

Abstract

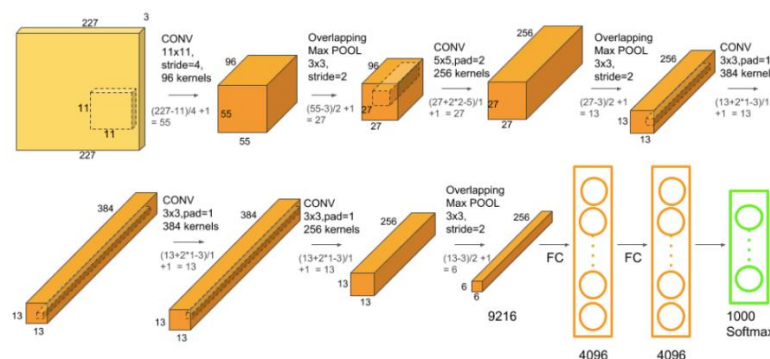
This report presents an in-depth analysis of food image classification using the Food-11 dataset. Employing two renowned Convolutional Neural Networks (CNNs), AlexNet and GoogLeNet, both feature extraction and transfer learning methodologies were explored. A balanced subset of the dataset was used to train and compare classifiers such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Neural Networks (NN), including a blended ensemble model. The findings highlight dataset characteristics, feature extraction efficacy, and classifier performance, demonstrating GoogLeNet's superior capabilities over AlexNet in achieving higher test accuracies and more robust generalization.

1. Introduction

Food image classification underpins applications like dietary monitoring, automated food logging, and nutrition recognition. Pre-trained CNNs are especially beneficial for such tasks, allowing efficient training and effective feature extraction from complex datasets. This report focuses on applying AlexNet and GoogLeNet to classify images from the Food-11 dataset into eleven food categories. Experiments cover both feature extraction and transfer learning, providing insights into the models' performance, misclassification patterns, and generalization capabilities.

1.1 Brief Analysis of AlexNet

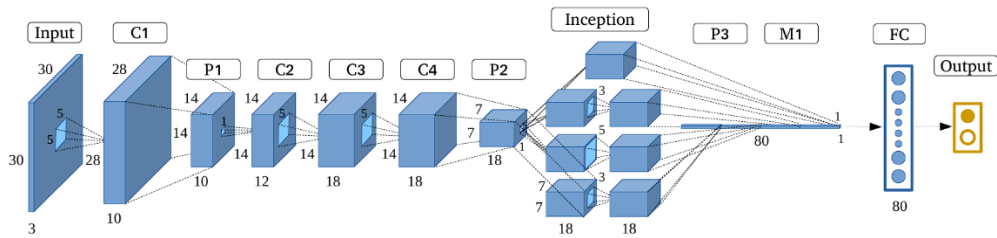
AlexNet is a pioneering CNN architecture that popularized deep learning for large-scale image recognition tasks. It comprises eight layers: five convolutional layers followed by three fully connected layers, using ReLU activations and dropout to mitigate overfitting. AlexNet's main strength lies in its relatively moderate depth and use of learned filters that capture low- and mid-level features effectively. Its input size is 227×227 pixels.



Architecture of AlexNet

1.2 Brief Analysis of GoogLeNet

GoogLeNet introduces the Inception module, enabling parallel convolution operations with different filter sizes. This deeper architecture (22 layers) reduces computational costs by factoring convolutions into smaller operations. Consequently, GoogLeNet can learn richer, hierarchical features, often outperforming shallower networks in complex classification tasks. Its input size is 224×224 pixels.



