

3.3. Traditional Machine Learning vs Deep Learning in Food Recognition

From the overall findings, the greatest accuracy in the conventional machine learning method is obtained in the RGB color space, particularly in the B channel. In this implementation, the prevalent approach involves utilizing texture features and neural networks with Manhattan distance, achieving 93.33%. However, it is noteworthy that deep learning models, such as MobileNet_v2, ResNet18, and ResNet50, have been shown to consistently achieve accuracies of 94%, with MobileNet_v2 demonstrating a notable 94.9% accuracy. This suggests that while conventional machine learning approaches can attain high accuracy on individual channels, their efficacy is constrained and contingent on the features or methods employed, as well as the selection of channel combinations. In contrast, deep learning demonstrates consistently high performance overall, without reliance on color channels or manually designed features.

Out of the 14 color channels, 13 indicate that texture is the most effective feature. This indicates that texture retains the most unique information pertinent to food image segmentation classification. Nevertheless, the highest outcomes of conventional machine learning still do not reach the average precision when compared to deep learning. Deep learning naturally comprehends the representation of texture, edge, spatial pattern, and color simultaneously, without requiring the explicit extraction of LBP or GLCM features. The benefits of deep learning go beyond just one type of visual data, including its ability to understand complex and contextually rich representations.

RGB–B channel gives the best result in traditional machine learning with an accuracy value of 93.33%. This shows that color channel B contains dominant contrast and texture, or reflects different important elements between food image classes. However, DL does not show any dependence on a particular color channel: CNN models work on full RGB combinations (or other augmentations), and have the ability to automatically weigh which channels are important, through an end-to-end learning

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

process. Traditional machine learning requires explicit experiments for each channel, while deep learning abstracts it in its internal structure (e.g. convolutional filters learn color and texture features simultaneously).

The selection of a combination of methods and features is crucial in traditional machine learning where users need to try many configurations (channels, features, methods) to find the best results. While on the deep learning side, models such as MobileNet, ResNet, EfficientNet only need to prepare the dataset and architecture, then learn the most optimal features and representations themselves. The performance of traditional machine learning fluctuates across color channels and methods. Color channels such as Cr and Cb (in YCbCr) experience performance degradation (90.13% and 91.47%) even when using the best texture features. Meanwhile, all deep learning models show performance stability above 94%, regardless of input channels because they work on full images. This shows that deep learning has a much higher generalization capacity, compared to traditional machine learning which is prone to local overfitting on certain channel-feature-method combinations. Table 4 shows the comparison of the traditional machine learning and deep learning based on this paper experiment.

Table 5. Comparison between traditional machine learning and deep learning approaches based on this paper experiment.

Criteria	Traditional machine learning	Deep Learning
Maximum Accuracy	93.33% (RGB-B + Texture + NN using Manhattan distance)	94.9% (MobileNet_v2) ; the **highest score**
Consistency	Fluctuates across channels and methods	Consistently high across all deep learning models
Feature Dependence	High since it requires feature and color channel exploration	Learns automatically from data
Feature Dominance	Texture dominates in almost all channels	No single dominant feature – end-to-end abstraction
Interpretability	Can be explained per feature and create deep analysis	Using complex models (black-box)
Exploration Efficiency	Manual experiments required	One training process can generate many insights
Adaptability	Weak with new data if dominant channel changes	Strong due to broader abstraction learning

4. DISCUSSIONS

A comparative overview of the various food picture classification approaches used between 2015 and 2025 is provided in Table 6, with an emphasis on feature extraction methods such as GLCM, LTP, and Convolutional Neural Network (CNN). With respect to conventional techniques, texture-oriented methods such as GLCM and LTP continue to be widely utilized, evidenced by the 2018 research by Wint Myat Thu et al. [37], which employed GLCM and LTP alongside SVM classification, resulting in an accuracy of 93%. However, other studies using GLCM on other datasets, such as Gade and Vyavahare (2018) [38], found accuracies of about 70-75%, indicating that the quality and characteristics of the used data are crucial for the effectiveness of texturing approaches.

