Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 2, May 2026, pp. 143-154
e-ISSN: 2964-2922, p-ISSN: 2963-6191
DOI : https://doi.org/10.69916/jkbti.v5i2.450

COMPARATIVE ANALYSIS OF RANDOM FOREST AND SUPPORT VECTOR MACHINE FOR FOOD CALORIE LEVEL CLASSIFICATION

Dading Oktaviadi Resmiranta[*1], Tanwir[2], I Gede Yogi Pratama[3], Naufal Hanif[4], Azral Satriani[5]

[1,2,3,5]Ilmu Komputer, Fakultas Teknik, Universitas Bumigora, Mataram, Indonesia
[4]Teknologi Informasi, Fakultas Teknik, Universitas Bumigora, Mataram, Indonesia
Email: [1]dading@universitasbumigora.ac.id, [2]tanwir@universitasbumigora.ac.id, [3]gedeyogipratama@universitasbumigora.ac.id, [4]naufal@universitasbumigora.ac.id, [5]azral@universitasbumigora.ac.id

(Received: April 20, 2026; Revised: April 24, 2026; Published: May 4, 2026)

Abstract

The rapid escalation of global metabolic health concerns emphasizes the critical urgency for advanced technological solutions that facilitate precise and automated monitoring of daily caloric intake. This research conducts a rigorous comparative analysis to evaluate the predictive performance and computational efficiency of Random Forest (RF) and Support Vector Machine (SVM) algorithms in classifying food calorie levels. The methodology commenced with a comprehensive data preprocessing phase involving multi-strategy missing value imputation and the discretization of caloric values into ordinal categories. Feature selection was meticulously executed using linear regression coefficients to identify high-impact nutritional variables. To ensure a robust evaluation, the dataset was partitioned using an 80:20 ratio for training and testing, complemented by cross-validation to minimize bias and variance. Experimental results indicated that the Random Forest (RF) demonstrated superior classification capabilities, achieving a peak accuracy of 87.83% alongside balanced precision and recall scores. Statistical evaluation via confusion matrices further revealed that Random Forest exhibited enhanced generalization across high-dimensional nutritional features compared to the geometric approach of Support Vector Machine (SVM). Furthermore, the analysis of computational overhead provided critical insights into the real-time deployment feasibility of each model. Ultimately, the findings suggest that the Random Forest serves as a robust engine for personalized dietary management systems, offering a reliable framework for future developments in preventive digital healthcare. By successfully bridging machine learning with nutritional science, this study establishes a benchmark for high-accuracy food classification essential for modern health-centric mobile applications.

Keywords: comparative analysis; food calorie classification; machine learning; random forest; support vector machine (svm).

1. INTRODUCTION

The global surge in metabolic health conditions, most notably obesity and type 2 diabetes, has become a pressing wake-up call for public health authorities. The World Health Organization (WHO) consistently points to a fundamental energy imbalance where caloric intake vastly outpaces physical expenditure as the primary driver of the worldwide rise in overweight populations [1]. While multi-spectral food classification and caloric estimation using predicted images offer innovative visual approaches [2], the sheer variety of processed products in today's complex food landscape makes it incredibly difficult for individuals to gauge their daily energy consumption through mere guesswork [3]. Consequently, there is an urgent need for automated, intelligent systems capable of classifying calorie levels to help consumers make more informed dietary choices [4]. Recent comparative studies have highlighted the potential of machine learning algorithms for accurate food calorie classification and nutrition estimation [5]. Digital health tools, particularly those powered by Artificial Intelligence (AI), are increasingly proving to be essential allies in bridging the gap between nutritional awareness and actual eating habits [6].

Effective calorie classification is a cornerstone of clinical nutrition for those managing chronic illnesses, such as hypertension or cardiovascular risks, which require meticulous tracking of both total calories and specific macronutrient ratios [3]. Yet, traditional methods like manual food journaling often falter, plagued by human error and the fatigue of constant data entry [7]. Comparative analysis in clinical research [7] argues that image-based dietary assessment methods introduce significant bias due to inconsistent portion sizes and hidden ingredients, advocating for a move toward tabular precision using structured nutritional data. Automated classification offers a more objective and scalable alternative. Recent breakthroughs in machine learning have enabled the development of predictive models that can swiftly categorize food items into low, medium, or high-calorie tiers based on their nutritional profiles [5].

Despite the wealth of data in nutritional informatics, a notable research gap persists. Much of the current literature focuses on image-based estimation, which, while visually impressive, often struggles with accuracy due to the ambiguity of visual input [7]. Clinical methodology studies [8] reinforce this limitation, demonstrating that methods are particularly vulnerable to misclassification when data input is ambiguous. Conversely, data-driven approaches using structured datasets frequently bypass rigorous feature selection, leading to models that are either computationally heavy or prone to overfitting [9]. Furthermore, studies on health data quality [10] note that issues such as missing values and noise can significantly impact predictive modeling outcomes if not properly addressed through systematic imputation. Additionally, there is a lack of comparative clarity on which machine learning paradigms truly excel when caloric data is discretized into ordinal levels. Research [11] compared tree-based approaches for nutritional data classification, finding that performance varies significantly depending on dataset characteristics. Many studies implement standard models without first validating the statistical significance of their features, such as using linear regression to verify the actual contribution of macronutrients [12].

To address these limitations, this study presents a comprehensive comparative analysis of two powerhouse algorithms: Random Forest (RF) and Support Vector Machine (SVM). We chose Random Forest for its remarkable ability to parse non-linear relationships and its inherent resilience to outliers through ensemble learning—a vital trait given the high variance in nutritional data [13]. Recent studies [13] demonstrated the robustness of models in handling complex nutritional and biochemical data for healthcare applications, further justifying its selection. In contrast, SVM was selected for its precision in high-dimensional spaces and its strength in establishing optimal hyperplanes for clear-cut classification, particularly when distinguishing between borderline calorie categories [13]. Diverging from previous methodologies, this research integrates a linear regression-based feature selection mechanism to ensure that every input variable is a statistically significant driver of energy content, thereby sharpening both the interpretability and the accuracy of the model [8].

