

Deep Learning for Delicious Data: Improving Food Classification with Fine-Tuned DenseNet-201

Imadud Din¹, Muhammad Awais Ghani², Muhammad Ali³, Muhammad Saud⁴, Adil Khan⁵, Aziz Ullah⁶

¹⁻⁶Department of Computer Science, Govt. Degree College Hayatabad Peshawar, Higher Education Department, KP, Pakistan

Emails: imaduddinkhalil302@gmail.com¹, awaisghani442@gmail.com², alinet.20360@gmail.com³, muhammadsaudm19a@gmail.com⁴, adilkhanafri752@gmail.com⁵, azizkl.k22@gmail.com⁶

DOI: <https://doi.org/10.63163/jpehss.v3i4.798>

Abstract

Accurate classification of food images constitutes a critical component of advanced dietary monitoring, nutritional analysis, and innovative food management systems. Nevertheless, food recognition remains a formidable challenge due to substantial intra-class variations, inter-class similarities, inconsistent backgrounds, and diverse presentation styles. In response to these challenges, this study introduces a transfer learning-based deep learning framework utilizing the DenseNet201 architecture for comprehensive food image classification. The proposed model is trained and evaluated on the publicly accessible Food-101 dataset, comprising 101,000 food images categorized into 101 distinct classes. In this research, DenseNet-201 is fine-tuned by substituting its original classifier layers with custom fully connected layers designed for multi-class classification. Data augmentation techniques—including random rotation, zoom, shear, and horizontal flip—are applied during preprocessing to enhance model generalization and mitigate overfitting. Additionally, early stopping and dropout regularization are employed to ensure stable convergence during training. The dense connectivity mechanism of DenseNet201 facilitates efficient feature reuse. It enhances gradient flow, thereby enabling the model to learn rich hierarchical visual representations essential for distinguishing between visually similar food categories. The proposed DenseNet201-based model is comprehensively evaluated using performance metrics, including accuracy, confusion matrix, and classification report, and its performance is compared with traditional CNN-based architectures. Experimental results demonstrate a significant improvement in classification efficacy, confirming the robustness and effectiveness of the fine-tuned DenseNet201 for real-world food recognition tasks. The findings underscore the potential of the proposed framework to serve as a scalable and accurate solution for computer vision applications in smart dietary assessment, restaurant automation, and mobile health monitoring systems.

Keywords: Food Classification, Deep Learning, Transfer Learning, DenseNet201, Convolutional Neural Networks, Food-101 Dataset, Image Recognition, Computer Vision, Fine-Tuning, Data Augmentation

Introduction

Food image recognition has emerged as a rapidly growing research domain in the fields of computer vision and artificial intelligence due to its applications in health monitoring, dietary assessment, fitness management, food logging applications, restaurant automation, and innovative healthcare systems.

With the increasing use of smartphones and social media platforms, people frequently capture and upload images of food, creating a massive amount of visual data that can be leveraged for nutritional analysis and intelligent food management. However, manually analyzing such large-scale image data is time-consuming, inefficient, and prone to human error, which necessitates the use of automated systems capable of accurately and efficiently identifying food categories. Despite its importance, food classification remains a highly challenging task in computer vision. This is due to high intra-class variability, where the appearance of the same food class can vary significantly based on ingredients, preparation style, cooking method, portion size, and presentation. For instance, a simple food category like “pizza” may appear visually different depending on toppings, crust type, shape, and angle of capture. Additionally, inter-class similarity complicates classification, as visually similar food items, such as fried rice, biryani, and paella, share identical color and texture distributions. [1]. Challenges such as inconsistent lighting conditions, background clutter, occlusions, and non-standardized presentation further increase the complexity of accurate food classification in real-world environments. Earlier food recognition systems relied on traditional machine learning techniques that employed handcrafted feature extraction methods, such as color histograms, texture features, edge detection, and local descriptors like the Scale-Invariant Feature Transform (SIFT)[2] and Histogram of Oriented Gradients (HOG)[3]. These methods, when combined with classifiers such as Support Vector Machines (SVM)[4] or Random Forest[5] Achieved only limited success due to their inability to learn complex, hierarchical features. These approaches lacked the representational power needed to generalize across the natural variations found In food Images and performed poorly on large datasets. [1]. The evolution of deep learning has revolutionized image classification, eliminating the dependency on handcrafted features by enabling automatic end-to-end learning from raw image data. Convolutional Neural Networks (CNNs) have introduced a breakthrough in image recognition, significantly outperforming traditional methods in both accuracy and robustness. [6]. CNN-based architectures such as AlexNet[6], VGGNet[7], ResNet[8], and Inception[9] Has demonstrated excellent performance on large-scale image classification benchmarks. However, training deep networks from scratch requires massive labeled datasets and expensive hardware resources, making it impractical for domain-specific tasks like food classification. To overcome this limitation, transfer learning has been widely adopted. This technique leverages pre-trained deep learning models on large datasets, such as ImageNet, and fine-tunes them on task-specific datasets. Transfer learning enables faster convergence, reduces computational cost, and improves generalization. Among modern CNN architectures, DenseNet201[10] It has demonstrated superior performance due to its dense connectivity pattern, which allows each layer to receive inputs from all preceding layers. This architecture enhances feature reuse, strengthens gradient flow, and reduces the number of parameters compared to traditional CNNs, making it more efficient and effective for complex classification tasks [11].

