Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1136~1141
e-ISSN: 2548-6861

Multilabel Machine Learning-Based Detection of Allergens in Food Recipes

Ratih Nur Esti Anggraini [1]*, Ahmad Hafizh Assa'ad[2]*, Shintami Chusnul Hidayati [3]*
* Teknik Informatika, Institut Teknologi Sepuluh Nopember
ratih@its.ac.id [1], 5025201106@student.its.ac.id [2], shintami@its.ac.id [3]

Article Info

Article history:

Received 2026-02-16
Revised 2026-03-24
Accepted 2026-04-10

Keyword:

Food Allergens,
KNN,
Machine Learning,
MLP,
SVM.

ABSTRACT

Food allergens are substances that can trigger allergic reactions or intolerances in some individuals. Recent data indicate that the global prevalence of food allergies is approximately 8%, with a comparable proportion observed in Indonesia. This research focuses on developing a machine learning model to detect allergens in food recipes, utilizing K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) methods with a multilabel classification approach. The primary challenge is the difficulty of identifying hidden allergens in the diverse ingredients of recipes, which can be harmful to individuals with food allergies. This study utilizes 15,823 data points from a food recipe dataset, labeled manually by annotator with five main types of allergens. After data preprocessing and feature extraction using TF-IDF, the models were trained and tested with an 80:20 ratio. Results indicate that the SVM with hyperparameter tuning on the manually labeled dataset performed the best across all allergen types, achieving average F1-Scores of 0,9776.

I. Introduction

Food allergy is an abnormal immune response that occurs when the body mistakenly identifies certain food proteins as harmful substances [1]. This condition can trigger a wide range of symptoms, ranging from mild reactions such as skin irritation to severe and potentially life-threatening anaphylaxis. Food allergies represent a growing global health concern, affecting both children and adults, and requiring careful dietary management to prevent adverse reactions. Recent studies have demonstrated the potential of computational approaches to predict allergy risk by analyzing clinical and demographic data, enabling early detection and preventive healthcare strategies [2,3]. Furthermore, epidemiological research indicates that the prevalence of food allergies has increased significantly over the past decade. The available data suggest that global food allergy prevalence is approximately 10% [4, 5]. A similar prevalence has been reported [6]. This condition is emphasizing the need for improved detection and monitoring mechanisms [7].

In addition, the increasing adoption of online food delivery platforms and digital dietary recommendation systems has intensified the need for reliable allergen identification. Individuals with food allergies depend on accurate ingredient information when selecting meals through these platforms, yet menu descriptions are often incomplete, inconsistent, or expressed in natural language. Automated detection of allergens from textual food information can therefore play a crucial role in enhancing food safety and supporting personalized dietary guidance, particularly in intelligent nutrition recommendation systems [8]. Such capabilities can reduce health risks associated with accidental allergen exposure and improve user trust in digital food services.

Several types of foods are widely recognized as major allergens, including milk, nuts, eggs, seafood, and wheat. Previous literature reviews highlight that these allergenic foods account for a substantial proportion of allergy cases worldwide and are subject to strict food labeling regulations to protect consumers [9]. Studies on allergenic protein characterization also show that variations in protein structure and food processing methods can influence allergenicity, making detection more complex in processed or composite food products [10]. Consequently, accurate identification of allergenic ingredients remains a critical requirement for food safety and public health protection.

Several approaches have been developed to detect allergens and monitor allergy reactions. Computer vision-based systems utilizing deep learning techniques, such as convolutional neural networks, have been proposed to detect allergenic food items from food images, demonstrating promising accuracy in identifying allergen-containing foods visually [11]. In clinical settings, ensemble learning methods have been applied to predict outcomes of oral food challenges by analyzing patient medical records, laboratory results, and clinical history [12]. Additionally, deep learning models such as Long Short-Term Memory (LSTM) networks have been employed to analyze longitudinal microbiome data to predict food allergy development, demonstrating the importance of temporal biological patterns in allergy progression [13]. Other studies have explored natural language processing and machine learning techniques for food safety monitoring, including automated extraction of ingredient information from textual food descriptions and labeling systems [14, 15]. These approaches highlight the expanding role of artificial intelligence in allergy detection across multiple data modalities.