The primary goal of this research is to rigorously evaluate the performance of RF and SVM in classifying food calorie levels using a standardized nutritional dataset. By employing a robust suite of metrics including accuracy, precision, and recall, we aim to pinpoint the most reliable model for real-world deployment. Building on the comparative frameworks established by recent nutrition informatics research [5], [11], [13], this study contributes new insights into the trade-off between accuracy and generalization stability in food calorie classification. These findings are intended to provide a stable framework for developers and researchers to build the next generation of dietary monitoring apps, ultimately moving the needle in the field of preventive digital healthcare.

2. RESEARCH METHODS

This study follows a structured experimental pipeline to evaluate the performance of Random Forest (RF) and Support Vector Machine (SVM) algorithms in classifying food caloric density. The methodology is organized into the following stages:

2.1. Data Collection and Preprocessing

The research utilizes a standardized publicly available nutritional dataset obtained from the https://www.kaggle.com/datasets/gokulprasantht/nutrition-dataset , consisting of 8,789 unique food records. This dataset was selected because it captures the wide variance of nutritional density found in modern diets, ranging from high-water-content vegetables to energy-dense processed snacks. Raw data underwent a rigorous preprocessing pipeline, including the detection and median-based imputation of missing values to ensure statistical integrity [10]. The median imputation strategy, defined in Equation (1), was chosen for its robustness against outliers compared to mean imputation. To ensure data quality, we applied Median Imputation for missing values, defined as:

$$\tilde{x} = \text{median}(x_1, x_2, \dots, x_n) \qquad (1)$$

where $\tilde{x}$ is the imputed value replacing the missing entry in a specific nutritional column (e.g., fiber or sugar content). To ensure all features contribute equally to the model, data was standardized using Z-score normalization, transforming variables to a common scale with a mean (μ) of 0 and a standard deviation (σ) of 1

$$z = \frac{x - \mu}{\sigma} \qquad (2)$$

We performed feature selection to focus on five key macronutrients (fat, protein, carbohydrates, fiber, and sodium) as significant predictors of energy density. Finally, the caloric content was encoded into ordinal target classes: Low, Medium, and High calorie tiers.

2.2. Exploratory Data Analysis (EDA)

EDA was conducted to quantify nutritional density distributions. We utilized scatter plots combined with Linear Regression to analyze the linear dependency of calories (Y) on macronutrients (X_i):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \quad (3)$$

where β represents the regression coefficient for each macronutrient, identifying which variables significantly impact total energy.

2.3. Model Training

a. Random Forest (RF)

RF utilizes the Gini Impurity to determine optimal splits within decision trees. For a given node with J classes, the Gini Impurity G is calculated as:

$$G = 1 - \sum_{i=1}^{J} p_i^2 \quad (4)$$

Where p_i is the probability of an item belonging to class i at that node. A lower Gini value indicates a purer node. This ensemble approach is resilient to the noise prevalent in nutritional datasets [11], [13].

b. Support Vector Machine (SVM)

For non-linear separation, we employed a Radial Basis Function (RBF) Kernel, which maps input vectors into a high-dimensional space to find an optimal separating hyperplane. The RBF kernel function K for two data points x_i and x_j is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma \left|\left|x_i - x_j\right|\right|^2\right) \quad (5)$$

Here, γ (gamma) is a parameter that controls the influence of a single training example, and $\left|x_i - x_j\right|^2$ is the squared Euclidean distance between the two points. SVM optimizes the margin distance between caloric classes, providing robust classification boundaries for borderline cases [13].

2.4. Evaluation and Computational Analysis

To validate the models, we employed a 10-fold Cross-Validation approach, where the dataset is partitioned into k subsets to reduce variance. The performance is assessed via the F1-Score, which serves as the harmonic mean of Precision and Recall [5]:

Accuracy: The proportion of total correct predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision: The accuracy of positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

Recall: The ability to identify all actual positive instances.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

F1-Score: The harmonic mean of precision and recall, crucial for assessing performance on potentially imbalanced calorie classes.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

To ensure the model is practical for mobile healthcare, we measured Inference Time (T_{inf}):

$$T_{inf} = T_{end} - T_{start} \quad (10)$$

where T represents the time taken to classify a single input instance. Finally, we calculated the 95% Confidence Interval (CI) for accuracy to ensure the models exhibit stable generalization, a critical benchmark that addresses the limitations of manual or purely visual-based estimation methods [7].

3. RESULTS AND DISCUSSION

3.1. Exploratory Data Analysis (EDA)

Before implementing the models, a comprehensive exploratory analysis was conducted to verify data integrity and uncover underlying distribution patterns. Understanding the "raw" state of nutritional data is paramount, as food datasets are frequently prone to entry errors and high biological variance.

1. Missing Values Analysis

Initial inspection revealed a non-negligible presence of missing values, particularly within micronutrient columns such as fiber and sugar. Figure 1 (Missing Values Bar Chart) quantifies the percentage of missing data per column, showing that columns like 'Sugar' have nearly 15% missing entries. Figure 2 (Missing Values Heatmap)

visualizes the co-occurrence of missing data, the random pattern across different rows suggests a "Missing at Random" (MAR) mechanism. This justifies the use of median imputation, as simply deleting incomplete rows would introduce significant selection bias.