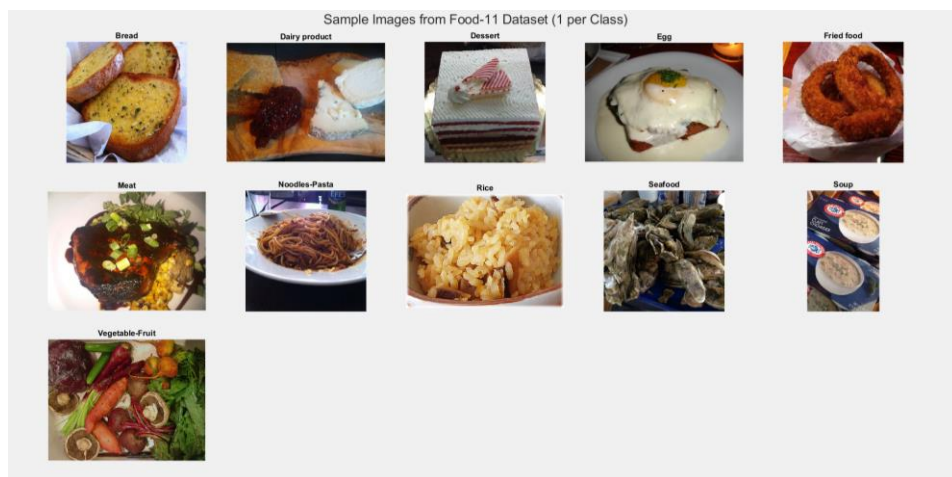
Architecture of GoogleLeNet

2. Dataset Description and Analysis

The **Food-11 dataset** contains images of eleven food categories: Bread, Dairy Product, Dessert, Egg, Fried Food, Meat, Noodles-Pasta, Rice, Seafood, Soup, and Vegetable-Fruit. Each class exhibits distinct visual properties in terms of color, texture, brightness, contrast, and sharpness.

2.1 Sample Images

Below is a representative sample from the Food-11 dataset, showing one image per class.

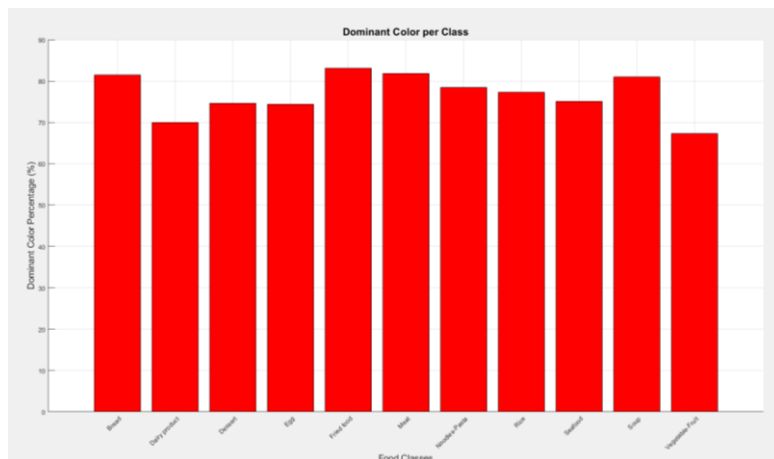


Sample images from Food-11Dataset with 1 image per class

2.2 Color Dominance

A **notable finding** is the overwhelming dominance of the red (R) channel across most classes. For example, Bread displays an R count of 324,018,001 out of 397,831,881 total pixels, and even Vegetable-Fruit, though less red-dominant at ~67%, still contains a substantial red count of 632,176,079 out of 939,233,717 pixels.

This **pervasive red dominance** may bias CNNs toward color-based features, potentially overshadowing texture and shape cues. Classes with similar red intensities, such as Dessert and Dairy Product, risk higher misclassification rates if the model relies too heavily on color.



Dominant Color per Class

2.3 Image Size Statistics

The Food-11 dataset exhibits **considerable variability** in image dimensions:

- **Minimum size** encountered: 207×270 pixels
- **Maximum size** encountered: 6144×9216 pixels
- **Most common size**: 512×512 pixels

Since CNNs require consistent input sizes, images were resized (using MATLAB's *augmentedImageDatastore*) to the networks' required input—**227×227 for AlexNet** and **224×224 for GoogLeNet**. This direct scaling does not preserve aspect ratio, potentially distorting objects and altering feature integrity.