Limited studies have explored allergen detection from textual data, particularly food recipes, which are widely used by individuals when preparing or selecting meals. Food recipes contain both structured and unstructured textual information describing ingredients and preparation methods that may implicitly indicate allergen presence. Prior research in food text mining has primarily focused on recipe recommendation, nutritional analysis, and ingredient substitution, with less attention given to allergen detection [16]. This indicates a research gap in utilizing textual recipe data for automated allergen identification, particularly for regional and culturally specific cuisine datasets.

In contrast to previous studies, this research investigates the effectiveness of traditional machine learning models combined with TF-IDF features for multilabel allergen detection from food recipe texts. This research proposes a machine learning-based approach to detect allergens in food recipes using K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). Previous comparative studies have shown that traditional machine learning classifiers remain competitive in text classification tasks due to their interpretability and efficiency when combined with feature extraction techniques such as TF-IDF [17-19].

II. METHODOLOGY

This study focuses on evaluating three machine learning algorithms applied during the modelling stage. The performance of KNN, SVM, and MLP is compared to identify the most effective method for representing allergen-related topics. As illustrated in Figure 1, the research begins with data collection by obtaining Indonesian food recipe from Kaggle. Subsequently, the dataset undergoes preprocessing, which includes removing stopword, punctuation and digit removal, tokenization, case folding, stopword removal, stemming, and unique word verification.

After preprocessing, feature extraction is performed using the TF-IDF method. The processed data is then splitted into data train and test. Next, we train three classifiers, namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) algorithms. Finally, the modelling performance is assessed and analyzed using a confusion matrix.

A. Data Collection

The first step was collecting data of food recipes from https://www.kaggle.com/code/dionisiusdh/indonesian-food-recipe-data-visualization/ where 15.823 rows of Indonesian food recipes was collected. The dataset was subsequently annotated by identifying the presence of five specified allergens: milk, nuts, eggs, seafood, and wheat. The allergen labeling process was conducted manually by trained annotators, who assigned a binary label to each data entry, where a value of 1 indicated the presence of the respective allergen and 0 indicated its absence. Table 1 illustrate the dataset with multilabel allergens and Figure 2 shows the number of allergens in the dataset. Based the illustration, we can conclude that most of the dataset contains seafood allergen but less nut allergen.

