e-ISSN 2540-9840 p-ISSN 1412-3770
Accredited by Ministry of Education, Culture, Research and Technology No: 105/E/KPT/2022
Jurnal Ilmiah Sains, April 2026, 26(1): 48-61
DOI: https://doi.org/10.35799/jis.v26i1.66334
Available online at https://ejournal.unsrat.ac.id/v3/index.php/jis

Object Recognition for Ingredient Detection and Recipe Retrieval Using Single Shot Multibox Detector

William James Olii[1], Debby Paseru[1*], Junaidy Sanger[1]

[1]Depatment of Informatics Engineering, Faculty of Engineering, De La Salle Catholic University
Manado, Indonesia
*Corresponding author: dpaseru@unikadelasalle.ac.id,

ABSTRACT

Object recognition has become an essential area of computer vision, enabling systems to identify and classify visual information for practical, everyday applications. This study aims to develop an object-recognition system capable of detecting common food ingredients and retrieving suitable recipes to assist users in cooking decisions. The research employs a dataset of 1,800 images representing six vegetable categories and implements the Single Shot Multibox Detector within an application that integrates automated recipe retrieval through a conversational artificial intelligence model. The system is evaluated through functional testing, scenario-based testing, and quantitative performance measurement using established object-recognition metrics. The results show that the model performs most effectively with an eighty–twenty training–testing split, achieving high precision and robust detection across varied lighting conditions, backgrounds, and object distances. The study also demonstrates that the system successfully identifies multiple ingredients within a single image and provides relevant recipe suggestions. These findings indicate that the approach can support domestic decision-making by connecting real-time ingredient recognition with intelligent retrieval of cooking information. The outcomes highlight the practical potential of object recognition for daily-use applications and suggest directions for expanding ingredient categories and improving detection accuracy in future work.

Keywords: Ingredient detection; object recognition; recipe retrieval

Pengenalan Objek untuk Deteksi Bahan dan Pengambilan Resep Menggunakan Single Shot Multibox Detector

ABSTRAK

Pengenalan objek telah menjadi bidang penting dalam visi komputer, yang memungkinkan sistem untuk mengidentifikasi dan mengklasifikasikan informasi visual guna mendukung aplikasi praktis dalam kehidupan sehari-hari. Penelitian ini bertujuan untuk mengembangkan sistem pengenalan objek yang mampu mendeteksi bahan makanan umum serta mengambil resep yang sesuai untuk membantu pengguna dalam pengambilan keputusan memasak. Penelitian ini menggunakan dataset yang terdiri dari 1.800 citra yang merepresentasikan enam kategori sayuran dan menerapkan *Single Shot Multibox Detector* dalam sebuah aplikasi yang mengintegrasikan pengambilan resep secara otomatis melalui model kecerdasan buatan percakapan. Sistem dievaluasi melalui pengujian fungsional, pengujian berbasis skenario, serta pengukuran kinerja kuantitatif menggunakan metrik pengenalan objek yang telah ditetapkan. Hasil penelitian menunjukkan bahwa model bekerja paling efektif pada pembagian data pelatihan–pengujian delapan puluh–dua puluh, dengan tingkat presisi yang tinggi serta kemampuan deteksi yang andal pada berbagai kondisi pencahayaan, latar belakang, dan jarak objek. Penelitian ini juga menunjukkan bahwa sistem berhasil mengidentifikasi beberapa bahan dalam satu citra dan menyediakan rekomendasi resep yang relevan. Temuan ini mengindikasikan bahwa pendekatan tersebut dapat mendukung pengambilan keputusan dalam konteks domestik dengan menghubungkan pengenalan bahan secara waktu nyata dengan pengambilan informasi memasak yang cerdas. Hasil penelitian menegaskan potensi praktis pengenalan objek untuk aplikasi sehari-hari serta memberikan arahan bagi pengembangan kategori bahan yang lebih luas dan peningkatan akurasi deteksi pada penelitian selanjutnya.

Kata kunci: Deteksi bahan; pengambilan resep; pengenalan objek

(Article History: Received 27-01-2026; Accepted 06-04-2026; Published 10-04-2026)

INTRODUCTION

Object recognition has become an essential domain within computer vision, providing computational systems with the capability to identify and locate objects within digital images and videos. Object recognition functions by mimicking aspects of human visual perception, enabling automated systems to distinguish objects based on visual cues (Batubara *et al.*, 2020; Kumar *et al*., 2023). In recent years, advancements in deep learning have accelerated the development of highly efficient object detection algorithms, allowing their application across diverse real-world scenarios, including autonomous vehicles, surveillance tools, navigation aids, and consumer-level applications. The relevance of object recognition continues to increase as societies adopt intelligent systems that can assist everyday decision-making. Within this growing landscape, integrating object detection into daily human activities—such as cooking—presents an opportunity to enhance convenience and reduce cognitive load in domestic routines.