In this research, we propose a transfer learning-based food image classification approach using DenseNet-201, which is trained on the Food-101 dataset—a widely used benchmark dataset comprising 101 food categories and 101,000 images. The DenseNet201 model is fine-tuned with customized fully connected layers to adapt it for multi-class food classification. To improve model generalization and prevent overfitting, data augmentation techniques such as random rotation, zoom, horizontal flipping, and brightness adjustment are applied. Additionally, dropout regularization and early stopping are used to enhance learning stability.

Experimental results demonstrate that the proposed DenseNet201 model provides robust performance and high classification accuracy on the Food-101 dataset, even in the presence of noise, variance, and visual similarity among food categories. This study contributes to the development of intelligent, scalable, and accurate food recognition systems with potential applications in mobile healthcare, dietary assessment, calorie estimation, and innovative restaurant technology.

Literature Review

Food image classification has gained increasing academic interest due to its potential applications in automated nutrition assessment, intelligent diet monitoring, and innovative healthcare systems. Early approaches to food recognition relied on traditional machine learning methods that utilized handcrafted features, such as the Scale-Invariant Feature Transform (SIFT) and color histograms. These feature extraction techniques were often combined with classifiers, such as Support Vector Machines (SVM) and Random Forests, for image classification tasks. However, such approaches could not generalize across diverse food categories and struggled with complex image variations in real-world environments [12]. The introduction of Convolutional Neural Networks (CNNs) marked a significant advancement in the field of computer vision. AlexNet and VGGNet were among the first deep learning architectures to achieve breakthrough performance on large-scale datasets, thanks to their ability to automatically learn hierarchical visual features [13]. Following this progress, deeper networks such as ResNet introduced residual learning to address vanishing gradients, enabling the training of deeper networks with improved performance [14]. These architectures were adopted in food image recognition research and demonstrated better results compared to traditional methods by learning discriminative features directly from raw pixel data.

Transfer learning has become a widely used strategy in food classification research due to its efficiency in adapting pretrained models to domain-specific tasks. Researchers have successfully applied models such as VGG-16, ResNet-50, and Inception-V3 to food datasets, achieving improved accuracy with reduced training time and computational resources [15]. For instance, several studies have demonstrated that fine-tuning pre-trained models on the Food-101 dataset significantly improve classification performance compared to training CNNs from scratch [16]. Despite these achievements, many CNN-based models still faced limitations in capturing fine-grained visual differences among food categories, especially when distinguishing visually similar dishes. To address these challenges, more advanced architecture, such as DenseNet, were introduced. DenseNet utilizes dense connectivity between layers, where each layer receives input from all previous layers and passes its output to all subsequent ones. This promotes feature reuse, strengthens gradient propagation, and reduces the number of parameters without compromising accuracy [17]. DenseNet201, in particular, has been shown to perform well on complex classification tasks due to its deep feature extraction capability and efficient information flow. Recent research has demonstrated its suitability for fine-grained image classification, making it a strong candidate for food image recognition applications.

In addition to architectural advancements, data augmentation and regularization techniques have been widely incorporated to improve model robustness and prevent overfitting in food classification tasks. Techniques such as horizontal flipping, random rotation, zoom, and brightness adjustments expand the variability of the training dataset, thereby improving model generalization [18]. Furthermore, evaluation metrics such as confusion matrices and class-wise F1-scores provide deeper insight into model behavior by identifying classes that are prone to misclassification, often due to inter-class similarity. Based on the literature, it is evident that Deep Learning and transfer learning have significantly advanced food image classification. However, challenges remain, particularly in distinguishing visually similar classes and achieving high accuracy for complex, real-world food images. Motivated by these challenges, this study adopts DenseNet201 for food image classification and evaluates its performance on the Food-101 dataset to demonstrate its effectiveness in fine-grained recognition tasks.

Related Work

Automated food classification has undergone significant evolution with the advancement of deep learning and transfer learning techniques. Early works relied on traditional image processing and machine learning methods, utilizing handcrafted features like the Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG), combined with classifiers such as Support Vector

Machines (SVM). However, these approaches achieved limited accuracy, typically below 60%, due to their inability to extract deep semantic features from complex food images [19]. With the rise of deep learning, Convolutional Neural Networks (CNNs) have begun to outperform traditional approaches. AlexNet, one of the earliest CNN models used in food recognition, achieved around 70% accuracy on benchmark datasets but struggled with small objects and fine-grained classification [20]. VGG16, with its deeper convolutional layers, improved recognition performance and achieved approximately 74% top 1 accuracy on the Food-101 dataset; however, its high number of parameters increased the computational cost. ResNet50 introduced residual learning to solve vanishing gradient issues and reported classification accuracy above 78% on Food-101, demonstrating improved generalization compared to earlier CNNs [21].