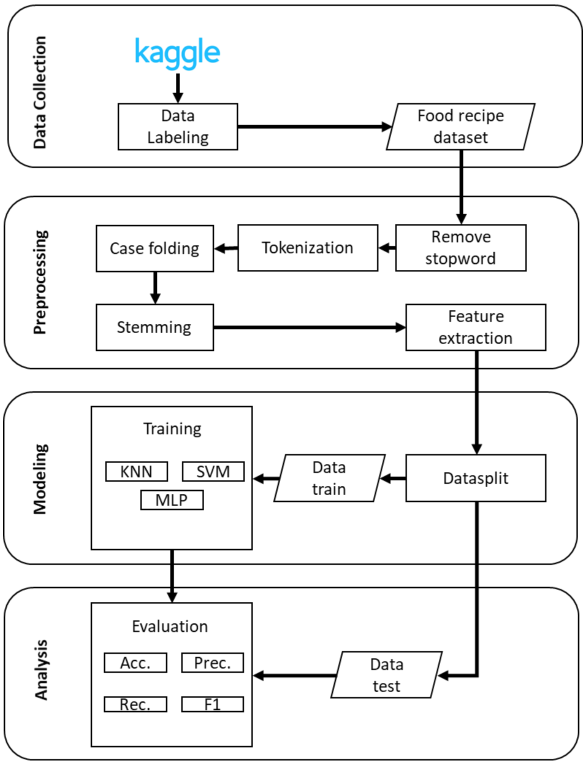

Figure 1. Proposed Methodology

TABLE 1.
DATASET SAMPLE

Title	Ingredients	Milk	Nut	Egg	Sfood	Wheat
Rolade Sehat **(*Healthy Rolade*)**	100 gr dada ayam 4 besar telur 1 sdt minyak wijen 200 gr wortel 1/2 sdm minyak goreng 130 gram tahu 50 ml susu uht full cream 40 g rolled oat 15 ml kecap asin (*Translation*: *100g chicken breast* *4 large eggs* *1 tsp sesame oil* *200g carrots* *1/2 tbsp cooking oil* *130g tofu* *50ml full-cream UHT milk* *40g rolled oats* *15ml soy sauce*)	1	0	1	0	1

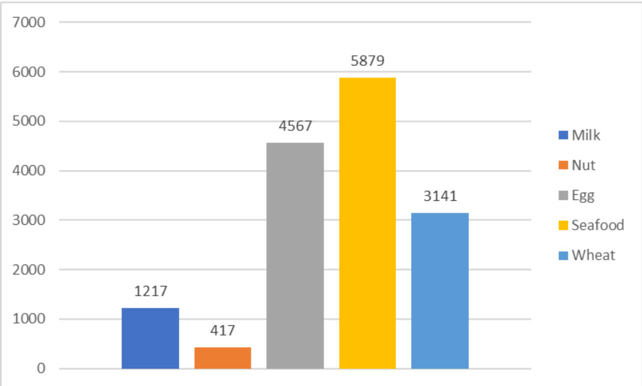

Figure 2. Number of allergens in dataset

B. Preprocessing

The next phase is text preprocessing which includes remove stopword, tokenization, case folding and stemming. For this research, a list of stopword was created as shown in Table 2. In addition, the numbers and punctuation were also removed. Subsequently, TF-IDF was implemented to extract the feature.

TABLE 2.
LIST OF STOPWORD

Stopword List				
Butir	Kg	Gram	Pada	Serta
Kecil	dengan	Untuk	Dan	Hingga
Sedang	ke	di	atau	dengan
besar	Seperti	Siung	lembar	secukupnya

C. Modeling

The preprocessed dataset then undergone data split (80:20) where 12.620 rows of data will be used as data train. Three machine learning classifiers, namely KNN, SVM and MLP were trained using the data train.

The K-Nearest Neighbors (KNN) method used to detect allergens in a multilabel dataset of food recipes is a non-parametric approach that utilizes similarities between data to perform classification. In this context, each food recipe is analyzed to determine the presence of multiple allergens simultaneously, making classification a multi-output task. The process begins by determining the optimal number of nearest neighbors (K) parameter through a grid search [20]. Each food recipe in the dataset is represented as a feature vector, and the distance between the unknown recipe vector and the recipe in the dataset is calculated using a specific metric, such as Cosine distance. The unknown recipe is then classified based on the majority of allergen classes present in its K nearest neighbors.

In this implementation, SVM is adapted to handle multi-output tasks by using scikit-learn's "MultiOutputClassifier", which allows simultaneous classification of multiple allergen labels. Each food recipe in the dataset is represented as a feature vector, and the SVM constructs a hyperplane in high feature space to separate different allergen classes [21]. The training process involves finding the optimal hyperplane that maximizes the margin between the classes of allergen data.

MLP consists of an input layer, multiple hidden layers, and an output layer. Each layer consists of a number of neurons that are connected to the neurons in the next layer through weights that are adjusted during the training process. In this context, each food recipe in the dataset is represented as a feature vector, and the MLP processes this vector through layers of neurons to generate multi-label predictions [22]. Each neuron in the output layer represents one allergen label, allowing the MLP to predict the presence of multiple allergens simultaneously.

D. Analysis

The final phase is evaluating the result using confusion matrix. To optimize the result, grid search was employed to find the best parameter for each classifier. To evaluate the result, we use six scenarios by comparing three classifiers with and without grid search parameter tuning, as shown in Table 3.