Cooking is a fundamental human activity that involves transforming raw ingredients into meals using various tools, processing techniques, and recipes. Recipes serve as structured guidelines that assist individuals in organizing available ingredients to create desirable dishes (Negoro *et al*., 2019; Iskandar & Nugra, 2022; Alfin *et al*., 2025). Prior studies have also explored the integration of recipes into digital applications with recommendation capabilities, enabling users to receive cooking suggestions based on predefined inputs such as selected ingredients, dietary preferences, or stored recipe databases. These systems demonstrate the potential of technology in supporting culinary decision-making by simplifying the process of matching ingredients with appropriate dishes. However, selecting an appropriate recipe still becomes challenging when individuals possess limited or uncertain ingredient combinations, particularly when the identification of available ingredients relies on manual input rather than automated recognition. Many home cooks experience difficulties in determining what meals can be prepared from the ingredients they have on hand, particularly when the process relies on manual identification and personal knowledge (Iskandar & Nugra, 2022; Negoro *et al*., 2019). Previous studies have shown that limitations in recognizing and organizing available ingredients can reduce efficiency in meal planning and increase the time required for decision-making (Alfin *et al*., 2025). These challenges indicate the need for technological solutions that are capable of automatically identifying available ingredients and providing appropriate recipe recommendations.

The primary problem addressed in this study concerns the lack of automated systems capable of assisting users in identifying food ingredients through image input and subsequently retrieving suitable recipes. Without technological assistance, individuals must manually recognize the ingredients they possess and cross-reference them with available recipes, which is time-consuming and error-prone, especially for users with limited culinary experience. This gap motivates the development of an application capable of detecting food ingredients directly from uploaded images and providing recipe suggestions that align with detected items.

A general solution to this problem involves adopting computer vision models for ingredient detection and integrating them with automated recipe retrieval mechanisms. Reliable food-ingredient recognition requires an object detection algorithm that can operate efficiently on real-world, user-generated images that vary in lighting, angle, and background. Furthermore, linking detection results with an intelligent recipe-retrieval component allows users to instantly obtain multiple cooking ideas based on identified ingredients. By combining these functionalities, such a system can significantly simplify the cooking decision-making process.

Prior research has demonstrated that the Single Shot Multibox Detector (SSD) is an effective approach for object detection tasks, including applications related to food and grocery recognition (Purwanto *et al*., 2022; Puspitasari & Pangaribuan, 2025; Tanujaya & Lina, 2023). SSD operates by generating multiple bounding boxes across different feature maps in a single forward pass, enabling efficient detection of objects at various scales (Bouazizi *et al.*, 2024; Kumar *et al*., 2023). Compared to alternative approaches such as YOLO and conventional convolutional neural network (CNN)-based classifiers, SSD offers a balanced trade-off between detection accuracy and computational efficiency, making it suitable for real-time applications (Dharma *et al*., 2025; Maulidiansyah & Abdillah, 2023; Prasetyo *et al.*, 2024).

Several empirical studies have validated the effectiveness of SSD in different domains. (Tanujaya & Lina, 2023) applied SSD to grocery item recognition and achieved an accuracy of 86%, demonstrating its capability in structured retail environments with limited datasets. Similarly, (Fuady *et al*., 2020) implemented SSD in a navigation system for visually impaired users, reporting an accuracy of 92%, which highlights its robustness in real-time and assistive scenarios. More recent studies further confirm SSD’s adaptability, showing strong performance in dynamic environments such as agricultural monitoring and smart retail systems (Bouazizi *et al*., 2024; Kumar *et al*., 2023). These works collectively establish SSD as a reliable detection model; however, most of them focus on general object categories or controlled environments rather than everyday household contexts involving food ingredients.

In parallel, research on ingredient recognition and recipe recommendation has explored the integration of computer vision with intelligent retrieval systems. (Chen & Ngo, 2016) pioneered this approach by proposing a deep learning framework that maps visual ingredients to recipe databases, demonstrating the feasibility of automated cooking assistance. More recent studies have extended this concept by incorporating recommender systems and semantic matching techniques to improve the relevance of suggested recipes (Alfin *et al*., 2025; Iskandar & Nugra, 2022; Purwanto *et al.*, 2022). Despite these advancements, many existing systems rely on predefined datasets or manual input of ingredients, limiting their applicability in real-world scenarios where ingredient recognition must occur automatically from user-generated images.

