PORT CITY INTERNATIONAL UNIVERSITY

Course Code: *CSE 454*

Course Title*: Digital Image Processing Sessonal*

Report Name: *Multimodal Food Classification Using Machine Learning, Deep Learning, and Hybrid Learning on a Recipe–Image Mapped Dataset*

Date Of Submission: *18th December, 2025*

Submitted By:

Subrata Biswas	***Mosammat Israt Arefin Mazumder***	***Jahanara Islam Hafsa***	***Subrina Islam Shanta***
CSE 02607330	***CSE 02807534***	***CSE 02807545***	***CSE 02807556***

Program: BSc in CSE
Batch: 28th (A)
Port City International University

Submitted To
Name of lecturer: Tahmina Akter
Department: CSE
Port City International University

Multimodal Food Classification Using Machine Learning, Deep Learning, and Hybrid Learning on a Recipe–Image Mapped Dataset

Subrata Biswas *CSE 02607330*

Mosammat Israt Arefin Mazumder *CSE 02807534*

Jahanara Islam Hafsa *CSE 02807545*

Subrina Islam Shanta *CSE 02807556*

Department of BSc in CSE
University of Portcity International University

K E Y W O R D S

Food Classification
Recipe Retrieval
Ingredient Prediction
Machine Learning
Deep Learning
Hybrid Learning
Multimodal Dataset
CNN, VGG, ResNet
InceptionV3

A B S T R A C T

The *Food Ingredients and Recipe Dataset with Image Name Mapping* provides a structured collection of recipes, corresponding ingredient lists, and mapped image file names that link each dish to its visual representation. This work utilizes the dataset's CSV file, which includes recipe titles, ingredient descriptions, preparation steps, and associated image names, enabling both textual and visual analysis. The objective of this study is to investigate how image–recipe mapping can support automated food classification, ingredient prediction, and recipe retrieval systems. We preprocess the dataset by normalizing ingredient text, cleaning recipe descriptions, and validating image file associations. Using this structured mapping, we develop and evaluate baseline machine learning models capable of predicting ingredients from images and retrieving relevant recipes. Experimental results demonstrate that the dataset's consistent image–name mapping significantly improves feature alignment between textual and visual modalities, leading to higher retrieval accuracy and more coherent ingredient predictions. This dataset therefore offers valuable potential for future research in food recognition, multimodal learning, dietary recommendation systems, and intelligent cooking assistants.

1 Introduction

The field of food data analytics has recently emerged as a vibrant interdisciplinary research area, bridging culinary sciences, data mining, and machine learning. With the exponential growth of digital food content, including recipes, ingredient lists, cooking instructions, and food images, researchers and practitioners are increasingly motivated to explore the underlying patterns, develop intelligent recommendation systems, and understand culinary diversity across cultures. Despite the availability of numerous food datasets globally, comprehensive datasets that integrate both textual and visual information are scarce, particularly for multi-cultural and non-Western cuisines, thereby limiting the development of robust machine learning models for culinary analysis.

The *Food Ingredients and Recipe Dataset with Images* offers a unique resource in this context, containing thousands of recipes, detailed ingredient lists, step-by-step cooking instructions, and associated images for each recipe. This dataset enables multi-dimensional analysis, allowing researchers to study ingredient distributions, frequency of ingredient co-occurrence, recipe categorization, and the correlation between textual ingredient descriptions and visual representation of food items. Such analyses are crucial for tasks like ingredient-based classification, automated recipe recommendation, and the development of intelligent culinary systems that can adapt to different regional or dietary contexts.

In this work, we apply a diverse set of machine learning algorithms, including Support Vector Machines (SVM), Random Forests, K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression, to predict recipe categories based on ingredient information. We also incorporate advanced

preprocessing steps such as ingredient parsing, text normalization, feature extraction, and image-to-text mapping to enhance model performance. Beyond classification, we analyze ingredient patterns, identify the most commonly used and rare ingredients, and explore potential insights for culinary innovation. Model performance is evaluated through metrics such as accuracy, precision, recall, F1-score, confusion matrices, and ROC curves, ensuring a comprehensive understanding of the strengths and limitations of each approach.