TABLE 3.
EXPERIMENT SCENARIO

No	Method	Model
1		KNN
2	Non GridSearchCV	SVM
3		MLP
4		KNN
5	GridSearchCV	SVM
6		MLP

III. RESULT AND DISCUSSIONS

A. Experiment Results

The experiment was conducted based on the scenarios in Table 3. First, three classifiers were trained to predict the allergen in the food ingredient of food recipes without implementing hyperparameter tuning. The result was shown in Table 4. It indicates that SVM has the best performance in predicting food allergens from food recipes with overall metrics showed above 95%.

TABLE 4.
RESULT WITHOUT HYPERPARAMETER TUNING

Model	Metric	Food Allergen				
		Milk	Nut	Egg	Seafood	Wheat
KNN	*Acc*	0,9496	0,9829	0,8624	0,8609	0,9379
	Prec	0,9310	0,9118	0,8190	0,8793	0,8995
	Rec	0,4170	0,3780	0,6853	0,7393	0,7773
	F1	0,5760	0,5345	0,7462	0,8032	0,8339
SVM	*Acc*	**0,9965**	**0,9940**	**0,9975**	**0,9823**	**0,9940**
	Prec	**0,9921**	**0,9846**	**1,0000**	**0,9874**	**1,0000**
	Rec	**0,9653**	**0,7805**	**0,9914**	**0,9662**	**0,9700**
	F1	**0,9785**	**0,8707**	**0,9957**	**0,9766**	**0,9848**
MLP	*Acc*	0,9927	0,9924	0,9932	0,9756	0,9911
	Prec	0,9876	0,9677	0,9945	0,9781	0,9871
	Rec	0,9228	0,7317	0,9785	0,9579	0,9684
	F1	0,9541	0,8333	0,9865	0,9679	0,9777

Next, we implemented grid search for hyperparameter tuning in order to get the best parameter configuration. The best configuration was in Table 5.

TABLE 5.
HYPERPARAMETER FROM GRID SEARCH

Model	Hyperparameter	Value	Result
KNN	K	[5, 10, 15, 20, 25]	20
	Metric	['uniform', 'distance']	distance
	Weights	['cosine','euclidian','manhattan']	cosine
SVM	C	[0.1, 1, 10]	10
	Gamma	[1, 0.5, 0.1, 0.01]	1
	Kernel	[rbf', 'linear']	linear
MLP	Hidden Layer Sizes	[50, 100, 150]	(150)
	Activation	['relu', 'tanh']	tanh
	Solver	['adam', 'sgd']	adam
	Alpha	[0.0001, 0.05]	0.05
	Learning Rate	['constant', 'adaptive']	adaptive

Table 6 shown the results of evaluating the model on each food allergens. SVM still outperformed other classifiers. In addition, hyperparameter tuning can increase the results of recall and F1-score specifically to predict minor class; nut allergen in this case.

TABLE 6.
RESULT WITH HYPERPARAMETER TUNING

Model	Metric	Food Allergen				
		Milk	Nut	Egg	Seafood	Wheat
KNN	*Acc*	0,3243	0,3659	0,7057	0,7624	0,8073
	Prec	0,4870	0,5357	0,7920	0,8370	0,8668
	Rec	0,9981	0,9962	0,9997	0,9883	0,9968
	F1	0,9922	0,9730	1,0000	0,9859	0,9968
SVM	*Acc*	**0,9846**	**0,8780**	**0,9989**	**0,9835**	**0,9874**
	Prec	**0,9884**	**0,9231**	**0,9995**	**0,9847**	**0,9921**
	Rec	**0,9965**	**0,9946**	**0,9987**	**0,9832**	**0,9959**
	F1	**0,9921**	**0,9848**	**1,0000**	**0,9883**	**0,9936**
MLP	*Acc*	0,9730	0,7927	0,9957	0,9728	0,9874
	Prec	0,9825	0,8784	0,9978	0,9805	0,9905
	Rec	0,3243	0,3659	0,7057	0,7624	0,8073
	F1	0,4870	0,5357	0,7920	0,8370	0,8668

B. Discussions

Based on Table 7 the results of the experiment scenarios show that scenario 5 has the highest values in several key metrics. The highest accuracy of 0.9958 was achieved in scenario 5, which used the GridSearchCV method with an SVM model. This shows that extensive parameter optimization through GridSearchCV is able to produce highly accurate models. The highest precision of 0.9830 was achieved in scenario 2, which used the SVM model without GridSearchCV. It indicates that SVM is the best classifiers for the case study, with or without hyperparameter tuning.