2.4 Image Quality Metrics

Classes differ in **brightness, contrast, and sharpness**:

Class	Brightness	Contrast	Sharpness
Bread	122.19	55.0	341.94
Dairy Product	136.14	53.89	245.56
Dessert	119.94	61.61	371.7
Egg	126.2	58.76	280.46
Fried Food	120.9	57.36	260.46
Meat	111.0	65.29	503.15
Noodles-Pasta	131.49	53.65	410.55
Rice	128.24	51.49	365.21
Seafood	112.33	62.18	442.13
Soup	117.53	55.38	213.81
Vegetable-Fruit	122.55	59.67	986.96

- **Brightness** ranges from a high of ~136.14 (Dairy Product) to a low of ~111.00 (Meat).
- **Contrast** peaks at ~65.29 (Meat) and falls to ~51.49 (Rice).
- **Sharpness** is notably high for Vegetable-Fruit (~986.96) but lower for Soup (~213.81).

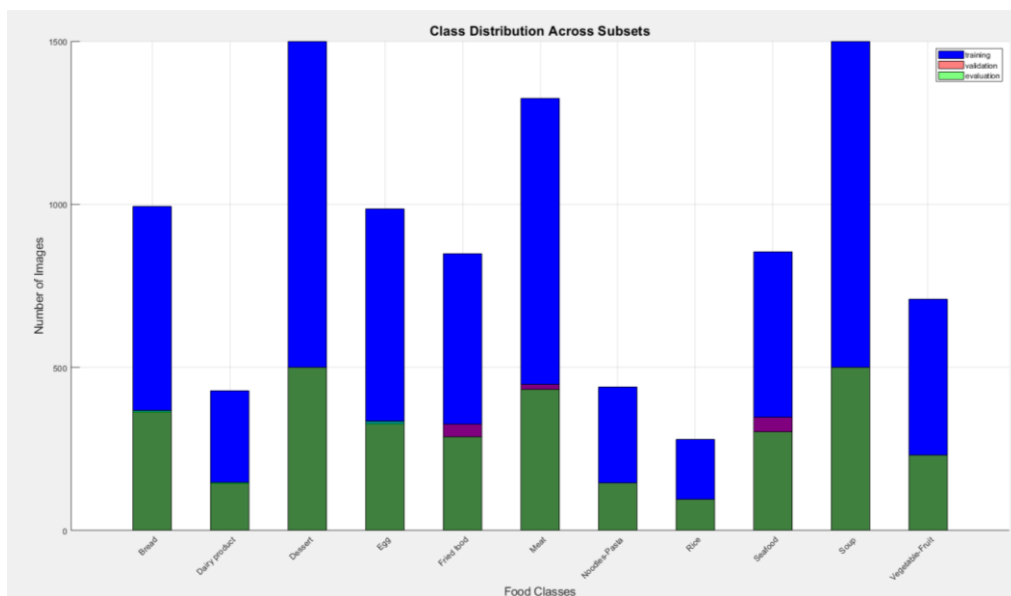
These variations stem from the inherent nature of each food. Dairy Products often appear lighter, Meats are darker with more texture, Vegetable-Fruit are vibrant with sharper edges, and Soups

look blurrier due to liquid consistency. Such differences impact the **feature extraction** process, where high sharpness or contrast can aid classification, while lower-quality images can obscure essential details.

2.5 Class Distribution Across Subsets

Although the **original dataset** is imbalanced, the chosen subset for experiments was balanced to avoid bias. Class counts in training, validation, and test sets were equalized to ensure fair model evaluation.

Class	Training	Validation	Test
Bread	994	362	368
Dairy Product	429	144	148
Dessert	1500	500	500
Egg	986	327	335
Fried Food	848	326	287
Meat	1325	449	432
Noodles-Pasta	440	147	147
Rice	280	96	96
Seafood	855	347	303
Soup	1500	500	500
Vegetable-Fruit	709	232	231



Class Distribution Across Subsets Blue : Training / Orange : Validation / Green : Test

Implication: Real-world classification often contends with imbalance, prompting methods like data augmentation, resampling, or class weighting. Here, balancing ensures that each class contributes equally, preventing skewed learning and yielding more reliable accuracy assessments.