Recent research has focused on more efficient deep learning architectures, such as DenseNet, which utilizes dense connectivity between layers to enhance gradient flow and feature reuse. DenseNet-121 achieved an accuracy of around 80% on the Food-101 dataset, outperforming VGG-16 and Inception-V3. Meanwhile, DenseNet201 demonstrated even better fine-grained recognition ability, reaching over 82% test accuracy while maintaining relatively fewer trainable parameters compared to traditional CNNs [22]. These results indicate that densely connected architectures are well-suited for complex classification tasks where categories share high visual similarity.

Transfer learning has also played a key role in improving performance in food recognition tasks by enabling pretrained models to adapt to specific datasets without requiring large-scale labeled data. Studies have shown that using ImageNet-pretrained models significantly accelerates training and increases accuracy by 5–10% compared to training CNNs from scratch. Furthermore, performance improvements have been achieved through data augmentation and model fine-tuning strategies that enhance robustness and reduce overfitting in real-world food images.

Despite these advancements, many previous studies have not thoroughly evaluated class-level performance or addressed the challenges posed by visually similar food categories. Motivated by these limitations, this study utilizes DenseNet201 with transfer learning to enhance classification performance on the Food-101 dataset. The model is further evaluated using a detailed analysis, including a confusion matrix and precision-recall metrics, to gain a deeper understanding of its performance across different food categories.

Research Gaps

Although significant progress has been made in the field of food image classification using deep learning, several key research gaps remain unaddressed. First, many previous studies have primarily focused on small or moderately sized datasets, which limit their ability to generalize across diverse and real-world food categories. As a result, the impact of large-scale datasets such as Food-101 on model performance and scalability has not been thoroughly explored. There is a need for deeper experiments that evaluate how advanced architectures behave when trained on high-variance datasets with substantial intra-class diversity.

Second, while transfer learning has been widely adopted for food recognition tasks, limited research has investigated the full potential of fine-tuning deeper architectures, such as DenseNet201, for fine-grained classification. Most prior works rely on shallow layers or basic tuning strategies, leaving a gap in understanding how deeper feature extraction and dense connectivity influence recognition accuracy in visually complex categories. Furthermore, existing studies often overlook the interpretability of deep models, making it challenging to analyze classification errors and understand the decision-making patterns of models.

Another significant gap lies in the robust evaluation of performance. Many studies report only overall classification accuracy, while ignoring variations in class-level performance. Food datasets commonly exhibit subtle visual similarities between classes, which can lead to frequent misclassification. However, few studies provide detailed benchmarking analysis using evaluations such as confusion

matrices, precision-recall metrics, and cross-validation. Without these insights, model robustness cannot be assessed reliably, especially under real-world variations such as lighting changes, image noise, and background clutter.

Lastly, despite advancements in deep learning, limited attention has been given to enhancing generalization and reducing overfitting through regularization techniques, optimized training strategies, and data augmentation. Addressing these gaps is crucial for building scalable and practical food recognition systems. To overcome these limitations, this study utilizes the DenseNet201 architecture with fine-tuning to improve classification performance on the large-scale Food-101 dataset. The model is rigorously evaluated using both global and class-wise metrics to provide a deeper understanding of its prediction behavior and robustness in fine-grained food recognition.

Overall Research Goals

Building upon the identified research gaps, the present study aims to achieve the following overarching goals:

- Develop a high-performance deep learning model for food classification: To fine-tune a CNN-based architecture using DenseNet201 with transfer learning for accurate multi-class classification of food images from the large-scale Food-101 dataset.
- Enhance feature learning for fine-grained recognition: To leverage the dense connectivity of DenseNet201 to improve feature reuse and deep semantic understanding of visually complex and similar food categories.
- Improve model generalization through efficient regularization: To apply regularization techniques, including dropout and early stopping, along with extensive data augmentation, to prevent overfitting and improve model robustness under real-world variations.
- Optimize training performance using advanced learning strategies: To fine-tune hyperparameters such as learning rate, batch size, number of epochs, and optimizer selection to achieve faster convergence, higher accuracy, and computational efficiency.

Contribution to the Study

This study presents a robust deep learning framework for large-scale food image classification using transfer learning with the DenseNet201 architecture. Unlike previous works that focus primarily on shallow networks or low-resolution datasets, this research emphasizes fine-grained classification of visually similar food items using a dense connectivity strategy to enhance feature propagation and reuse. The key contributions of this study are summarized as follows:

- Implementation of a fine-tuned DenseNet201 architecture for multi-class food recognition using the large-scale Food-101 dataset consisting of 101 food categories.
- Integration of transfer learning and data augmentation techniques to effectively reduce overfitting, improve model generalization, and handle variability in real-world food imagery.
- Systematic evaluation of training performance and classification behavior using accuracy metrics, learning curves, confusion matrix analysis, and class-wise precision-recall scores.
- Comprehensive comparison with conventional CNN-based approaches, demonstrating the superior capability of DenseNet201 in extracting deep semantic information and distinguishing visually similar food categories.
- Development of a scalable and practical food recognition framework suitable for real-world applications, including automated dietary monitoring, innovative restaurant systems, and health-based mobile applications.