TABLE 7.
SCENARIO RESULTS

Scenario	Average			
	Accuracy	Recall	Precision	F1-score
1	0,9187	0,8881	0,5994	0,6988
2	0,9929	**0,9928**	0,9347	0,9613
3	0,9890	0,9830	0,9119	0,9439
4	0,9308	0,9486	0,5931	0,7037
5	**0,9958**	0,9896	**0,9665**	**0,9776**
6	0,9938	0,9918	0,9443	0,9659

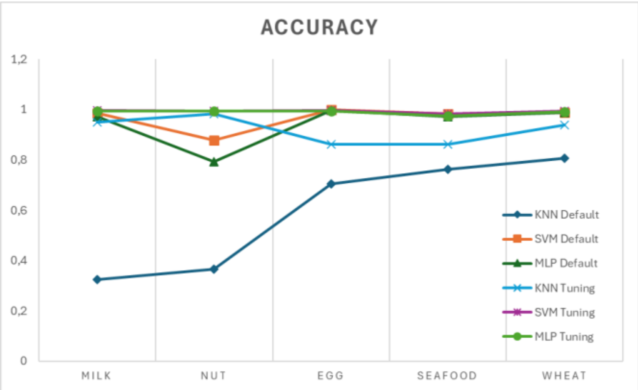

Figure 3. Accuracy comparison

Figure 3 illustrates the comparison between model's accuracy. The result indicates that SVM outperformed other models. The hyperparameter tuning was not given a significant improvement on SVM performance. On the other hand, though KNN shows the worst performance, hyperparameter tuning does increase the performance.

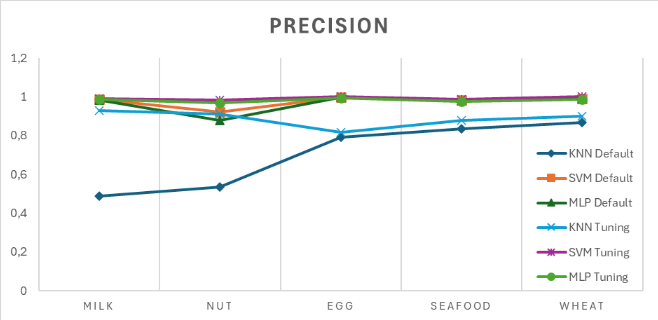

Figure 4. Precision comparison

The model performance through precision comparison in Figure 4 amplify the previous statement that SVM is the best model for predicting allergen in food recipes. Hyperparameter tuning through grid search affecting classifiers with lower performance.

Based Table 4, Figure 3 and Figure 4, there are substantial variations in detection difficulty across allergen categories. Among all labels, Nut is consistently the most challenging to detect across the three models, as indicated by notably lower recall and F1-scores, precision, and recall, suggesting a higher rate of false negatives. In contrast, Egg and Wheat achieve very high performance, indicating that their feature patterns are more distinctive and consistently captured by the models. Milk and Seafood show intermediate difficulty, although Milk exhibits reduced recall under KNN. These differences are likely influenced by label distribution, variability in textual representation, and potential feature overlap among allergens. Overall, this per-label evaluation provides meaningful insight into class-specific errors and identifies nut allergens as the most difficult category to detect, even without an explicit confusion matrix.