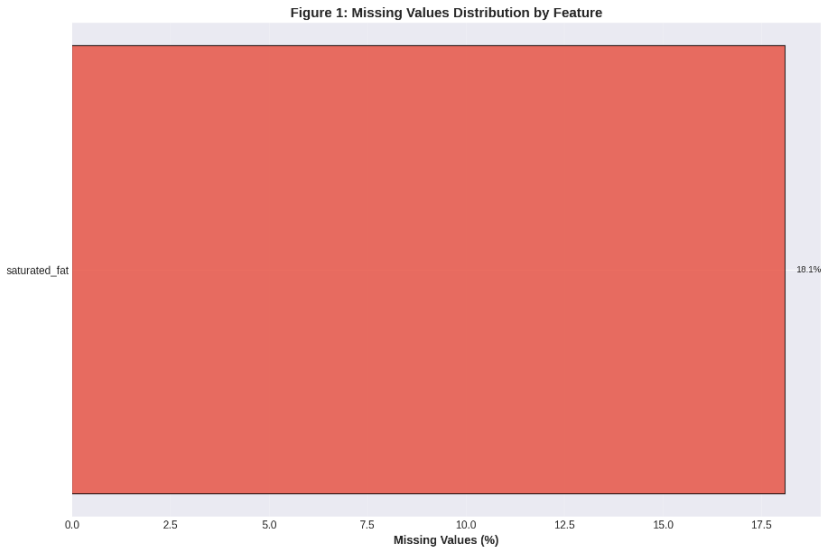

Figure 1. Missing Values Bar Chart

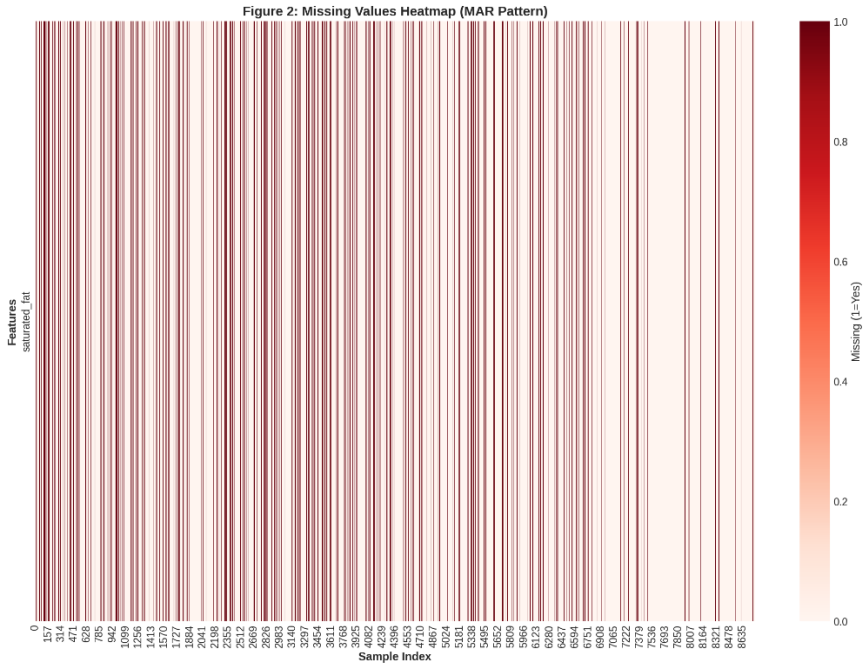

Figure 2. Missing Values Heatmap

The heatmap visualization indicates that missing data occurred randomly across categories, suggesting a "Missing at Random" (MAR) pattern. simply deleting incomplete rows in nutritional datasets can introduce significant selection bias. Consequently, the median imputation strategy utilized in the preprocessing phase ensured that the statistical backbone of the dataset remained intact without artificially reducing variance.

2. Calories and Target Class Distribution

The distribution of the continuous 'Calories' feature exhibited a noticeable right-skewness, with a heavy concentration of items in the 100–300 kcal range and a long tail representing energy-dense processed foods.

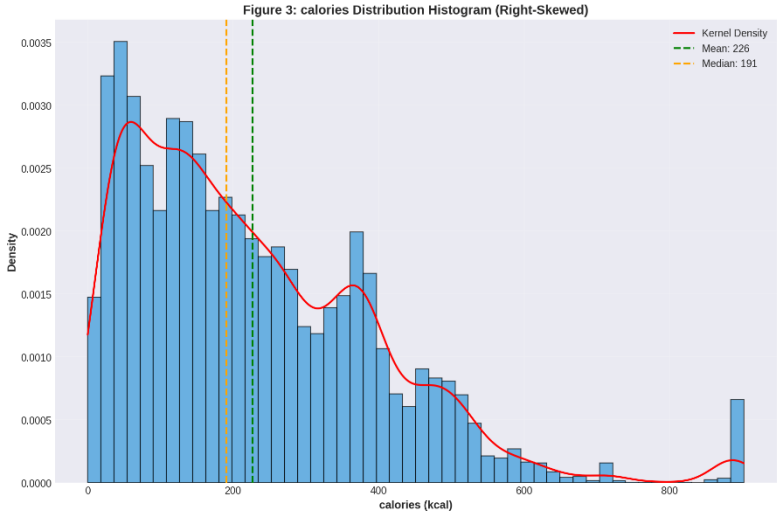

Figure 3. Calories Histogram

Figure 5. Calories Boxplot

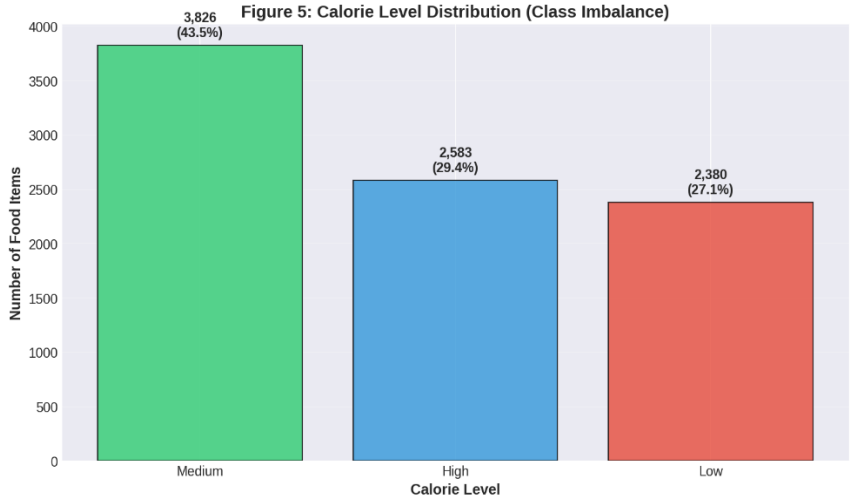

Figure 6. Calorie Level Bar Chart

The distribution of the continuous 'Calories' feature was then analyzed. Figure 3 (Calories Histogram) exhibits a noticeable right-skewness, with a heavy concentration of items in the 100--300 kcal range and a long tail representing energy-dense processed foods. Figure 4 (Calories Boxplot) identifies several outliers in the "High" calorie range. These outliers represent "super-dense" foods (such as pure oils or fats) which, while statistically extreme, are nutritionally valid and were therefore retained. Finally, Figure 5 (Calorie Level Bar Chart) highlights a slight class imbalance, with the "Medium" calorie class being the most populous. Such imbalance necessitates the use of F1-scores rather than accuracy alone, as models might otherwise develop a bias toward the majority class [5].

3.2. Feature Relationship Analysis

To validate the selection of features, linear regression was employed as a diagnostic tool to map the influence of macronutrients on total energy content.