Furthermore, contemporary research trends emphasize the integration of computer vision with artificial intelligence services to enhance user-centered applications. Recent studies within the last five years highlight the growing importance of combining deep learning-based object detection with intelligent recommendation systems for practical domains such as smart kitchens and food computing (Liu *et al*., 2025; Min *et al*., 2022; Zhou et al., 2021). These studies suggest that the convergence of visual recognition and AI-driven recommendation can significantly improve user experience; however, the integration remains underexplored in lightweight, real-time applications using open-source conversational models.

Based on this review, it can be observed that previous studies have either focused on object detection performance using SSD or on recipe recommendation systems using predefined inputs. Limited research has integrated both components into a unified system that operates on real-world images captured under diverse conditions. Therefore, this study positions itself at the intersection of these domains by combining SSD-based ingredient detection with an AI-driven recipe retrieval mechanism. The novelty of this research lies in the use of a custom-built dataset reflecting real household conditions, along with the integration of an open-source conversational AI model to generate dynamic recipe suggestions. This approach not only extends prior work but also addresses current gaps in practical applicability and system integration within the evolving research landscape.

Despite existing contributions, several gaps remain in the literature. Prior SSD research commonly focuses on general object categories, waste classification (Asmawati *et al*., 2025; Pernando *et al.*, 2023; Ramadhan, 2021; Salsabila *et al*., 2025), or agricultural challenges such as weed detection (Maulidiansyah & Abdillah, 2023; Saputra *et al*., 2022), leaving limited studies that apply SSD specifically to common household food ingredients. Furthermore, existing ingredient-recognition systems often rely on large, publicly available datasets, whereas practical household applications require models trained on real images captured directly by end-users under uncontrolled conditions. Additionally, current systems rarely integrate open-source conversational AI tools—such as HuggingChat (Greyling, 2023) to provide flexible and context-aware recipe suggestions. These limitations underscore the need for a system that can detect multiple frequently used vegetables under diverse conditions and generate recipe recommendations seamlessly.

Responding to these gaps, the present study aims to develop an application that applies the Single Shot Multibox Detector to detect six commonly used vegetables—cabbage, carrots, cucumbers, potatoes, tomatoes, and chayote—based on a dataset of 1,800 images collected and processed directly by the researcher. The study further integrates SSD with an open-source AI tool, HuggingChat, to retrieve three recipe suggestions for each successful detection. The novelty of this work lies in combining a custom-built ingredient dataset, SSD-based object detection, and AI-driven recipe retrieval into a unified application intended for everyday use. The scope of this study includes implementing SSD for ingredient detection, developing the necessary application framework, and conducting performance evaluation using metrics such as Mean Average Precision (mAP), Average Recall (AR), and Average Precision (AP). Through this integration, the study seeks to contribute to the advancement of object recognition applications that support daily human activities while demonstrating the practical potential of combining computer vision and AI-based retrieval systems.

RESEARCH METHOD

The research applied a structured methodological framework to develop and evaluate an object-recognition system using the Single Shot Multibox Detector (SSD) for identifying food ingredients and retrieving recipe suggestions.

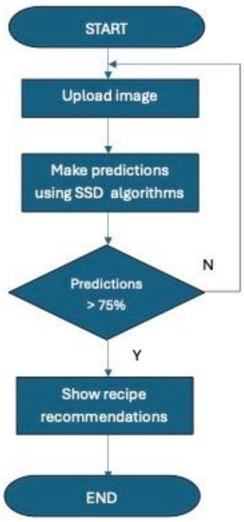

Figure 1. Research Workflow

The methodological design integrated principles from computer vision, deep learning, and software engineering. The Rapid Application Development (RAD) framework was selected as the overarching software development methodology, emphasizing accelerated prototyping and iterative validation (Habibi & Aprilian, 2020). In addition, this study adopts a data science methodology based on the Cross-Industry Standard Process for Data Mining (CRISP-DM), which structures the analytical process into stages including data understanding, data preparation, modeling, evaluation, and deployment. The integration of RAD and CRISP-DM enables the study to address both system development and data-driven model construction, ensuring that the object detection model is developed systematically while the application is iteratively refined to meet user requirements (Figure 1).

Dataset Preparation and Collection

Data collection constituted a crucial methodological component since the SSD model required a sufficiently large and representative dataset to learn the distinguishing features of food ingredients. The dataset consisted of 1,800 images representing six vegetable classes commonly used in daily cooking, namely cabbage, carrots, cucumbers, potatoes, tomatoes, and chayote. The selection of these vegetables was based on their high frequency of use in household cooking practices, as reported in previous studies on food preparation and recipe-based applications (Iskandar & Nugra, 2022; Negoro *et al*., 2019; Purwanto *et al*., 2022), ensuring the practical relevance of the system. Each class contained 300 images, resulting in a balanced dataset to support consistent model training. The dataset was self-curated by the researcher and subsequently reviewed through manual validation to ensure correct labeling and class consistency.