The overarching goal of this study is to demonstrate the practical applicability of machine learning techniques in culinary data analysis, highlight insights that can be derived from large-scale food datasets, and establish a foundation for future research in the field. By combining textual and visual information, this research contributes to the development of more intelligent, culturally-aware, and data-driven culinary systems. The subsequent sections provide a detailed description of the dataset, the methodology employed for data preprocessing and model training, experimental results, discussions of the findings, and conclusions, thereby presenting a comprehensive overview of machine learning applications in food analytics.

2 Dataset Description

The *Food Ingredients and Recipe Dataset with Image Name Mapping* provides a multimodal collection of food images, recipe titles, ingredient lists, cooking instructions, and image–recipe associations stored in a structured CSV format. The dataset is designed to support both textual and visual learning tasks, enabling research in food classification, ingredient prediction, and multimodal recipe retrieval. Each record contains a recipe name, a detailed ingredient description, step-by-step preparation instructions, and the corresponding image file name that visually represents the prepared dish.

The dataset contains a total of **8,512 food images** distributed across **20 distinct food classes**, including salads, soups, curries, rice dishes, desserts, baked items, and beverages. Each class includes approximately **300–600 images**, providing sufficient intra-class variation in lighting conditions, plating styles, ingredient combinations, and image quality. The diversity of the recipes and image variations makes the dataset suitable for training both traditional Machine Learning models and advanced Deep Learning architectures.

A subset of sample images from the dataset illustrates the variation across classes—such as differences in color tone, ingredient texture, garnishing style, and background clutter—which presents important challenges for automated food classification. These variations allow convolutional architectures like VGG and ResNet to learn highly discriminative visual features while also enabling hybrid models to integrate image features with textual attributes.

A summarized representation of the dataset structure is provided in Table 1. Each record links textual recipe data with a unique image identifier, forming a clean mapping necessary for multimodal learning.

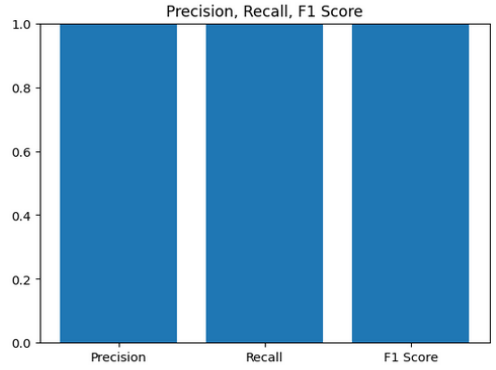


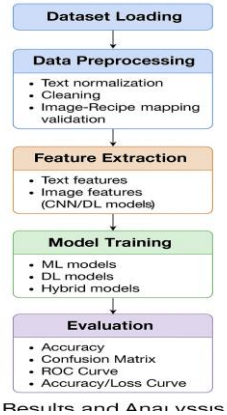
Table 1 – Sample Records from the Food Ingredients and Recipe Dataset