On the other hand, convolutional neural network (CNN)-based methods have become more and more advantageous in recent years. Feng et al. (2023) found that fine-grained CNN approaches could categorize Chinese food with an accuracy of 93.5% [13]. However, not all CNN implementations perform optimally. A previous investigation by Matsuda et al. (2015) employing covariance pooling of CNN feature maps attained a mere 58.65% accuracy [39]. This finding indicates that CNN methods necessitate the implementation of suitable architectures and training configurations to ensure optimal performance. Another study, published in 2023, sought to make a comparison between CNN and GLCM features. The findings indicated that CNN attained 67%, while GLCM achieved 69%. This suggests that, under specific circumstances, conventional media outlets can maintain a competitive edge against CNN [40].

Table 6. Comparison results of previous research of this research

No	Title	Year	Method(s)	Features Used	Accuracy or Result
1	Intelligent classification models for food products basis on morphological, colour and texture features [41]	2017	ML (Intelligent classification models): Multilayer Perceptron (MLP), Support Vector Machines (SVM), Random Forest (RF), Simple Logistic (SLOG) and Sequential Minimal Optimization (SMO)	Morphological, color, texture	80% to 96% for the training
2	GLCM and LTP Based Classification of Food Types [37]	2018	GLCM, LTP + classifiers (SVM)	Texture (GLCM, LTP)	93%
3	Feature Extraction using GLCM for Dietary Assessment Application [38]	2018	GLCM + classifier (not specified)	Texture (GLCM)	70–75% (from text)
4	A fine-grained recognition technique for identifying Chinese food images [13]	2023	CNN + fine-grained recognition	Deep CNN features	85.4% and 93.5%
5	Selection of Food Identification System Features Using Convolutional Neural Network (CNN) Method [40]	2023	CNN	Deep CNN features	67% of CNN and 69% of GLCM
6	Food Image Recognition Using Covariance of Convolutional Layer Feature Maps [39]	2015	CNN (Covariance pooling)	Deep CNN features (covariance maps)	58.65%
7	**Ours**	**2025**	**NN Euclidean, NN Manhattan 3-NN Euclidean, 3-NN Manhattan, Naive Bayes (NB), J48 (DT), SVM Polynomial Kernel,CNN**	**Texture (combined GLCM and LBP), CNN features**	**GLCM and LBP with Machile learning is 93.33%, MobileNet_v2 is 94.94%**

This paper's research shows the comparison between of CNN techniques and traditional machine learning approaches, including Euclidean NN, Manhattan, and Naive Bayes, along with the application

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

of joint GLCM and LBP features. The outcomes achieved are remarkable, with the GLCM + LBP + Manhattan NN approach attaining a competitive accuracy of 93.33%, and the MobileNet_v2 model showing an accuracy of 94.94%. This finding serves to reinforce the prevailing trend in the field, which suggests that a combination of modern and traditional approaches can result in high performance in the domain of food image classification, contingent upon the judicious selection of features and methods that align with the distinctive characteristics inherent in the specific dataset under consideration.

The better performance of MobileNet_v2 indicates that deep learning models can generalize more effectively through hierarchical feature representations, as opposed to the model being able to learn color-texture features in a hand-crafted manner. The subtle gap between MobileNet_v2 and ResNet18 may be because of the depthwise separable convolutions that MobileNets uses, helping against overfitting in somewhat smaller datasets.

Interestingly, a common trend in our internal results was the high accuracy of texture-based features extracted from the B (Blue) channel in RGB color space for many traditional methods. Since shadows and illumination variations might affect the skull color channel, this could be due to lighting conditions of the dataset and therefore intrinsic contrast between food items and background that makes blue channel less susceptible to variations. Because the blue color channel is generally less intense in food items with naturally warm tones (reds and yellows), both GLCM and LBP are better able to capture more discriminative structural properties (such as edges, textures), without being primarily concerned with capturing universal intensity signals.