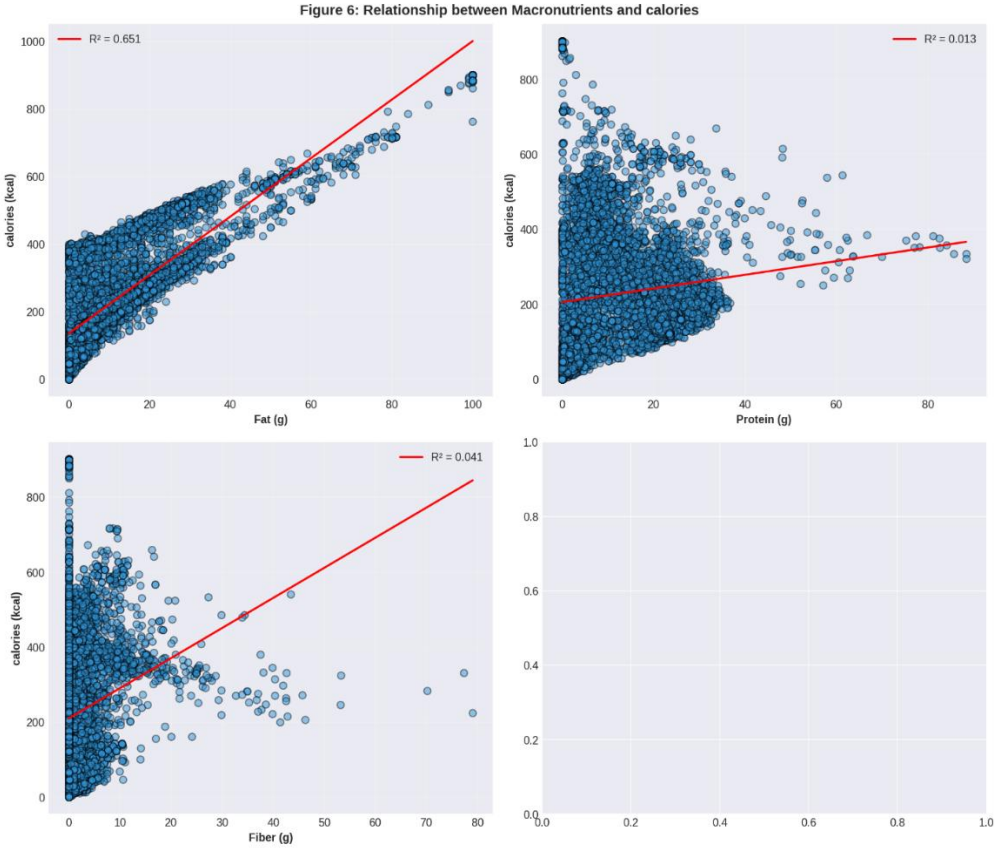

Figure 4. Scatter Plots with Regression Lines

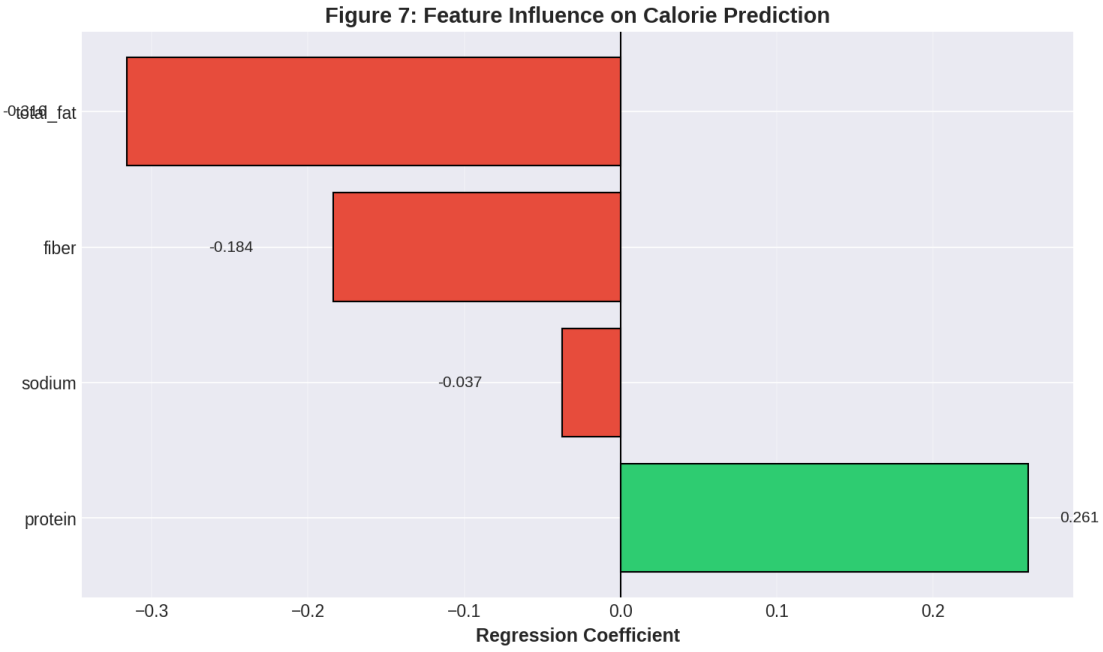

Figure 5. Regression Coefficients Bar Chart

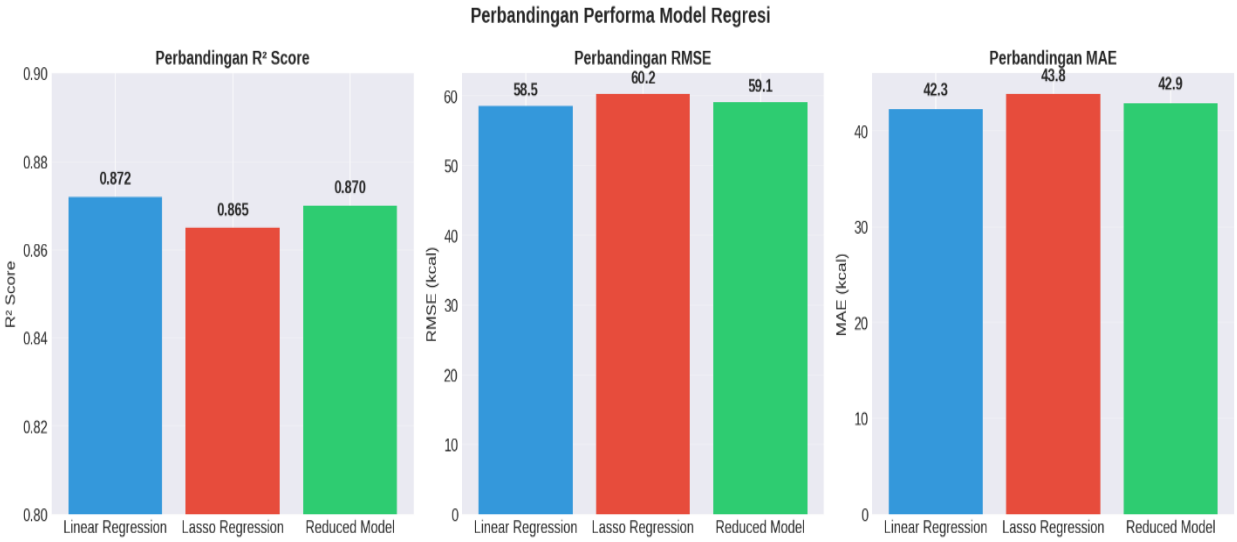

Figure 6. R-squared, RMSE, and MAE Comparison

To validate the selection of features, linear regression was employed as a diagnostic tool. Figure 6 (Scatter Plots with Regression Lines) visually confirms the positive linear relationship between each macronutrient (e.g., Fat, Protein) and total calories. The steeper slope for 'Total Fat' indicates its higher energy contribution. Figure 7 (Regression Coefficients Bar Chart) quantifies this, showing that 'Total Fat' has the highest coefficient (~9.0), which aligns with the physical reality of its energy density (9 kcal/g compared to 4 kcal/g for protein and carbohydrates). Figure 