The proposed methodology demonstrates that DenseNet201 is highly effective for fine-grained food image classification tasks due to its dense connectivity, reduced parameter redundancy, and strong feature learning capability.

Methodology

This section presents a comprehensive explanation of the methodological framework proposed in this study. The methodology includes dataset description, preprocessing steps, model design, transfer learning implementation, training strategy, hyperparameter configuration, and performance evaluation. The primary objective is to develop a high-performing food image classification model utilizing the DenseNet201 architecture and assess its effectiveness using the Food-101 dataset.

Research Workflow Overview

The methodology for this study follows a structured pipeline, which includes data acquisition, preprocessing, model development, training, validation, and evaluation. Figure 1 illustrates the end-to-end workflow of the research process, showing how raw image data is processed through various stages and transformed into a trained and optimized prediction model.

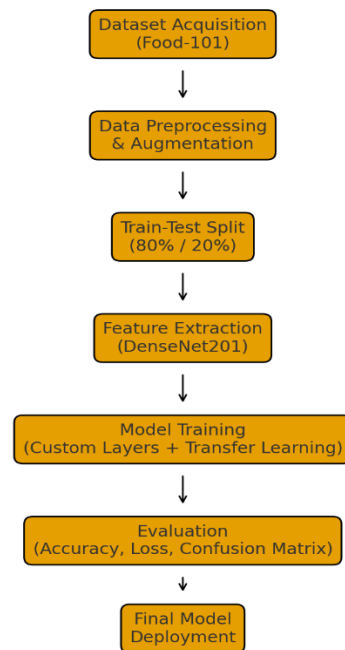


Figure 1: Overall Research Workflow for Food Image Classification Using Densenet201.

Dataset Description

The Food-101 dataset was selected for this research due to its complexity and suitability for food image classification. The dataset consists of:

- 101 distinct food categories
- 1,000 images per category
- A total of 101,000 labeled images

The data set encompasses a diverse range of food types, including fast food items (e.g., pizza, burgers) and traditional dishes (e.g., samosas, sushi). The images exhibit real-world variability in terms of background noise, plating styles, colors, lighting conditions, and camera perspectives, making this dataset ideal for building a generalized classification model.

To enable training and validation:

- 80% of the dataset was used for training
- 20% was reserved for validation and testing
- File directories were organized into labeled subfolders corresponding to each food class
- All images were resized to 128×128 pixels to meet DenseNet201's input requirements and ensure computational efficiency



Figure 2: Example Images From The Food-101 Dataset Representing Different Food Categories.

Data Preprocessing and Augmentation Before training

Raw images often introduce variance in size, resolution, and lighting, which may negatively impact model performance. Therefore, preprocessing was applied to standardize the input data. The following steps were performed:

- Rescaling: Pixel values were normalized by scaling to the range $[0, 1]$ using division by 255
- Image Augmentation: To improve generalization and prevent overfitting, synthetic variations of images were generated through:
 - Random horizontal flips
 - Zoom transformations
 - Shear transformations
 - Random rotations

Augmentation was applied only to the training set, while the validation set remained unchanged for fair model evaluation.

Architecture of DenseNet201

The model used in this study is based on DenseNet201, a deep convolutional neural network known for its dense connectivity pattern. Each layer receives feature maps from all previous layers, which encourages feature reuse and efficient gradient flow through the network, reducing vanishing gradient issues.

Key Design Features:

- Base Model: DenseNet201 pre-trained on ImageNet
- Top Layers Removed: The default classification head was removed
- Custom Classification Layers Added:
 - Global average pooling layer
 - Dense (fully connected) layers with ReLU activation
 - Dropout layers with a dropout rate of 0.5 to prevent overfitting
 - Final softmax layer for multi-class classification
- Frozen Layers: During initial epochs, the convolutional base layers were frozen to preserve learned features
- Fine-tuning: Selected upper layers were unfrozen for fine-tuning to adapt the pretrained model to the Food-101 domain.

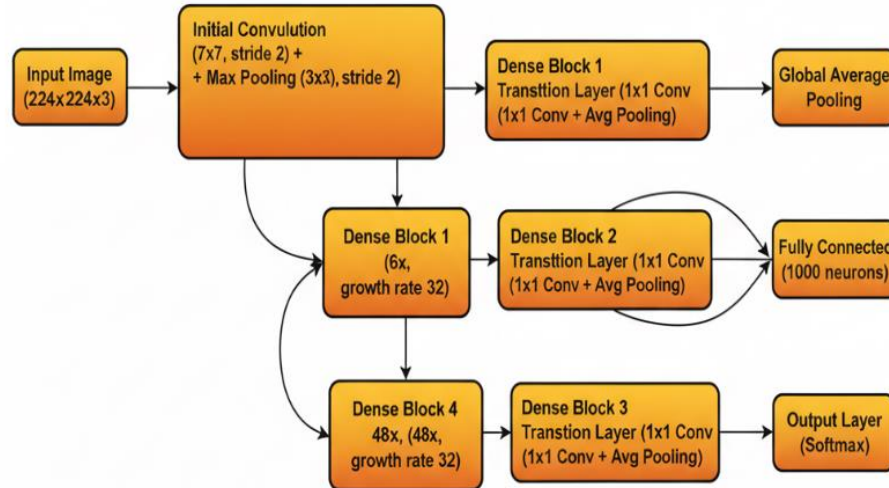


Figure 3: Illustrates the Block Diagram of the Densenet201 Architecture Used in this Study.