Overall, the use of TF-IDF combined with machine learning model was able to extract important allergen information from the text representation of food recipes. It is reflected by the high accuracy across most experimental scenarios. However, the observed variations in precision, recall, and F1-score—particularly in scenarios where accuracy remains high, but precision is relatively low—highlight the inherent limitations of TF-IDF in capturing semantic context. Because TF-IDF relies on surface-level term frequencies, it cannot model semantic equivalence or contextual meaning, making it sensitive to lexical variation such as synonyms, alternative ingredient names, and ambiguous culinary terms. Different expressions referring to the same ingredient (e.g., "groundnuts" vs. "peanuts") may be treated as unrelated features, while context-dependent terms or implicit references may not be recognized as allergen indicators. Such linguistic variability is common in food descriptions and can lead to misclassification despite overall high accuracy. Addressing these challenges would likely require semantic-aware representations, such as word embeddings, domain ontologies, or contextual language models, which could improve robustness in real-world applications but fall outside the scope of the present study.

To address the limitations of TF-IDF in capturing semantic relationships, future research may explore more advanced text representations such as word embeddings [23, 24] and pre-trained language models [25, 26]. Unlike TF-IDF, these approaches can encode contextual meaning, semantic similarity, and linguistic nuances, enabling the model to better handle synonyms, ambiguous culinary terminology, and implicit ingredient references commonly found in recipe texts. Such context-aware representations have the potential to improve detection accuracy and robustness, particularly for complex or linguistically diverse descriptions. However, these methods typically require greater computational resources and larger datasets and therefore were beyond the scope of the present study.

IV. Conclusions

In this research, we predict the five major allergens in the food ingredient of food recipes. The investigation was conducted on three classifiers namely KNN, SVM and MLP both with and without hyperparameter tuning. The result shows that SVM outperforms other classifiers even without hyperparameter tuning with all metrics shows the performance above 0.98. On the other hand, KNN shows poor performance based on its recall especially in minority class such as milk and nut. However, hyperparameter tuning helps to improve its performance. It indicates that KNN is more sensitive to data imbalance, which results in decreased performance, especially in recall.

Future work will include a detailed evaluation of computational complexity, runtime efficiency, Subset accuracy and Hamming Loss, as this aspect was not investigated in the present study. We also will investigate multilabel stratified sampling or cross validation as alternative validation schemes to assess potential effects on performance stability, since the present work used a random hold-out split given the challenges of maintaining label distribution across multiple labels. Moreover, the focus can be shifted to addressing the imbalance data to improve model's performance. In addition, the implementation of multimodality can be implemented by combining text data from ingredients and image of the food recipes.

References

[1] Q. Liu, S. Lin, and N. Sun, "How does food matrix components affect food allergies, food allergens and the detection of food allergens? A systematic review," Trends in Food Science & Technology, vol. 127, pp. 280–290, 2022, doi: 10.1016/j.tifs.2022.07.009

[2] K. Kamphorst, A. Lopez-Rincon, A. M. Vlieger, J. Garssen, E. van 't Riet, and R. M. van Elburg, "Predictive factors for allergy at 4–6 years of age based on machine learning: A pilot study," PharmaNutrition, vol. 23, 2023, doi: 10.1016/j.phanu.2022.100326.

[3] S. H. Sicherer and H. A. Sampson, “Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment,” Journal of Allergy and Clinical Immunology, vol. 141, no. 1, pp. 41–58, 2018, doi: 10.1016/j.jaci.2017.11.003.

[4] R. L. Warren, J. M. Dyer, and R. S. Gupta, “Prevalence and characteristics of food allergy in the United States,” Journal of Allergy and Clinical Immunology, vol. 142, no. 2, pp. 394–403.e11, Aug. 2018.

[5] A. A. Loh and K. Tang, “Food allergy worldwide: Epidemiology, mechanisms, and prevention,” World Allergy Organization Journal, vol. 11, no. 1, 2018.

[6] Universitas Gadjah Mada, “Childhood allergy cases rising in Indonesia,” Oct. 2025. [Online]. Available: https://ugm.ac.id/en/news/childhood-allergy-cases-rising-in-indonesia-ugm-pediatrician-stresses-accurate-diagnosis-and-prevention/

[7] R. S. Gupta et al., “The public health impact of parent-reported childhood food allergies in the United States,” Pediatrics, vol. 142, no. 6, 2018, doi: 10.1542/peds.2018-1235.