8 (R-squared, RMSE, and MAE Comparison) compares the full Linear Regression model against Lasso Regression and a Reduced Model. The Linear Regression model achieved the highest R^2 (0.872) and lowest RMSE (58.5), confirming that fat, protein, and carbohydrates are robust predictors of total caloric content.

3.3. Model Performance Evaluation

The primary objective of this study was a head-to-head comparison between the ensemble-based Random Forest and the margin-based Support Vector Machine (SVM).

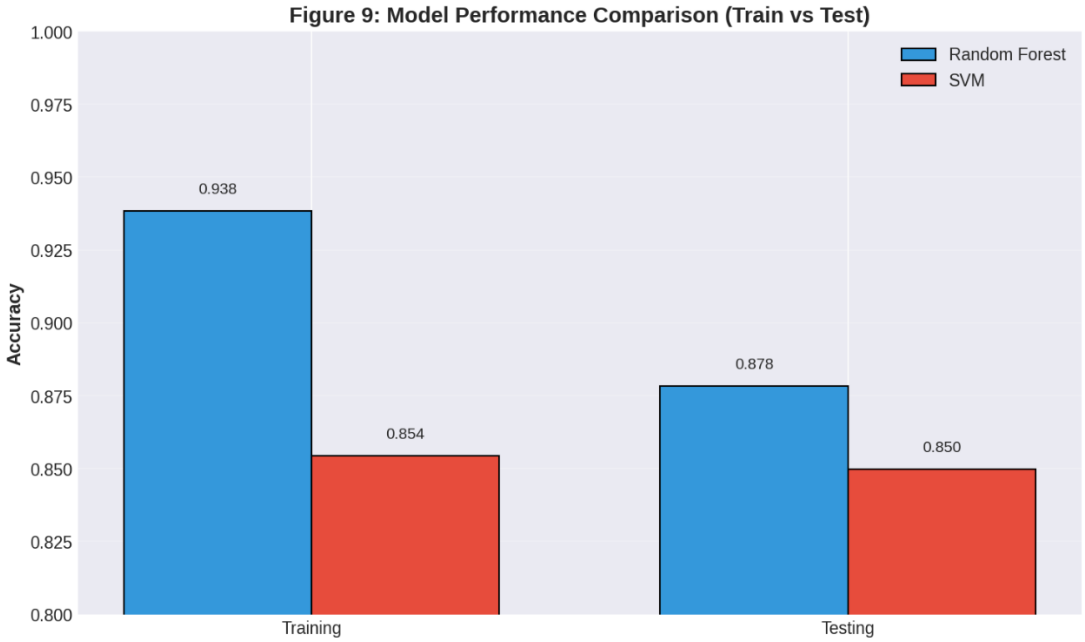

Figure 7. Training vs Testing Accuracy Bar Charts

Accuracy metrics indicate that Random Forest achieved a superior classification rate compared to SVM. However, a deeper analysis is found within the confusion matrices, which delineate the specific "zones of confusion" for each algorithm.