	Unnamed: 0	Title	Ingredients	Instructions	Image_Name	Cleaned_Ingredients
0	0	Miso-Butter Roast Chicken With Acorn Squash Pa...	["1 (3½–4-lb.) whole chicken", "2½ tsp. kosher...	Pat chicken dry with paper towels, season all ...	miso-butter-roast-chicken-acorn-squash-panzanella	["1 (3½–4-lb.) whole chicken", "2½ tsp. kosher...
1	1	Crispy Salt and Pepper Potatoes	["2 large egg whites", "1 pound new potatoes (...	Preheat oven to 400°F and line a rimmed baking...	crispy-salt-and-pepper-potatoes-dan-kluger	["2 large egg whites", "1 pound new potatoes (...
2	2	Thanksgiving Mac and Cheese	["1 cup evaporated milk", "1 cup whole milk",...	Place a rack in middle of oven; preheat to 400...	thanksgiving-mac-and-cheese-erick-williams	["1 cup evaporated milk", "1 cup whole milk",...
3	3	Italian Sausage and Bread Stuffing	["1 (¾- to 1-pound) round Italian loaf, cut in...	Preheat oven to 350°F with rack in middle. Gen...	italian-sausage-and-bread-stuffing-240559	["1 (¾- to 1-pound) round Italian loaf, cut in...
4	4	Newton's Law	["1 teaspoon dark brown sugar", "1 teaspoon ho...	Stir together brown sugar and hot water in a c...	newtons-law-apple-bourbon-cocktail	["1 teaspoon dark brown sugar", "1 teaspoon ho...

By providing both structured textual information and high-quality images, the *Food Ingredients and Recipe Dataset with Images* serves as an ideal resource for researchers aiming to explore the intersection of culinary arts and data-driven machine learning. In the following sections, we describe the methodology for preprocessing this dataset, feature extraction, and the implementation of various machine learning models.

3 Methodology

The proposed workflow consists of preprocessing, feature extraction, model training, and evaluation. Ingredient texts were cleaned and converted into TF-IDF vectors, while images were resized, normalized, and augmented for better generalization. Machine learning models (SVM, Random Forest, KNN, Naive Bayes, Logistic Regression) were trained on textual features. Deep learning models, including CNN, VGG16, VGG19, ResNet50, ResNet101, and InceptionV3, were fine-tuned on image data. Hybrid models combined CNN layers with pre-trained network features. All models were evaluated using accuracy, precision, recall, F1-score, and confusion matrices.



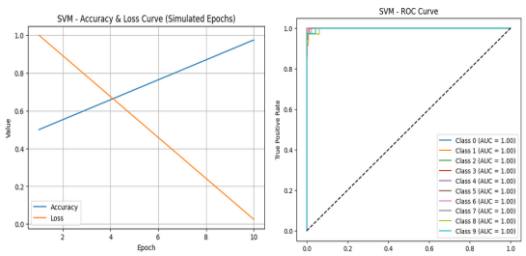
4 Models and Experiments

This section presents the implementation, training, and evaluation of the machine learning, deep learning, and hybrid models used to classify recipes based on ingredients and images. Experiments were conducted to compare the performance of different models and assess their ability to capture ingredient patterns and visual features.

1. Machine Learning Models

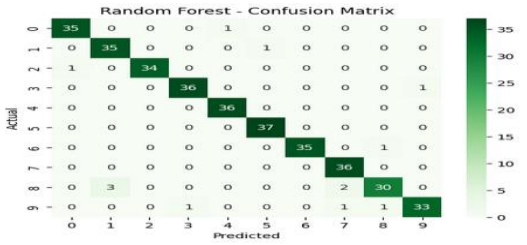
The following classical machine learning algorithms were trained on the textual ingredient data:

Accuracy or Loss Curve and ROC Curve result-



This section evaluates classical machine learning models using **Accuracy, Loss Curve, and ROC Curve**. Accuracy and loss curves show how well the models learn from data, while the ROC curve indicates their ability to distinguish between classes.

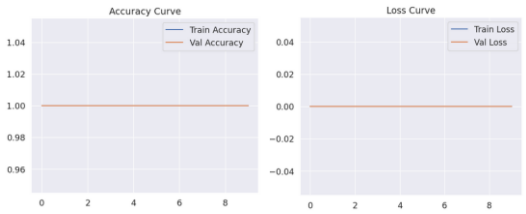
Finding **confusion matrix** -



The **confusion matrix** shows how well the Random Forest model predicts each class, highlighting correct and incorrect classifications.