[8] R. Zhou, J. Wang, Y. Li, A. Chen, and M. Wong, “Personalized nutrition recommendation system based on artificial intelligence and federated learning,” European Journal of Public Health and Environmental Research, vol. 1, no. 1, pp. 67–72, 2025.

[9] J. Muthukumar, P. Selvasekaran, M. Lokanadham, and R. Chidambaram, “Food and food products associated with food allergy and food intolerance – An overview,” Food Research International, vol. 138, 2020, doi: 10.1016/j.foodres.2020.109780.

[10] K. Verhoeckx et al., Food Processing and Allergenicity. Boca Raton, FL, USA: CRC Press, 2015.

[11] M. Mishra, T. Sarkar, T. Choudhury, N. Bansal, S. Smaoui, M. Rebezov, M. A. Shariati, and J. M. Lorenzo, “Allergen30: Detecting food items with possible allergens using deep learning-based computer vision,” Food Analytical Methods, vol. 15, no. 11, pp. 3045–3078, 2022, doi: 10.1007/s12161-022-02353-9.

[12] J. Zhang, D. Lee, K. Jungles, D. Shaltis, K. Najarian, R. Ravikumar, G. Sanders, and J. Gryak, “Prediction of oral food challenge outcomes via ensemble learning,” Informatics in Medicine Unlocked, vol. 36, 2023, doi: 10.1016/j.imu.2022.101142.

[13] A. A. Metwally, P. S. Yu, D. Reiman, Y. Dai, P. W. Finn, and D. L. Perkins, “Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via long short-term memory networks,” PLoS Computational Biology, vol. 15, no. 2, 2019, doi: 10.1371/journal.pcbi.1006693.

[14] W. Min, S. Jiang, L. Liu, Y. Rui, and S. Jain, “A survey on food computing,” ACM Computing Surveys, vol. 52, no. 5, 2019, doi: 10.1145/3329168.

[15] J. Chen, L. Pang, and J. Luo, “Cross-modal recipe retrieval: How to cook this dish?” IEEE Transactions on Multimedia, vol. 23, pp. 447–460, 2021, doi: 10.1109/TMM.2020.2976817

[16] J. Marin, A. Biswas, F. Ofli, N. Hynes, A. Salvador, Y. Aytar, I. Weber, and A. Torralba, “Recipe1M+: A dataset for learning cross-modal embeddings for cooking recipes and food images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 187–203, 2021, doi: 10.1109/TPAMI.2019.2927476.

[17] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” Neurocomputing, vol. 408, pp. 189–215, 2020.

[18] A. Gasparetto, M. Erba, A. Roldan, and F. Esposito, “A survey on text classification algorithms: From text to labels,” Information, vol. 13, no. 2, 2022

[19] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown,

[20] U. I. Shabrina, R. Sarno, R. N. E. Anggraini, A. T. Haryono and A. F. Septiyanto, "Sentiment Analysis of Presidential Candidate Debates from YouTube Videos," 2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), Bandung, Indonesia, 2024, pp. 1-6, doi: 10.1109/AIMS61812.2024.10512640.

[21] R. N. E. Anggraini, A. Nugroho, R. Wahyuwidayat and R. Sarno, "Non-Compliance Level of Motor Vehicle Taxpayer Classification," 2023 14th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 2023, pp. 261-264, doi: 10.1109/ICTS58770.2023.10330868.

[22] J. Vijaya, N. Jajam and D. Padhy, "Fine-Tuning Multilayer Perceptron Classifiers for Enhanced Heart Disease Prediction," 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2025, pp. 1-6, doi: 10.1109/IATMSI64286.2025.10984498

[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

[24] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global vectors for word representation,” in Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP), Oct. 2014, pp. 1532–1543.

[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), vol. 1 (Long and Short Papers), Jun. 2019, pp. 4171–4186.

[26] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP,” in Proc. 28th Int. Conf. Computational Linguistics (COLING), Dec. 2020, pp. 757–770.