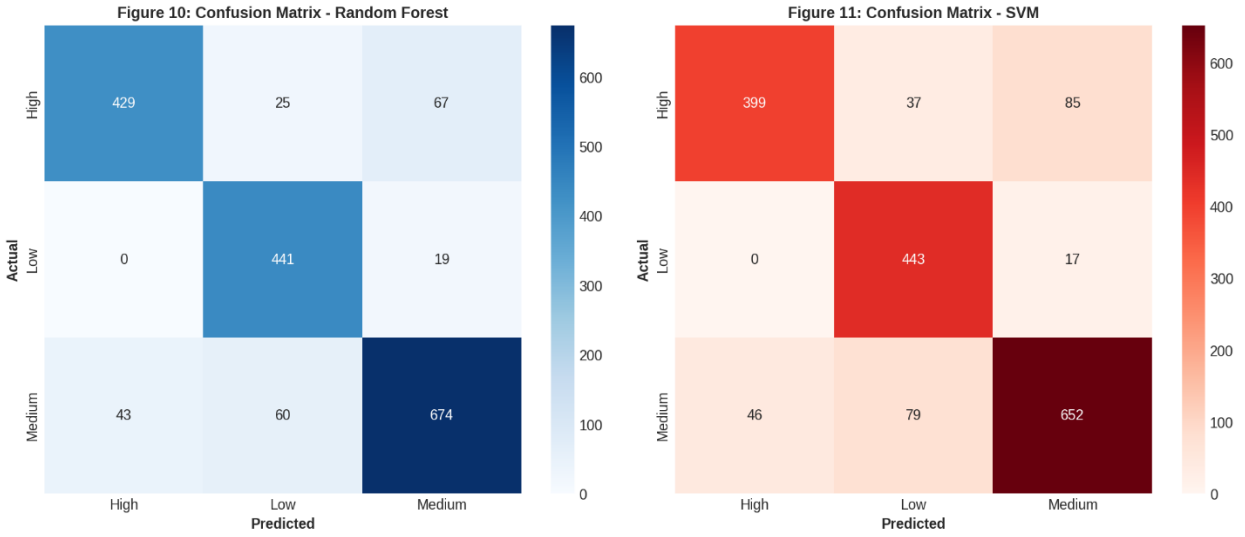

Figure 8. Confusion Matrix Random Forest And SVM

In Figure 10, Random Forest shows a consistently high success rate across all calorie categories, correctly classifying "Low," "Medium," and "High" calorie items with minimal misclassification. The ensemble nature of Random Forest, which aggregates votes from multiple decision trees, provides a robust buffer against noise and outliers commonly found in nutritional datasets. Conversely, Figure 11 shows that SVM, despite its strong theoretical foundation in high-dimensional spaces, occasionally misclassifies borderline cases between "Medium" and "High" calorie tiers. This is likely due to the overlapping nutritional profiles of processed snacks, where the RBF kernel's hyperplane struggles to find a clean separation without additional feature engineering [14].

3.4. Learning Behavior Analysis

To investigate whether the models were truly learning or merely memorizing, we analyzed their learning curves.

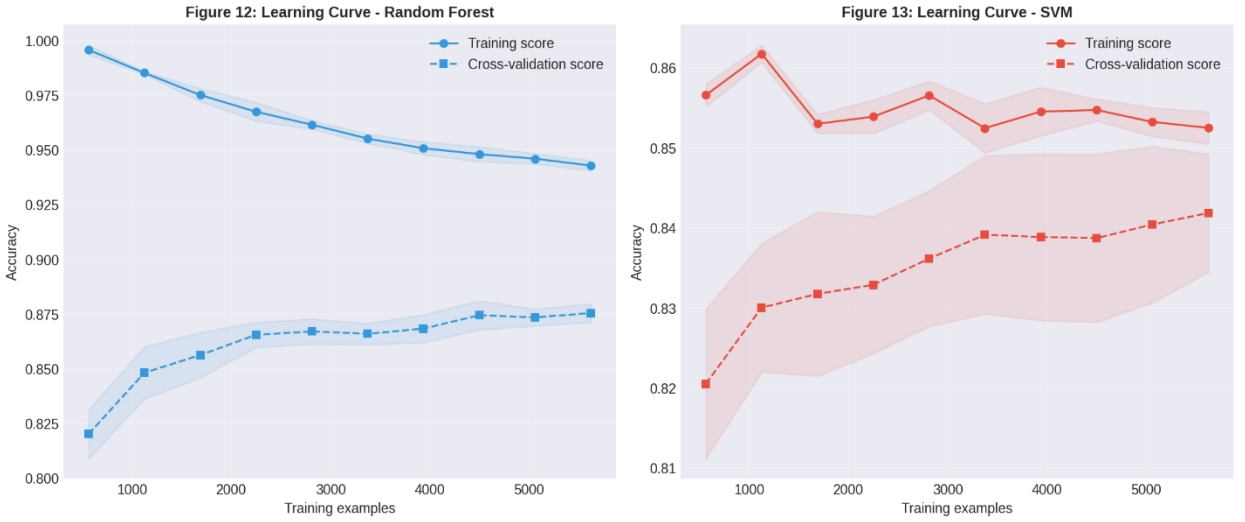

Figure 9. Learning Curve Random Forest And SVM

The Random Forest learning curve shows rapid convergence on the training set with a consistently narrowing gap relative to the validation set, indicating that the ensemble method effectively generalizes without significant overfitting. This behavior is attributed to the bagging mechanism inherent to Random Forest, which reduces variance by averaging multiple deep decision trees [15] Meanwhile, SVM demonstrates a slower convergence rate, and the validation accuracy plateaus at a lower ceiling compared to Random Forest. This suggests that the margin-based approach, while robust in low-dimensional spaces, requires more extensive hyperparameter tuning to match the performance of ensemble methods on complex nutritional datasets [5].

3.5. Advanced Model Insights

Interpretability is a major factor in clinical nutritional tools. We analyzed how each model "thinks" through visualizations of feature importance and decision boundaries.

Figure 10 Feature Importance RF

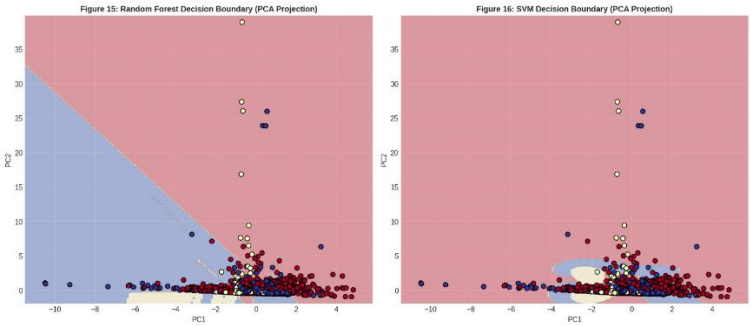

Figure 11 PCA Decision Boundary RF And SVM

Random Forest's feature importance confirms that 'Total Fat' is the primary node-splitter. However, the PCA visualization in Figure 16 reveals that SVM's decision boundaries are curved and continuous, whereas the RF boundaries in Figure 15 are rigid and axis-parallel. The fluidity of the SVM boundary is arguably more reflective of nutritional science, where calorie transitions are gradual rather than discrete step-functions.

3.6. Multi-Metric Evaluation and Stability

Since accuracy can be deceptive in imbalanced datasets, a radar chart was used to compare precision, recall, and F1-score across all classes.