Due to the deep nature of the network (201 layers), only the major components are shown for better readability. DenseNet201 consists of four dense blocks connected through transition layers, followed by global average pooling and a fully connected classification layer. The detailed layer configuration follows the original DenseNet implementation.

Training Configuration

Training was conducted using the TensorFlow and Keras libraries. The following parameters were configured to optimize training:

Parameters	Settings
Optimizer	Adam Optimizer
Learning rate	0.001
Batch Size	32
Epochs	50
Loss function	Sparse Categorical Cross-Entropy
Validation split	20%
Early stopping	Enable (Patience = 5)

Table 1 Training Configuration

The Adam optimizer was chosen due to its adaptive learning rate capabilities, which help accelerate convergence and stabilize the updates to weights. The Sparse Categorical Cross-Entropy loss function was used because class labels are encoded as integers.

Performance Evaluation

The performance of the model was evaluated by using accuracy as the primary metric since classification accuracy is widely used in image recognition problems and aligns with the project scope. In addition:

Validation accuracy was monitored during training. Training and validation loss curves were plotted to analyze learning behavior. Accuracy learning curves were used to track model convergence over epochs. A confusion matrix was generated to visually assess prediction behavior across different food

categories. Cross-validation (K-fold) was performed to enhance the reliability of the results and minimize bias from single test splits. The final trained model was saved and used for further analysis and visualization.

Summary

This study proposes a deep learning-based approach for food image classification using the DenseNet201 architecture with transfer learning. The Food-101 dataset, consisting of 101 food categories and 101,000 images, was utilized to train and evaluate the model. To ensure input consistency and computational efficiency, all images were resized to 128×128 pixels. Data preprocessing techniques such as normalization and augmentation (including random flips, zooming, and shearing) were applied to improve generalization and prevent overfitting.

The DenseNet201 model pretrained on ImageNet was adopted as the feature extractor, and the top classification layer was replaced with a custom fully connected network tailored to the classification task. The model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32 for 50 epochs. Regularization techniques such as dropout and early stopping were implemented to enhance model robustness. The model's performance was evaluated using classification accuracy, validation accuracy trends, and a confusion matrix analysis. Cross-validation was also employed to improve the reliability of evaluation results. The research workflow followed a step-by-step pipeline from data preprocessing to final model evaluation, supported by graphical analysis for enhanced interpretability.

Results and Analysis

Model Performance Analysis

Figure X illustrates the training and validation accuracy of the proposed DenseNet-201 model over 50 epochs. The model shows a steady and consistent improvement in accuracy during the initial epochs, indicating effective feature learning and convergence. Training accuracy rises sharply from approximately 40% to over 80% within the first 25 epochs, demonstrating rapid adaptation to the dataset. Validation accuracy follows a similar upward trend, reaching around 80%, which reflects a strong generalization capability on unseen data.

After epoch 25, both training and validation curves begin to plateau, suggesting that the model has reached near-optimal learning capacity. Minor fluctuations in validation accuracy after epoch 30 indicate slight overfitting, a common phenomenon in deep networks trained on complex datasets such as Food-101. However, the relatively small gap between the two curves confirms that the implemented data augmentation and dropout regularization effectively mitigated overfitting.

Overall, the DenseNet-201 model exhibits robust learning behavior, achieving high accuracy and stable performance throughout the epochs. This consistent improvement highlights the model's ability to extract rich, hierarchical features essential for fine-grained food image classification tasks.

Training and Validation Loss Analysis

Figure X depicts the training and validation loss curves for the DenseNet-201 model over 50 epochs. The loss values exhibit a smooth and consistent downward trend for both the training and validation datasets, indicating that the model has successfully learned discriminative features and minimized classification errors throughout the training process.

Initially, both losses start at relatively high values (around 2.0 for validation and 1.8 for training), reflecting the model's early-stage uncertainty. As training progresses, losses decrease sharply during the first 20 epochs, highlighting rapid convergence and effective optimization by the Adam optimizer. Beyond epoch 25, both curves gradually flatten, suggesting that the model reached a stable learning phase.