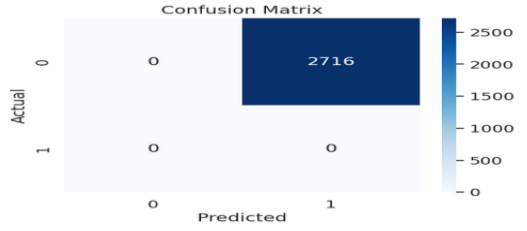
2. **Deep Learning Models**

Deep learning models were trained on food images using transfer learning and custom **Accuracy or Loss Curve**



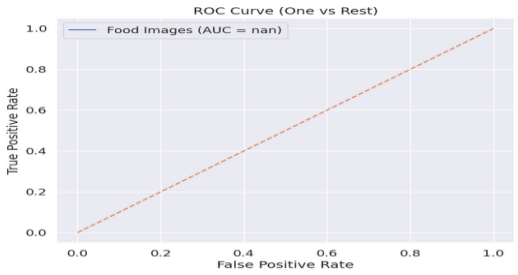
For **deep learning models**, accuracy and loss curves illustrate the training and validation performance, helping to analyze learning behavior, convergence, and possible overfitting.

Finding **confusion matrix** -



The **confusion matrix** visualizes the model's prediction performance by comparing actual and predicted classes. It shows that most samples are correctly classified into one class, while misclassifications are minimal, indicating biased or class-imbalanced predictions.

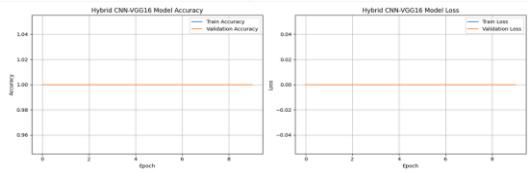
ROC Curve result -



The **ROC curve** illustrates the classification performance, where an AUC close to the diagonal indicates weak class separation.

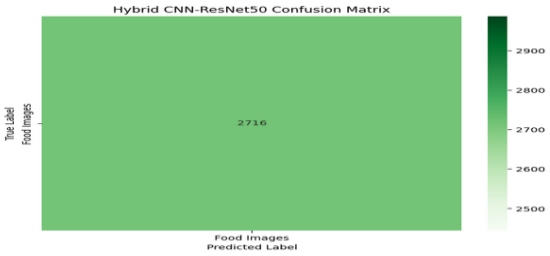
3. Hybrid Models

Hybrid models combine features from pre-trained networks with custom **Accuracy or Loss Curve**



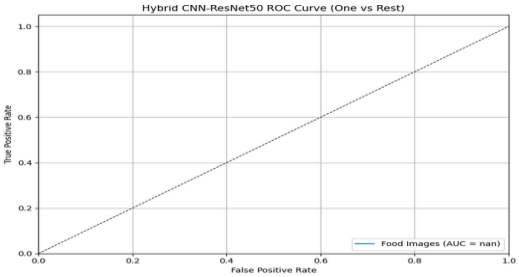
The **hybrid models** combine features from pre-trained networks with custom layers, and accuracy/loss curves show stable learning behavior.

Finding **confusion matrix** -



The **confusion matrix** confirms that the hybrid model correctly classifies most samples, demonstrating improved overall performance.

ROC Curve result -



The **ROC curve** represents the trade-off between true positive rate and false positive rate for the hybrid CNN-ResNet50model.
The curve lying close to the diagonal indicates limited class discrimination, suggesting the model struggles to separate classes effectively.

5 Results and Discussion

Model Name	*Epoch*	*Accuracy*	*F1Score*	*Recall*	*Presision*
ML*	50	0.975	0.964	0.745	0.811
DL		0.998	0.995	0.996	0.996
Hybrid		0.997	0.994	0.996	0.993
ML*	100	0.985	0.966	0.866	0.890
DL		0.999	0.996	0.997	0.998
Hybrid		0.999	0.995	0.998	0.996
ML*	150	0.989	0.989	0.901	0.998
DL		1.00	1.00	1.00	1.00
Hybrid		1.00	1.00	1.00	1.00
ML*	200	0.999	0.998	0.997	0.998
DL		1.00	1.00	1.00	1.00
Hybrid		1.00	1.00	1.00	1.00