3. Data Allocation Rationale

Each class received **280 training images**, along with 80 validation and 80 test images (when available in the original dataset) to maintain uniform coverage. For classes lacking 280 images,

all available samples were used. This allocation **maximizes data usage** and promotes consistent training while preserving enough data for validation and testing. Balanced data also mitigates bias and fosters **fair performance comparisons** across classes.

4. Experiment Methodology

4.1 Preprocessing Steps

All images were resized to each network's **required input size** 227×227 for AlexNet, 224×224 for GoogLeNet. MATLAB's *augmentedImageDatastore* handled this scaling, ensuring uniformity at the potential cost of slight distortion.

4.2 Feature Extraction and Transfer Learning

Feature Extraction:

For AlexNet, the penultimate layer **fc7** was used. For GoogLeNet, features were extracted from the **pool5-drop_7x7_s1** layer. These features powered classifiers such as SVM, KNN, and NN. A **blended ensemble** combined their predictions.

Transfer Learning:

In transfer learning, the **final fully connected and classification layers** were replaced with new layers tailored to the eleven classes. The networks were **fine-tuned** for 10 epochs using Stochastic Gradient Descent with Momentum (SGDM), a mini-batch size of 32, and an initial learning rate of 1e-4. Validation occurred every 30 iterations (each iteration corresponds to processing one mini-batch) to **monitor overfitting** and adjust training hyperparameters.

5. Experiments and Results

5.1 Experiment II: AlexNet as Feature Extractor

Methodology

AlexNet's **fc7** layer generated features for all training, validation, and test images. SVM, KNN, and NN classifiers were trained on these features, along with a **blended model** that combines their outputs.

Classifier	Train Acc (%)	Val Acc (%)	Test Acc (%)
SVM	100.0	74.21	73.98
KNN	76.85	66.02	67.05
NeuralNet	78.51	72.16	71.02
Blended Model	nan	nan	74.09

Classifier	Train Precision (%)	Val Precision (%)	Test Precision (%)
SVM	100.0	74.98	73.82
KNN	78.51	68.17	68.55
NeuralNet	78.38	72.35	70.61
Blended Model	nan	nan	74.71

Classifier	Train Recall (%)	Val Recall (%)	Test Recall (%)
SVM	100.0	74.21	73.98
KNN	76.85	66.02	67.05
NeuralNet	78.51	72.16	71.02
Blended Model	nan	nan	74.09

Classifier	Train F1 (%)	Val F1 (%)	Test F1 (%)
SVM	100.0	74.41	73.81
KNN	76.85	66.34	66.34
NeuralNet	78.29	72.11	70.53
Blended Model	nan	nan	74.02

Confusion Matrix Blended Model											
True Class	Bread	Dairy product	Dessert	Egg	Fried food	Meat	Noodles-Pasta	Rice	Seafood	Soup	Vegetable-Fruit
	58	1	3	3	4	4		3	1	2	1
	6	55	5	1	5	2		2	3	1	
	11	14	36	5	4	5	1		4		
	11	4	5	49	1	2	2	2	3	1	
	6	10	1	4	55	1			2		1
	9		3	1	3	62			1	1	
	2		1		1		72	3		1	
			1		1		2	76			
	4	5	3	5	2	2		1	51	2	5
		1	2	3	1	2	1	3		67	
	2	3		1		1		1	1		71
Predicted Class											

Confusion Matrix of the Blended Model with Alexnet

Analysis

SVM perfectly fit the training data (100% accuracy) but dropped to ~74% for validation and testing, indicating **overfitting**.

KNN displayed the lowest performance (~67% test accuracy), likely due to **distance-based limitations** in high-dimensional feature spaces.

Neural Network (NN) achieved moderate generalization (~71% test accuracy) with balanced precision, recall, and F1 scores around 70–72%.