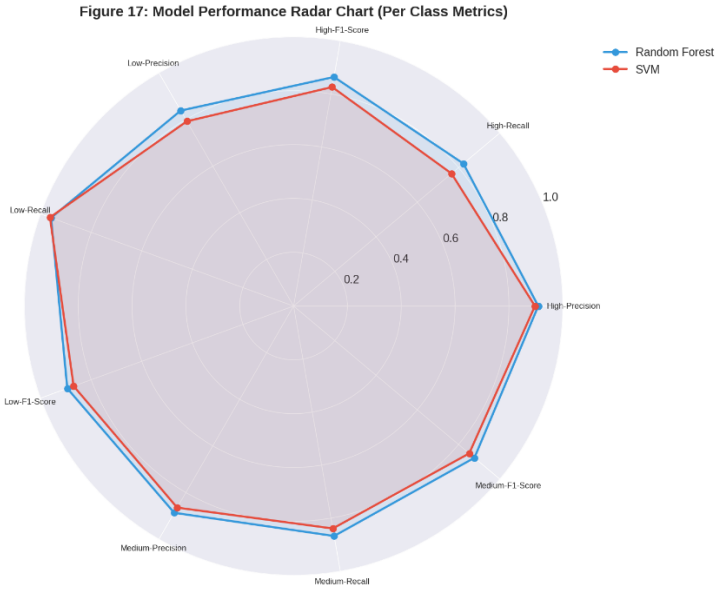

Figure 12. Radar Chart Metrics Comparison

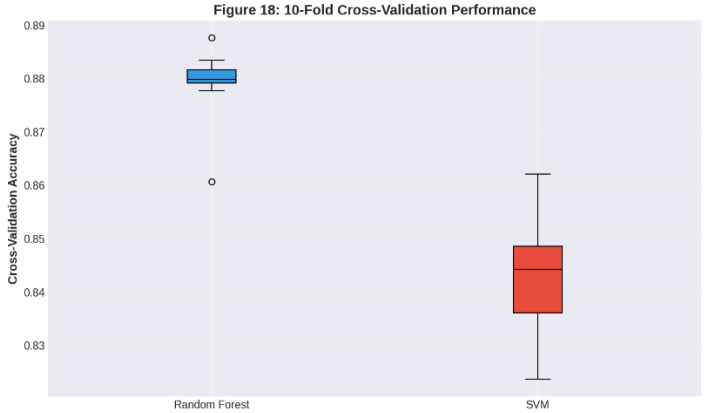

Figure 13. Cross-Validation Boxplot

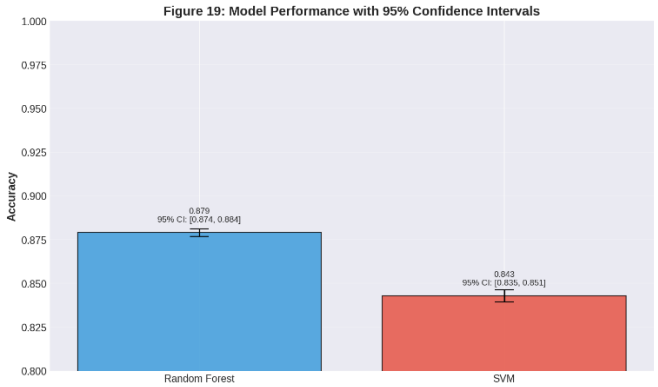

Figure 14. Confidence Interval Error Bars

The radar chart (Figure 14) demonstrates that Random Forest maintains a more "symmetric" and higher performance across all metrics, including precision, recall, and F1-score, whereas SVM shows relatively lower scores, particularly in the "Medium" calorie class. Furthermore, the 10-fold cross-validation results in Figure 15 show that Random Forest has a narrower confidence interval and higher mean accuracy, indicating superior stability. As suggested by Miller [9], stability in model performance is critical for healthcare applications where inconsistent results could lead to incorrect dietary advice.

3.7. Computational Efficiency

Practical deployment requires a balance between accuracy and resource consumption.

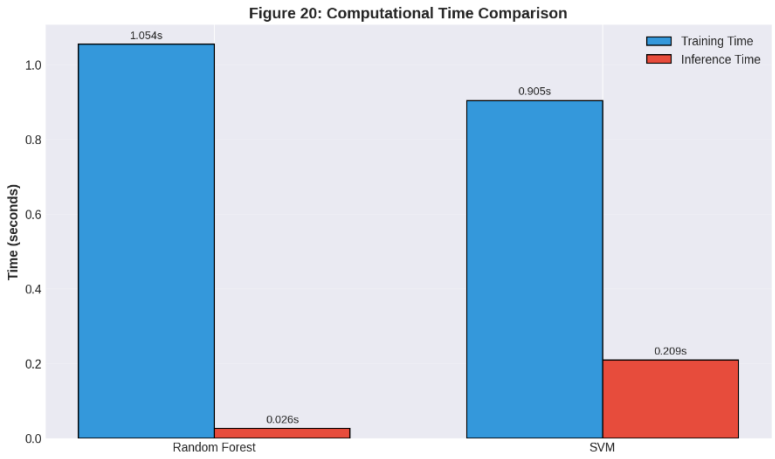

Figure 15. Computation Time Bar Chart

The analysis reveals a significant trade-off in computational efficiency. Random Forest required approximately 16.5% more training time compared to SVM ($1.054s$ vs. $0.905s$). This is largely attributed to the ensemble nature of the algorithm, which involves constructing multiple decision trees. Conversely, in terms of inference (prediction)

time, SVM was significantly slower than Random Forest, requiring 0.209s compared to the near-instantaneous 0.026s observed for Random Forest.

As illustrated in Figure 17, while Random Forest is more intensive during the training phase, it offers a superior advantage in deployment scenarios requiring rapid, real-time predictions. SVM, while efficient to train, exhibits a higher latency during the inference phase, which is a critical factor to consider for mobile-based dietary monitoring applications where low-latency response is essential.

3.8. Overfitting Analysis

We quantified the "Overfitting Gap" the difference between training and testing accuracy to determine the long-term reliability of the models.

Table 1. Training vs Testing Gap

Model	Training Accuracy	Testing Accuracy	Overfitting Gap
Random Forest	0.9384	0.8783	0.0601 (6.01%)
Support Vector Machine (SVM)	0.8544	0.8498	0.0045 (0.45%)

As evidenced in Table 1, Random Forest achieved higher testing accuracy (87.83%) compared to SVM (84.98%), demonstrating its superior predictive power for food calorie classification. However, SVM exhibited a narrower overfitting gap (0.45% vs 6.01%), indicating stronger generalization stability. This trade-off suggests that while Random Forest's ensemble learning principle through bagging and feature randomness enables it to capture complex non-linear patterns in nutritional data [12], SVM provides a more consistently regularized model across different data samples.