This is not a large margin even though this translates to the highest accuracy MobileNet_v2 (94.94%) outperformed the best traditional method, GLCM + LBP + Manhattan NN (93.33%), by about 1.6%. A more practical trade-off in real applications is then between accuracy and the computational cost. Traditional methods are less accurate but can provide speed-based inferences with high interpretability and low memory consumption. This implies that for resource-starved environments like embedded systems in hospitals or dietary apps on the mobile, old well-tuned methods may still be a good choice without much of a blow to classification performance. Together, these results suggest that a hybrid approach using color-texture features for rapid and interpretable classification, and deep learning in high-stakes or ambiguous cases may optimally balance efficiency and accuracy.

The importance of the work lies in three aspects: methodological, as it presents a consistent benchmarking framework for performance gains (dropout) derived from feature engineering–entailing manual selection and tuning of handcrafted features–versus deep learning performance, which can be utilized by other researchers on similar computer vision tasks; practical, indicating that deep learning models are likely preferred to optimize accuracy through MobileNet_v2 while traditional methods optimized for either high resource utilization or interpretability concerns such as clinical validation remain relevant; and scientific, demonstrating empirical estimates of how imputations around black-box models and interpretable methodologies fare against each other in food recognition with potential realizations to ongoing topics within the field rather than relying on subjective perspectives..

5. CONCLUSION

This study proposes a technique for comparative analysis of traditional machine learning and deep learning methods. In traditional machine learning itself, what is done is more complex related to the influence of feature extraction caused by segmentation, as well as from the selection of classifiers. It turns out that color space and color channel selection have an effect and texture features have an effect on food image classification. Based on the experiments that have been carried out, deep learning methods, especially CNN such as MobileNet_v2, have succeeded in achieving the highest accuracy of up to 94.94% in food image classification, showing significant advantages over traditional methods based on texture features such as GLCM or LBP which generally range from 70–93.33%. The advantage

Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

of CNN lies in its ability to perform end-to-end feature extraction without the need for a manual feature selection process, resulting in consistent performance across various data schemes. However, traditional methods have important advantages in terms of interpretability, computational efficiency, and reliance on relatively small datasets, where methods such as Euclidean NN or Manhattan NN are still able to record competitive accuracy in the range of 93.33% when combining texture features (GLCM + LBP) and optimal color channel selection (eg B in RGB).

Although CNN has proven to be superior in accuracy, significant weaknesses lie in the need for large training data, longer training time, and its "black-box" nature, making it difficult to interpret. Therefore, the direction of future research development needs to be focused on the integration of a hybrid approach that combines the advantages of deep learning in feature abstraction with traditional methods that are easier to interpret, while exploring explainable AI methods so that classification results can be explained transparently. In addition, further research can be directed to testing the robustness of the model on more varied real-world data, as well as minimizing the model's dependence on certain color channels or dominant features so that the food image classification system becomes more adaptive and practical for implementation on various application platforms, such as dietary assessment as one of the systems that plays an important role in hospital settings.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

The author would like to express sincere gratitude to Yudi Arimba Wani from the Faculty of Health Sciences, Universitas Brawijaya, for his valuable collaboration throughout this research. Special thanks are also extended to the Faculty of Computer Science, Universitas Brawijaya for providing funding support through the "FILKOM Berkarya" research grant..

REFERENCES

[1] S. Han, J. Xiao, and S. Lee, "From Kitchen to Clinic: Integrating Food as a Vital Component of Healthcare," *Future Postharvest Food*, vol. 2, no. 2, pp. 139–146, Jun. 2025, doi: 10.1002/fpf2.70005.

[2] P. Mishra and G. Singh, "Internet of Medical Things Healthcare for Sustainable Smart Cities: Current Status and Future Prospects," *Appl. Sci.*, vol. 13, no. 15, p. 8869, Aug. 2023, doi: 10.3390/app13158869.

[3] J. Bassaganya-Riera *et al.*, "Goals in Nutrition Science 2020-2025," *Front. Nutr.*, vol. 7, p. 606378, Feb. 2021, doi: 10.3389/fnut.2020.606378.