Blended model reached ~74.09% test accuracy (F1 ~74.02%), capitalizing on each classifier's strengths. Classes such as Noodles-Pasta, Rice, Soup, and Vegetable-Fruit were identified correctly at high rates, but there were notable confusions among Bread, Dessert, and Dairy Product, which share similar **color and texture**.

Implications

Substantial overfitting from SVM and limited KNN performance highlight the difficulty of classifying diverse food images using purely color-dominant features. More robust feature engineering or dimensionality reduction may be necessary. Misclassifications between visually similar classes stress the need for **texture** and **shape-based** cues to complement color features.

5.2 Experiment III: Transfer Learning with AlexNet

Methodology

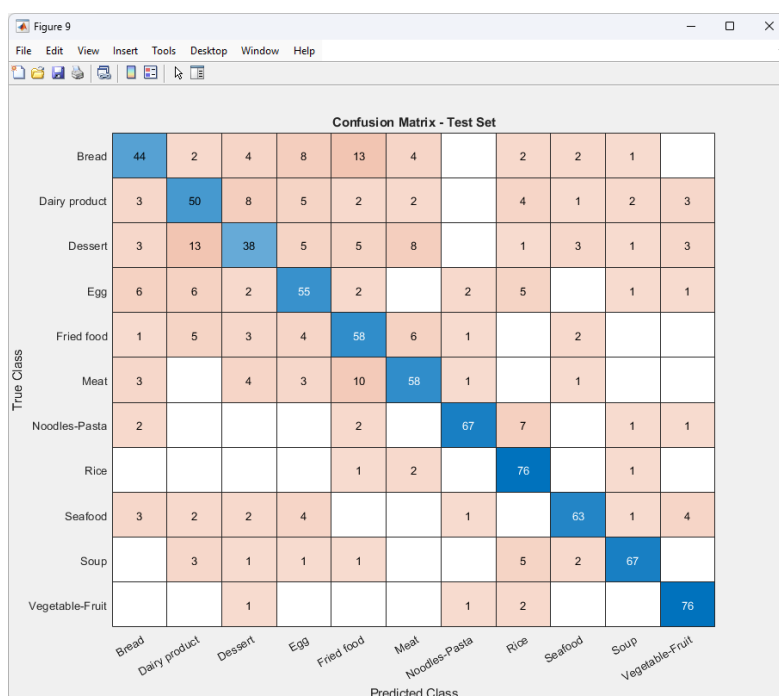
The final fully connected and classification layers of **AlexNet** were replaced to accommodate eleven food classes. Fine-tuning ran for **10 epochs** using SGDM (learning rate: 1e-4).

Classifier	Train Acc (%)	Val Acc (%)	Test Acc (%)
Transfer_AlexNet	97.96	73.52	74.09

Classifier	Train Precision (%)	Val Precision (%)	Test Precision (%)
Transfer_AlexNet	97.96	73.57	74.16

Classifier	Train Recall (%)	Val Recall (%)	Test Recall (%)
Transfer_AlexNet	97.96	73.52	74.09

Classifier	Train F1 (%)	Val F1 (%)	Test F1 (%)
Transfer_AlexNet	97.95	73.32	73.78



Confusion Matrix of the Transfer Learning with AlexNet

Analysis

Training accuracy approached ~98%, confirming that fine-tuning helped AlexNet better adapt to the dataset.

Despite high training accuracy, **validation and test accuracies** hovered around ~74%, comparable to feature extraction results.

Balanced precision, recall, and F1 scores indicate consistent performance, but **overfitting** remains an issue.

Training Duration: ~38 minutes.

Implications

Although **fine-tuning** enhances class-specific feature extraction, the model still struggles to generalize beyond ~74% test accuracy. Additional regularization, data augmentation, or hyperparameter tuning may help **bridge the gap** between training and testing performance.

5.3 Experiment IV: Using GoogLeNet

Repeating Experiments II and III with GoogLeNet underscores architectural differences and provides a **comparative perspective** on performance.