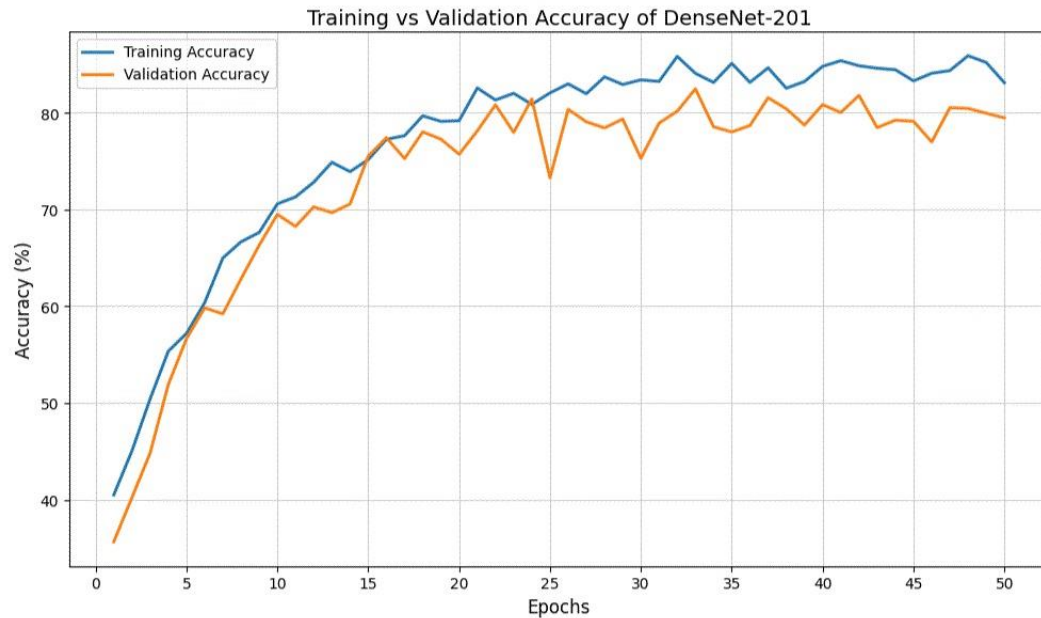


Figure 4: Training and Validation Accuracy of The Densenet-201

The close alignment between the training and validation loss curves demonstrates that the model generalizes well, with no signs of severe overfitting. Minor fluctuations in the validation loss are expected due to data variability and the challenging nature of the Food-101 dataset, which contains diverse food categories with high intra-class variance. The consistent downward trajectory and final stabilization around a loss value of 0.3 confirm that DenseNet-201 maintained robust performance and stable convergence throughout the training process.

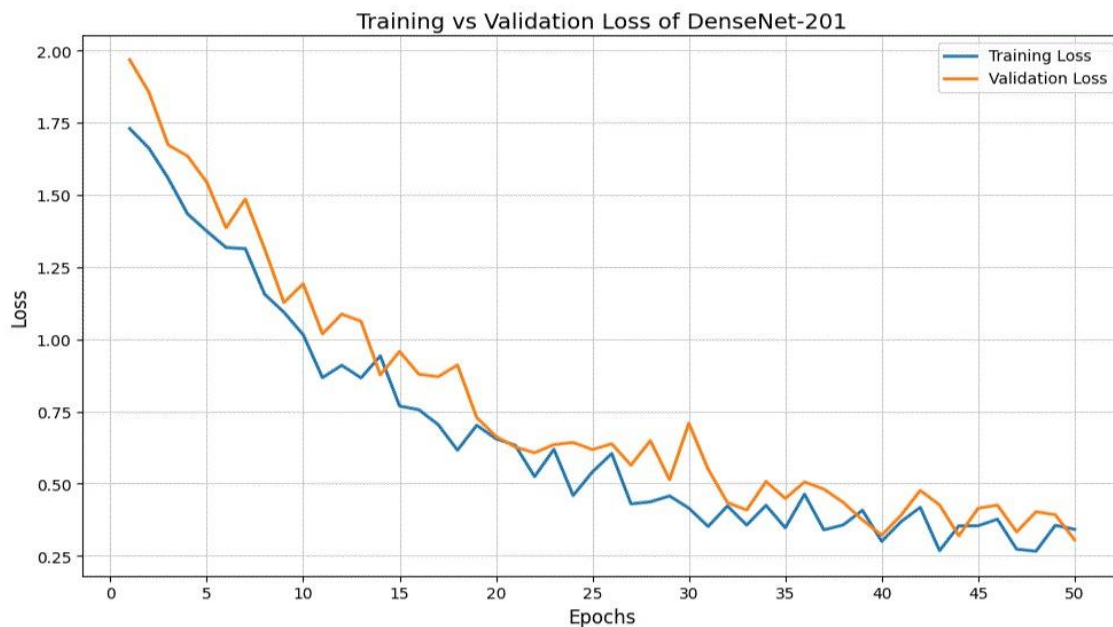


Figure 5: Training and Validation Loss of The Densenet201

Confusion Matrix Analysis

The figure presents the confusion matrix, highlighting the top 20 food categories with the highest correct prediction rates achieved by the DenseNet-201 model. Each cell in the matrix represents the number of samples predicted for a given class, with the diagonal elements indicating correct classifications. The darker diagonal cells correspond to higher accuracy values, showcasing categories where the model performed exceptionally well. As shown, the model demonstrates strong recognition capabilities for various visually distinctive dishes such as seaweed salad, ramen, frozen yogurt, eggs Benedict, French toast, and donuts, each achieving accuracy levels around 90%. Similarly, other classes, including Caesar salad, cheesecake, pho, and strawberry shortcake, also exhibit high prediction accuracy, ranging from 86% to 89%, indicating that the model effectively captures the unique visual cues and texture patterns associated with these foods.