[4] S. M. Donovan *et al.*, "Personalized nutrition: perspectives on challenges, opportunities, and guiding principles for data use and fusion," *Crit. Rev. Food Sci. Nutr.*, pp. 1–18, Feb. 2025, doi: 10.1080/10408398.2025.2461237.

[5] G. Prasad, M. Padhiary, A. Hoque, and K. Kumar, "AI-Driven Personalized Nutrition Apps and Platforms for Enhanced Diet and Wellness:," in *Food in the Metaverse and Web 3.0 Era*, A. M. Alhussaini Hamad and R. Soni, Eds., IGI Global, 2025, pp. 125–158. doi: 10.4018/979-8-3693-9025-2.ch006.

[6] L.-I. Coman, M. Ianculescu, E.-A. Paraschiv, A. Alexandru, and I.-A. Bădărău, "Smart Solutions for Diet-Related Disease Management: Connected Care, Remote Health Monitoring Systems, and Integrated Insights for Advanced Evaluation," *Appl. Sci.*, vol. 14, no. 6, p. 2351, Mar. 2024, doi: 10.3390/app14062351.

Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

[7] A. Abeltino *et al.*, “Digital applications for diet monitoring, planning, and precision nutrition for citizens and professionals: a state of the art,” *Nutr. Rev.*, vol. 83, no. 2, pp. e574–e601, Feb. 2025, doi: 10.1093/nutrit/nuae035.
[8] P. Khatiwada, B. Yang, J.-C. Lin, and B. Blobel, “Patient-Generated Health Data (PGHD): Understanding, Requirements, Challenges, and Existing Techniques for Data Security and Privacy,” *J. Pers. Med.*, vol. 14, no. 3, p. 282, Mar. 2024, doi: 10.3390/jpm14030282.
[9] H. S. Nogay, N. H. Nogay, and H. Adeli, “Image-based food groups and portion prediction by using deep learning,” *J. Food Sci.*, vol. 90, no. 3, p. e70116, Mar. 2025, doi: 10.1111/1750-3841.70116.
[10] S.-T. Cheng, Y.-J. Lyu, and C. Teng, “Image-Based Nutritional Advisory System: Employing Multimodal Deep Learning for Food Classification and Nutritional Analysis,” *Appl. Sci.*, vol. 15, no. 9, p. 4911, Apr. 2025, doi: 10.3390/app15094911.
[11] Y. Feng, Y. Wang, X. Wang, J. Bi, Z. Xiao, and Y. Luo, “Large-scale image classification and nutrient estimation for Chinese dishes,” *J. Agric. Food Res.*, vol. 19, p. 101733, Mar. 2025, doi: 10.1016/j.jafr.2025.101733.
[12] U. Chanthanuwatkun, A. Jitpattanakul, and N. Hnoohom, “Image Classification and Segmentation for Estimating Calories of Food,” in *2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, Nan, Thailand: IEEE, Jan. 2025, pp. 525–530. doi: 10.1109/ECTIDAMTNCON64748.2025.10962099.
[13] S. Feng, Y. Wang, J. Gong, X. Li, and S. Li, “A fine-grained recognition technique for identifying Chinese food images,” *Heliyon*, vol. 9, no. 11, p. e21565, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21565.
[14] H. He, F. Kong, and J. Tan, “DietCam: Multiview Food Recognition Using a Multikernel SVM,” *IEEE J. Biomed. Health Inform.*, vol. 20, no. 3, pp. 848–855, May 2016, doi: 10.1109/JBHI.2015.2419251.
[15] P. McAllister, H. Zheng, R. Bond, and A. Moorhead, “Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets,” *Comput. Biol. Med.*, vol. 95, pp. 217–233, Apr. 2018, doi: 10.1016/j.compbiomed.2018.02.008.
[16] A. J. Moshfegh, D. G. Rhodes, and C. L. Martin, “National Food Intake Assessment: Technologies to Advance Traditional Methods,” *Annu. Rev. Nutr.*, vol. 42, no. 1, pp. 401–422, Aug. 2022, doi: 10.1146/annurev-nutr-062320-110636.
[17] N. Shinozaki and K. Murakami, “Evaluation of the Ability of Diet-Tracking Mobile Applications to Estimate Energy and Nutrient Intake in Japan,” *Nutrients*, vol. 12, no. 11, p. 3327, Oct. 2020, doi: 10.3390/nu12113327.
[18] L. Lin, J. He, F. Zhu, E. J. Delp, and H. A. Eicher-Miller, “Integration of USDA Food Classification System and Food Composition Database for Image-Based Dietary Assessment among Individuals Using Insulin,” *Nutrients*, vol. 15, no. 14, p. 3183, Jul. 2023, doi: 10.3390/nu15143183.
[19] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, “Food image recognition using very deep convolutional networks,” in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, 2016, pp. 41–49.
[20] V. B. Raju, M. H. Imtiaz, and E. Sazonov, “Food Image Segmentation Using Multi-Modal Imaging Sensors with Color and Thermal Data,” *Sensors*, vol. 23, no. 2, p. 560, Jan. 2023, doi: 10.3390/s23020560.
[21] M. Anthimopoulos, J. Dehais, P. Diem, and S. Mougiakakou, “Segmentation and recognition of multi-food meal images for carbohydrate counting,” in *13th IEEE International Conference on BioInformatics and BioEngineering*, Chania, Greece: IEEE, Nov. 2013, pp. 1–4. doi: 10.1109/BIBE.2013.6701608.
[22] X. Huang *et al.*, “Application of Image Computing in Non-Destructive Detection of Chinese Cuisine,” *Foods*, vol. 14, no. 14, p. 2488, Jul. 2025, doi: 10.3390/foods14142488.