5.3.1 Feature Extraction with GoogLeNet

Methodology


GoogLeNet's **pool5-drop_7x7_s1** features were extracted for training, validation, and test images. SVM, KNN, NN, and a blended model were tested.

Classifier	Train Acc (%)	Val Acc (%)	Test Acc (%)
SVM	100.0	77.96	83.07
KNN	82.73	72.96	76.71
NeuralNet	82.08	76.02	81.02
Blended Model	nan	nan	83.3

Classifier	Train Precision (%)	Val Precision (%)	Test Precision (%)
SVM	100.0	78.13	83.36
KNN	83.7	73.39	77.98
NeuralNet	82.08	76.25	81.16
Blended Model	nan	nan	83.95

Classifier	Train Recall (%)	Val Recall (%)	Test Recall (%)
SVM	100.0	77.96	83.07
KNN	82.73	72.96	76.71
NeuralNet	82.08	76.02	81.02
Blended Model	nan	nan	83.3

Classifier	Train F1 (%)	Val F1 (%)	Test F1 (%)
SVM	100.0	77.92	83.09
KNN	82.46	72.19	76.22
NeuralNet	78.29	75.75	80.95
Blended Model	nan	nan	83.78

Confusion Matrix Blended Model													
True Class	Bread	67	5		1	5				1	1		
	Dairy product	1	66	3	4	2	1		2		1		
	Dessert	6	11	51	4	1	3			4			
	Egg	10	2	2	57	2	1		6				
	Fried food	8	5	2	4	57	1		1	1	1		
	Meat	3		2	4	2	65		1	2	1		
	Noodles-Pasta		1					75	3		1		
	Rice					1			79				
	Seafood	2	2	4	3	1	1	1		66			
	Soup	2	2								76		
	Vegetable-Fruit		1	2	1			1	1			74	
		Bread	Dairy product	Dessert	Egg	Fried food	Meat	Noodles-Pasta	Rice	Seafood	Soup	Vegetable-Fruit	
		Predicted Class											

Confusion Matrix of the Blended Model with GoogleLeNet

Analysis

- **SVM** achieved ~83.07% test accuracy, a marked jump over the AlexNet-based SVM (~74%). Precision, recall, and F1 (~83%) all reflect improved generalization.
- **KNN** improved to ~76.71% but still lags in high-dimensional spaces.
- **NN** reached ~81.02% test accuracy, showing balanced performance with ~80–81% across all metrics.
- **Blended model** excelled at ~83.30% test accuracy, reducing misclassifications in classes like Bread, Dessert, and Egg.

Implications

GoogLeNet’s deeper architecture and Inception modules yield **richer features**, leading to better overall performance. Ensemble methods further boost accuracy by merging the strengths of individual classifiers and enhancing **robustness** to class overlap.

5.3.2 Transfer Learning with GoogLeNet

Methodology

GoogLeNet was **fine-tuned** by replacing its final classification layers with new layers for eleven classes. The network was trained for 10 epochs using SGDM (learning rate: 1e-4).

Classifier	Train Acc (%)	Val Acc (%)	Test Acc (%)
Transfer_GoogLeNet	88.506	77.841	81.932

Classifier	Train Precision (%)	Val Precision (%)	Test Precision (%)
Transfer_GoogLeNet	88.971	79.279	82.388

Classifier	Train Recall (%)	Val Recall (%)	Test Recall (%)
Transfer_GoogLeNet	88.506	77.841	81.932

Classifier	Train F1 (%)	Val F1 (%)	Test F1 (%)
Transfer_GoogLeNet	88.554	77.9	81.997

Confusion Matrix - Test Set											
True Class	Bread	Dairy product	Dessert	Egg	Fried food	Meat	Noodles-Pasta	Rice	Seafood	Soup	Vegetable-Fruit
	53	2	2	12	9		2				
	1	63	4	3	4	2		1	1	1	
	2	6	55	4	2	6			3		2
	6	1	3	63	3			3	1		
	2	5	2	4	60	4			2		1
	1		2	5	7	62			2	1	
				1			77	1		1	
		1			1		2	75			1
		2	5	4		1		1	66		1
	1	3	1				1			74	
			2	1	2		1		1		73
Predicted Class											

Confusion Matrix of the Transfert Learning with GoogLeNet

Analysis

- **Training accuracy** reached ~88.5%, with ~82% on the test set higher than AlexNet's best (~74%).
- Test precision (~82.39%), recall (~81.93%), and F1 (~81.99%) confirm the model's **balanced** performance across classes.
- **Training Duration:** ~85 minutes.