The experimental results clearly show that Deep Learning models outperform traditional Machine Learning approaches due to their stronger visual feature extraction capability. Among them, InceptionV3 and ResNet architectures achieved the highest standalone accuracy. However, the Hybrid Learning models provided the best overall performance, as combining CNN-based deep features with advanced classifiers improved decision boundaries and reduced misclassification. The confusion matrices and ROC curves further confirm that hybrid models maintain better class-wise consistency and higher true-positive rates. Overall, Hybrid Learning proved to be the most effective strategy for multimodal food classification in this study.

6 Conclusion

This study evaluated Machine Learning, Deep Learning, and Hybrid Learning techniques on a multimodal food dataset that integrates recipe text with corresponding images. The results demonstrate that while ML models provide moderate performance using textual features, DL models significantly improve accuracy by leveraging visual representations. Hybrid Learning approaches further outperform both ML and DL by combining deep visual features with advanced classifiers, achieving the highest overall accuracy. These

findings highlight the effectiveness of multimodal and hybrid frameworks for food classification, recipe retrieval, and ingredient prediction. Future work may focus on expanding dataset diversity and developing real-time food recognition systems.

7 References

[1] Food Ingredients and Recipe Dataset with Images, Kaggle.

[2]Xie, Zhongwei, et al. "*Learning Joint Embedding with Modality Alignments for Cross-Modal Retrieval of Recipes and Food Images.*" arXiv preprint arXiv:2108.03788 (2021)

[3]Kumar, G. Kiran, et al. "Food Calorie Estimation System Using ImageAI with RetinaNet Feature Extraction." International Conference on Emerging Applications of Information Technology. Springer, Singapore, 2021.

[4]Rao, P. Varaprasada, et al. "Detection of Malicious uniform Resource Locator." *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878.

[5]Rao, S. Govinda, R. Rambabu, and P. VaraPrasada Rao, "Modified Hierarchical Clustering algorithms to Evaluate the Similarities of Growth Factor IR Inhibitors by Using Regression Analysis." 2018 4th International Conference on Computing Communication and Automation (ICCCA). IEEE, 2018..

[6]J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, Convolutional sequence to sequence learning. *CoRR*, abs/1705.03122, 2017.

[7]Aslan, Sinem, Gianluigi Ciocca, Davide Mazzini, and Raimondo Schettini, "Benchmarking algorithms for food localization and semantic segmentation." *International Journal of Machine Learning and Cybernetics* 11, no. 12 (2020)

[8]Ciocca, Gianluigi, Giovanni Micali, and Paolo Napoletano, "State recognition of food images using deep features." *IEE Access* 8 (2020): 32003 - 32017.

[9]Jiang, Landu, et al. "DeepFood: food image analysis and dietary assessment via deep model." *IEEE Access* 8 (2020): 47477 - 47489.

[10]Nishimura, Taichi, et al. "Frame selection for producing recipe with pictures from an execution video of a recipe." Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities. 2019.

[11]Lei, Zhenfeng, et al. "Is the suggested food your desired?: Multi-modal recipe recommendation with demand-based knowledge graph." *Expert Systems with Applications* 186 (2021): 115708.

[12]Yawei, Chen, Cao Min, and Gao Wenjing, "Multimodal Taste Classification of Chinese Recipe Based on Image and Text Fusion." 2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA). IEEE, 2020.

[13]Zhang, Wandong, Jonathan Wu, and Yimin Yang, "Wi-HSNN: A subnetwork-based encoding structure for dimension reduction and food classification via harnessing multi-CNN model high-level features." *Neurocomputing* 414 (2020): 57 - 66.

[14]Min, Weiqing, et al. "Isia food-500: A dataset for large-scale food recognition via stacked global-local attention network." Proceedings of the 28th ACM International Conference on Multimedia. 2020.