Vol. 6, No. 4, August 2025, Page. 2484-2499
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.5001

[23] A. Mjahad, A. Polo-Aguado, L. Llorens-Serrano, and A. Rosado-Muñoz, "Optimizing Image Feature Extraction with Convolutional Neural Networks for Chicken Meat Detection Applications," *Appl. Sci.*, vol. 15, no. 2, p. 733, Jan. 2025, doi: 10.3390/app15020733.
[24] K. Shehzad, U. Ali, and A. Munir, "Computer Vision for Food Quality Assessment: Advances and Challenges," *SSRN Electron. J.*, 2025, doi: 10.2139/ssrn.5196776.
[25] M. Rashed, S. Fakhry, and R. Satour, "Revolutionizing Food Quality With Machine Vision and Machine Learning Techniques:," in *Food in the Metaverse and Web 3.0 Era*, A. M. Alhussaini Hamad and R. Soni, Eds., IGI Global, 2025, pp. 71–124. doi: 10.4018/979-8-3693-9025-2.ch005.
[26] Z. Zhu and Y. Dai, "Improving Food Segmentation Through Selecting Suitable Feature Representations of Image Pixels," in *Pattern Recognition. ICPR 2024 International Workshops and Challenges*, vol. 15617, S. Palaiahnakote, S. Schuckers, J.-M. Ogier, P. Bhattacharya, U. Pal, and S. Bhattacharya, Eds., in Lecture Notes in Computer Science, vol. 15617. , Cham: Springer Nature Switzerland, 2025, pp. 34–48. doi: 10.1007/978-3-031-88217-3_3.
[27] S. Ghanghas, N. Kumar, V. K. Singh, S. Kumar, S. Birania, and A. Kumar, "Image Processing Technology, Imaging Techniques, and Their Application in the Food Processing Sector," in *Nonthermal Food Engineering Operations*, 1st ed., N. Kumar, A. Panghal, and M. K. Garg, Eds., Wiley, 2024, pp. 193–223. doi: 10.1002/9781119776468.ch6.
[28] U. S. Khan *et al.*, "Unsupervised Color Segmentation with Reconstructed Spatial Weighted Gaussian Mixture Model and Random Color Histogram," *Comput. Mater. Contin.*, vol. 78, no. 3, pp. 3323–3348, 2024, doi: 10.32604/cmc.2024.046094.
[29] Y. Lu, S. Young, H. Wang, and N. Wijewardane, "Robust plant segmentation of color images based on image contrast optimization," *Comput. Electron. Agric.*, vol. 193, p. 106711, Feb. 2022, doi: 10.1016/j.compag.2022.106711.
[30] K. Kiswanto, H. Hadiyanto, and E. Sediyono, "Meat Texture Image Classification Using the Haar Wavelet Approach and a Gray-Level Co-Occurrence Matrix," *Appl. Syst. Innov.*, vol. 7, no. 3, p. 49, Jun. 2024, doi: 10.3390/asi7030049.