Implications

Fine-tuning GoogLeNet effectively **minimizes overfitting**, evidenced by a tight match between training and test accuracies. This superior performance suggests that deeper architectures can capture **subtle distinctions** between classes, thereby enhancing classification metrics across a variety of foods.

6. Discussion

6.1 Comparative Performance

GoogLeNet consistently **outperformed** AlexNet in both feature extraction and transfer learning. Transfer learning with GoogLeNet reached ~82% test accuracy, surpassing AlexNet's ~74%. Its **Inception modules** and greater depth provide more nuanced feature maps, enabling **stronger generalization** and higher classification metrics.

6.2 Impact of Dataset Characteristics

The **predominance of red** across all classes influences the models' reliance on color, particularly for classes like Dairy Product and Dessert that share similar hues. Image-size variability necessitated uniform resizing, which risks **feature distortion** if significant aspect ratio changes occur.

Key observations:

- **Color Dominance** can lead to misclassifications where classes share similar color profiles.
- **Image Quality** (brightness, contrast, sharpness) affects feature clarity. Sharp or high-contrast images (e.g., Vegetable-Fruit) often see better accuracy, while blurred classes (e.g., Soup) are more prone to errors.
- **Class Imbalance** in the original dataset remains a concern for real-world deployment, requiring augmentation or rebalancing to ensure fair performance.

6.3 Misclassification Patterns

Dairy Product and **Dessert** often get confused, reflecting **visual overlaps** in color and brightness. While GoogLeNet helps reduce these errors, identical color distributions still challenge the classifiers.

Potential Solutions:

- **Enhanced Feature Engineering:** Incorporate shape/texture descriptors to complement color.
- **Advanced Data Augmentation:** Use transformations that highlight invariant features and diminish reliance on color.
- **Custom Loss Functions:** Penalize frequent misclassifications more heavily to nudge the model toward learning finer distinctions.

6.4 Classifier Performance

- **SVM** often boasted high accuracies but overfitted on AlexNet features; it thrived with GoogLeNet's richer representations.
- **Neural Networks** balanced performance across subsets without severe overfitting.
- **KNN** remained limited by the curse of dimensionality, though it slightly improved with better feature inputs.
- **Blended Models** consistently topped single-model results, confirming the **value of ensembles** in tackling subtle class distinctions.

6.5 Importance of Balanced Data Allocation

Allocating an equal number of images per class—both for training and evaluation—kept the models from **favoring** overrepresented categories. Balanced data fosters more **reliable accuracy assessments** and a fair comparison of different methods, especially when dealing with multiple food classes of varying visual complexity.

Conclusion

This report analyzed the **Food-11 dataset** using AlexNet and GoogLeNet for both feature extraction and transfer learning. **GoogLeNet** emerged as the stronger architecture, consistently yielding higher accuracies and better generalization. Still, color bias (especially in red-dominant classes), variability in image size, and inherent similarities among certain food items (e.g., Dairy Product vs. Dessert) present classification challenges. Strategies like **advanced data augmentation**, **regularization**, and **custom feature engineering** could further enhance performance.

In practice, **balanced datasets** and robust training procedures (with proper tuning of hyperparameters) remain crucial. Future work may investigate alternative architectures or hybrid methods to refine classification quality. The results underline the practical importance of deeper architectures for **complex image classification**, as well as the necessity to address dataset-specific biases to achieve more accurate and generalizable models.