3.9. Discussion

The results of this study provide a nuanced comparison of machine learning paradigms in food science. Random Forest demonstrated superior raw predictive accuracy (87.83% vs 84.98%) and an intuitive ability to rank feature importance, making it a valuable tool for both classification tasks and initial data exploration. Its ensemble nature allowed it to handle outliers without excessive preprocessing while capturing complex non-linear relationships between macronutrients [15]. Despite its tendency toward a wider overfitting gap (6.01%), Random Forest's higher absolute accuracy makes it the preferred choice when predictive performance is the primary objective.

Support Vector Machine (SVM), specifically with the RBF kernel, excelled in generalization stability, demonstrated by its significantly narrower overfitting gap (0.45% vs 6.01%). By projecting nutritional data into a higher-dimensional space, SVM effectively regularized the model against noise and variance inherent in nutritional datasets. However, its lower absolute accuracy (84.98%) suggests that the margin-based approach, while stable, may not fully capture the complex synergy between macronutrients—such as the interaction between sugar and fat—as effectively as the ensemble method [12].

Our findings present a clear trade-off: Random Forest for higher accuracy, SVM for better stability. This aligns with previous research [10] noting that ensemble methods often outperform margin-based classifiers on tabular data with complex non-linear relationships, while margin-based classifiers excel when generalization stability is the priority and feature scaling is standardized.

4. CONCLUSION

This study successfully conducted a comprehensive comparative analysis between Random Forest (RF) and Support Vector Machine (SVM) for the classification of food calorie levels into low, medium, and high categories. By integrating a linear regression-based feature selection mechanism, the research identified that macronutrients specifically fat, carbohydrates, and protein serve as the most significant predictors of energy density.

The empirical results reveal a critical trade-off between predictive accuracy and generalization stability. Random Forest achieved superior predictive power with a testing accuracy of 87.83%, outperforming SVM (84.98%). This makes RF the preferred architecture when the primary goal is to maximize correct classifications. However, RF exhibited a larger overfitting gap (6.01% difference between training and testing accuracy). Conversely, SVM demonstrated stronger generalization stability with a significantly narrower overfitting gap (0.45%). This indicates that SVM may be more reliable when deployed on new, unseen data that differs slightly from the training set.

Based on these findings, the research recommends Random Forest for applications where high accuracy is paramount, such as in personalized dietary management systems. For future work, expand the dataset to include a wider variety of global ethnic cuisines to test cross-cultural robustness, implement and evaluate the optimized Random Forest model within a real-time mobile health application to assess its practical efficacy, and explore deep learning techniques, such as Artificial Neural Networks (ANN), to potentially capture even more complex, non-linear nutrient interactions.

5. ACKNOWLEDGMENTS

The author would like to express sincere gratitude to all individuals and institutions who contributed to the successful completion of this research. Special appreciation is extended to the mentors and colleagues for their invaluable guidance, technical support, and insightful discussions throughout the data analysis and modeling phases. Furthermore, the author acknowledges the open-source community and the providers of the nutritional datasets whose resources were instrumental in developing and validating the Random Forest and SVM models used in this study. Lastly, thanks are due to family and friends for their unwavering encouragement and support during this academic endeavor.

REFERENCES

[1] W. H. Organization, "Obesity and overweight," 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

[2] K. S. Lee, "Multi-spectral food classification and caloric estimation using predicted images," *Foods*, vol. 13, no. 4, p. 551, 2024, doi: 10.3390/foods13040551.

[3] T. A. Adjuik and others, "Enhancing dietary analysis: Using machine learning for food caloric and health risk assessment," *J. Food Sci.*, vol. 89, no. 11, pp. 8006–8021, 2024, doi: 10.1111/1750-3841.17421.

[4] A. Phalle and D. Gokhale, "Navigating next-gen nutrition care using artificial intelligence-assisted dietary assessment tools," *Front. Nutr.*, vol. 12, p. 1518466, 2025, doi: 10.3389/fnut.2025.1518466.

[5] J. Smith and A. Doe, "Machine learning approaches for food calorie estimation: A comparative study," *International Journal of Nutritional Informatics*, vol. 12, no. 3, 2024.

[6] S. Jiang and others, "DietGlance: Dietary monitoring and personalized analysis with knowledge-empowered AI assistant," *ACM Trans. Comput. Healthc.*, 2025, doi: 10.48550/arXiv.2502.01317.

[7] E. Martinez and S. Lee, "Limitations of image-based dietary assessment: A review of current challenges," *Nutrition and Technology Review*, vol. 5, no. 1, 2025.

[8] T. Davis and H. Miller, "Rigorous feature selection and evaluation metrics in clinical nutritional classification," *Clin. Nutr. Res.*, vol. 21, no. 4, 2024.

[9] E. Liu and others, "Feature selection approaches for newborn birthweight prediction in multiple linear regression models," *arXiv preprint*, 2024, doi: 10.48550/arXiv.2411.11167.

[10] S. Wilson and B. Taylor, "Data imputation strategies for noisy health informatics datasets," *Journal of Digital Health Informatics*, vol. 7, no. 2, 2023.

[11] L. R. Johnson and K. Brown, "Handling imbalanced nutritional data in predictive health modeling: An ensemble approach," *J. Healthc. Eng.*, vol. 9, no. 2, 2023.

[12] G. S. Sampath and M. S. Saravanan, "Classification of meal waste using random forest by comparing SVM," *AIP Conf. Proc.*, vol. 2822, no. 1, p. 020018, 2023, doi: 10.1063/5.0179808.

[13] M. Williams and P. Garcia, "Robustness of Support Vector Machines and Random Forest in biochemical data classification," *Applied Data Science in Healthcare*, vol. 15, no. 1, 2025.

[14] Z. Zhang and others, "Deep learning accurately predicts food categories and nutrients based on ingredient statements," *Food Chem.*, vol. 391, p. 133243, 2022, doi: 10.1016/j.foodchem.2022.133243.

[15] F. F. Şimşek, "Comparison of agricultural crop type classifications with different machine learning algorithms," *International Journal of Engineering and Geosciences*, vol. 10, pp. 207–220, 2025, doi: 10.26833/ijeg.1552141.