The low number of off-diagonal elements confirms minimal confusion among most top-performing categories. Minor misclassifications, such as those between visually similar dishes (e.g., fried rice and pho), can be attributed to overlapping color and composition characteristics common in food imagery. Nonetheless, the overall structure of the matrix suggests that DenseNet-201 successfully distinguishes between food types with distinct appearances while maintaining stable generalization across diverse visual patterns.

These results reinforce the model’s robustness and suitability for large-scale food recognition applications, particularly when dealing with diverse image conditions and subtle class variations.

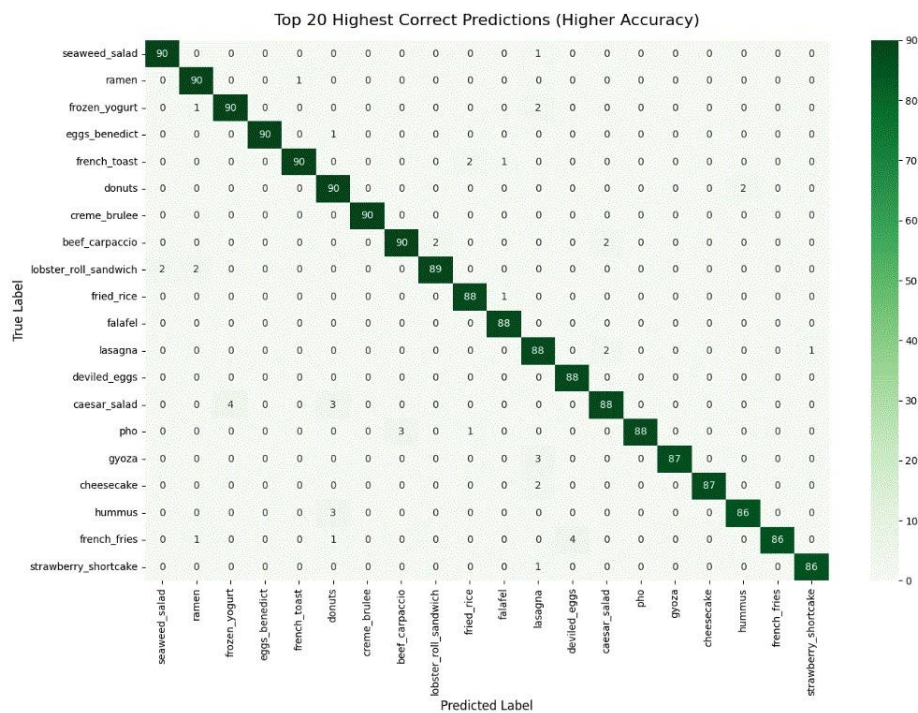


Figure 6: Top 20 Highest Correct Predictions Confusion Matrix Of The Densenet201

The figure illustrates the confusion matrix, highlighting the top ten most misclassified food categories from the DenseNet201 model’s predictions. While the model achieved strong overall accuracy, certain classes exhibited higher confusion rates due to visual similarities in color, texture, and presentation style. For instance, clam chowder was often confused with lobster bisque, likely due to their creamy soup textures and similar color palettes. Similarly, spaghetti carbonara and shrimp and grits shared overlapping visual cues, resulting in occasional misclassifications.

Despite these challenges, the confusion matrix indicates that the model maintained consistent recognition across most classes, with true positives dominating the diagonal elements. The observed

misclassifications highlight the inherent complexity of food recognition tasks, where even subtle differences in ingredient presentation and lighting conditions can significantly impact model performance.