[31] J. Ling, D. R. Heldman, and 1. School of Computer Science, Guangzhou Maritime University, Guangzhou 510725, China2. Department of Food Science and Technology, The Ohio State University, Columbus, OH 43210, USA, "Integration of machine learning technologies in food flavor research: Current applications, challenges, and future perspectives," *Int. J. Agric. Biol. Eng.*, vol. 18, no. 3, pp. 1–11, 2025, doi: 10.25165/j.ijabe.20251803.9623.
[32] R. S and G. Battineni, "A Survey on Recent Advancements in Auto-Machine Learning with a Focus on Feature Engineering," *J. Comput. Cogn. Eng.*, May 2023, doi: 10.47852/bonviewJCCE3202720.
[33] W. Zhang, E. Chen, M. Anderson, S. Thompson, and J. Lee, "Multimodal Deep Learning-Based Intelligent Food Safety Detection and Traceability System," *Int. J. Manag. Sci. Res.*, vol. 8, no. 3, pp. 71–76, Mar. 2025, doi: 10.53469/ijomsr.2025.08(03).09.
[34] W. Ji, K. Zhai, B. Xu, and J. Wu, "Green Apple Detection Method Based on Multidimensional Feature Extraction Network Model and Transformer Module," *J. Food Prot.*, vol. 88, no. 1, p. 100397, Jan. 2025, doi: 10.1016/j.jfp.2024.100397.
[35] G. A. Tahir and C. K. Loo, "Progressive Kernel Extreme Learning Machine for Food Image Analysis via Optimal Features from Quality Resilient CNN," *Appl. Sci.*, vol. 11, no. 20, p. 9562, Oct. 2021, doi: 10.3390/app11209562.
[36] M. Luo, W. Min, Z. Wang, J. Song, and S. Jiang, "Ingredient Prediction via Context Learning Network With Class-Adaptive Asymmetric Loss," *IEEE Trans. Image Process.*, vol. 32, pp. 5509–5523, 2023, doi: 10.1109/TIP.2023.3318958.
[37] W. M. Thu, Dr. P. E. San, and D. W. T. Tun, "GLCM and LTP Based Classification of Food Types," *Int. J. Sci. Eng. Appl.*, pp. 218–222, Apr. 2023, doi: 10.7753/IJSEA0708.1018.
[38] A. A. Gade and A. J. Vyavahare, "Feature Extraction using GLCM for Dietary Assessment Application," *Int. J. Multimed. Image Process.*, vol. 8, no. 2, pp. 409–413, Jun. 2018, doi: 10.20533/ijmip.2042.4647.2018.0050.

[39] A. Tatsuma and M. Aono, “Food Image Recognition Using Covariance of Convolutional Layer Feature Maps,” *IEICE Trans. Inf. Syst.*, vol. E99.D, no. 6, pp. 1711–1715, 2016, doi: 10.1587/transinf.2015EDL8212.
[40] A. Arnita, F. Marpaung, Z. A. Koemadji, M. Hidayat, A. Widianto, and F. Aulia, “Selection of Food Identification System Features Using Convolutional Neural Network (CNN) Method,” *Sci. J. Inform.*, vol. 10, no. 2, pp. 205–216, May 2023, doi: 10.15294/sji.v10i2.44059.
[41] N. Veernagouda Ganganagowder and P. Kamath, “Intelligent classification models for food products basis on morphological, colour and texture features,” *Acta Agronómica*, vol. 66, no. 4, pp. 486–494, Oct. 2017, doi: 10.15446/acag.v66n4.60049.