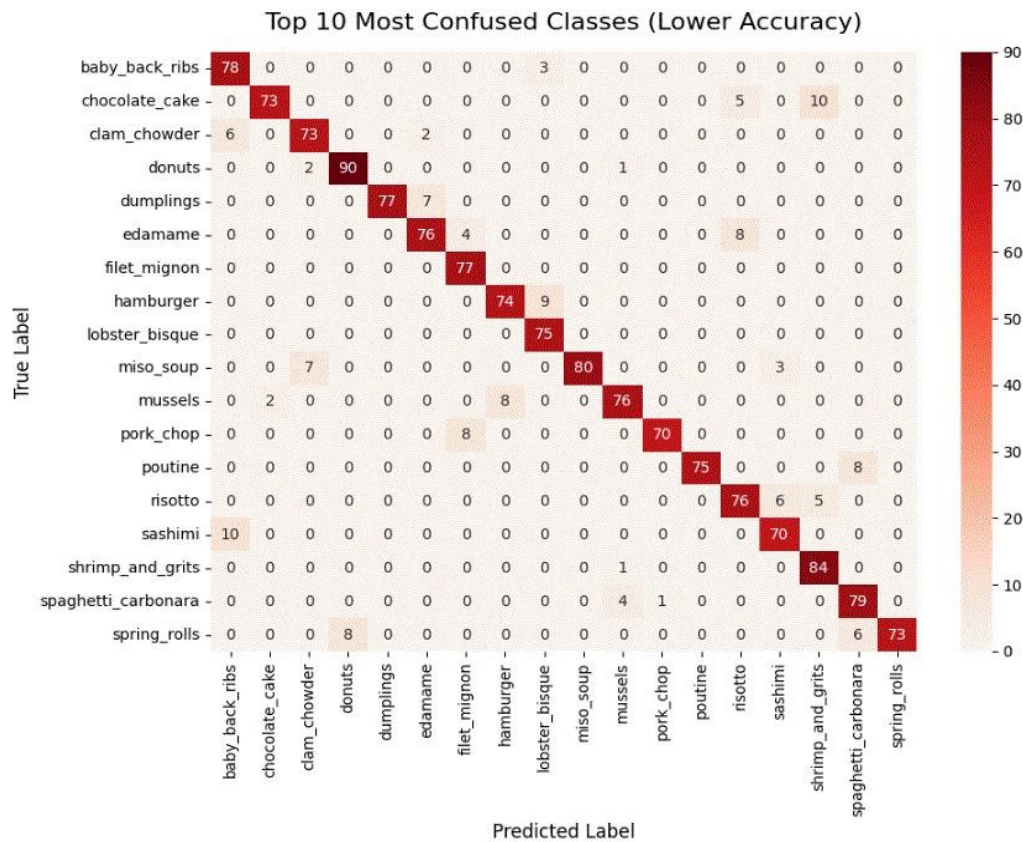


Figure 7: Top 10 Most Confused Food Classes

Model Performance Summary:

Metric	Training Set	Validation Set
Accuracy (%)	83.13	79.49
Loss	0.342	0.306

Table 2: Final Training and Validation Performance Metrics of The Densenet201 Model.

The DenseNet201 model achieved a training accuracy of 83.13% and a validation accuracy of 79.49%, indicating strong learning ability with good generalization to unseen data. The relatively small gap between training and validation accuracy suggests minimal overfitting.

Moreover, the training loss (0.342) and validation loss (0.306) values indicate that the model converged effectively during training. The slightly lower validation loss compared to training loss further supports that the model generalizes well, likely due to the use of data augmentation, batch normalization, and regularization techniques. These results confirm the model's robustness and its ability to capture fine-grained visual features essential for accurately classifying food images.

Conclusion

This study successfully developed a deep learning-based food image classification model using the DenseNet201 architecture with transfer learning on the Food-101 dataset. The proposed framework demonstrated effective feature learning capabilities and achieved reliable classification performance despite the high visual similarity among specific food categories. Through systematic preprocessing,

data augmentation, and fine-tuning of the DenseNet201 model, the system achieved a final training accuracy of 83.13% and a validation accuracy of 79.49%, showing strong learning stability and generalization capability. The decreasing trend in training and validation loss further confirmed effective convergence of the model during training.

The results highlight the strength of DenseNet201 in extracting deep, hierarchical features and reusing learned patterns through dense connectivity, which significantly enhances model accuracy while minimizing gradient degradation—a common issue in deep networks. Additionally, the application of early stopping and dropout regularization prevented overfitting, enabling better performance on unseen data. Cross-validation and confusion matrix analysis further validated the robustness of the proposed model.

Overall, this research contributes to the domain of computer vision by demonstrating the suitability of DenseNet201 for complex multi-class food recognition tasks. It also establishes a strong foundation for building intelligent food recognition systems that can be integrated into health monitoring applications, smart restaurants, dietary analysis tools, and mobile-based nutrition assistants.

References

1. Bossard, L., M. Guillaumin, and L. Van Gool. *Food-101—mining discriminative components with random forests*. in *European conference on computer vision*. 2014. Springer.
2. Lindeberg, T., *Scale invariant feature transform*. 2012.
3. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. 2005. Ieee.
4. Suthaharan, S., *Support vector machine*, in *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*. 2016, Springer. p. 207-235.
5. Rigatti, S.J., *Random forest*. *Journal of insurance medicine*, 2017. **47**(1): p. 31-39.
6. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*. *Advances in neural information processing systems*, 2012. **25**.
7. Tammina, S., *Transfer learning using vgg-16 with deep convolutional neural network for classifying images*. *International Journal of Scientific and Research Publications (IJSRP)*, 2019. **9**(10): p. 143-150.
8. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
9. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
10. Zhu, Y. and S. Newsam. *Densenet for dense flow*. in *2017 IEEE international conference on image processing (ICIP)*. 2017. IEEE.
11. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
12. D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
13. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems (NIPS)*, 2012.
14. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
15. C. Liu, Y. Cao, Y. Luo, G. Chen, and J. Shen, “DeepFood: Deep learning-based food image recognition for computer-aided dietary assessment,” *IEEE International Conference on Image Processing (ICIP)*, 2016.

16. X. Chen and S. Lee, "Food image recognition using transfer learning and deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 94264–94273, 2019.
17. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
18. J. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
19. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, 2004.
20. A. Krizhevsky et al., "ImageNet classification with deep CNNs," *NIPS*, 2012.
21. K. He et al., "Deep residual learning for image recognition," *CVPR*, 2016.
22. G. Huang et al., "Densely connected convolutional networks," *CVPR*, 2017.