All images were captured from real food items in their raw state to maintain consistency in visual features and avoid variability introduced by cooking processes (Table 1). Data collection was conducted under controlled yet diverse conditions to improve model generalization. Variations in lighting conditions included natural daylight, indoor ambient lighting, and low-light environments. Backgrounds were intentionally varied to simulate real-world usage scenarios, including common kitchen surfaces, dining tables, traditional market settings, and supermarket environments, rather than controlled photobox conditions. Object positioning was also diversified by capturing images from multiple angles (top view, side view, and tilted perspectives) and distances (close-up and medium range).

These variations were systematically designed as part of the dataset preparation process described in the data collection and preprocessing stage of this study, with the aim of enhancing the robustness of the SSD model when applied to user-generated images in uncontrolled environments.

These variations were intentionally designed to improve the model's generalization ability by exposing it to diverse visual conditions during training. By incorporating images captured under different lighting conditions, backgrounds, object orientations, and distances, the model learns more robust feature representations rather than overfitting to a specific environment. This approach aligns with established practices in computer vision, where dataset diversity is used to simulate real-world variability and improve model performance on unseen data. Furthermore, the effectiveness of this strategy was evaluated through scenario-based testing, which demonstrated that the model was able to maintain stable detection performance under varying environmental conditions, including low lighting and complex backgrounds. The dataset was manually processed, labeled, and prepared for the SSD architecture. The following reconstructed Table 1 summarizes the dataset composition.

Table 1. Dataset Composition

Food Items	Total Images	Notes
Cabbage	300	Includes varied lighting and backgrounds, includes multiple object angles, variations in distance, contains texture variations, contains color variation and represents various positions
Carrot	300	
Cucumber	300	
Potato	300	
Tomato	300	
Chayote	300	
Total	**1800**	—

The dataset served as the core training and testing material for SSD. The inclusion of diverse environmental factors supported the model's resilience toward changes in user-generated images.

SSD Computation and Model Configuration

The core computational methodology relied on the Single Shot Multibox Detector (SSD) architecture, which performs object detection in a single forward pass by integrating feature extraction, bounding-box generation, classification, and Non-Maximum Suppression (NMS). The computational processes referenced key theoretical principles in convolutional operations (Ayyadevara, 2019), Rectified Linear Unit (ReLU) activations (Mellyssa *et al.*, 2022), Sigmoid functions (Dangeti, 2017), and padding (Millstein, 2018). The SSD implementation followed the operational stages described in prior studies (Juhartini *et al*., 2025; Younis *et al*., 2020), consisting of feature extraction, multi-scale detection layers, and prediction refinement using NMS.

To provide a clearer understanding, the overall architecture of the SSD model used in this study is explained below. The input image is first resized into a fixed dimension and passed through a series of convolutional layers that serve as the backbone network for feature extraction. These layers generate hierarchical feature maps that capture both low-level visual patterns and high-level semantic information. Subsequently, multiple feature maps at different scales are used to perform object detection. Each feature map is associated with a set of default bounding boxes (anchor boxes) of varying sizes and aspect ratios, enabling the model to detect objects of different scales, such as small tomatoes and larger cabbages. In each detection layer, the model simultaneously predicts bounding-box offsets and class probabilities for each default box. This multi-scale detection mechanism is a key strength of SSD, as it allows efficient detection without requiring multiple passes over the image.

After passing through the detection layers, the model produces a set of raw predictions consisting of bounding-box coordinates, confidence scores, and class labels. These predictions are then refined using the Non-Maximum Suppression (NMS) algorithm, which removes redundant and overlapping bounding boxes by selecting only the most confident detections. The final output consists of labeled bounding boxes with associated confidence scores, representing the detected ingredients within the image.

This architectural design enables SSD to achieve a balance between detection accuracy and computational efficiency, making it suitable for real-time applications involving user-generated images.

Testing and Evaluation Procedure

The testing methodology included functional testing, performance evaluation, and scenario-based image testing. Functional testing verified that the application operated across various browsers and mobile devices, could detect different ingredients, and could provide recipe suggestions as expected. Scenario testing evaluated detection reliability under multiple conditions, such as dim lighting, varying distances, dual-object images, and different backgrounds. These scenarios ensured alignment with real-world usage.

Performance evaluation relied on established object-detection metrics, including Mean Average Precision (mAP), Average Precision (AP), and Average Recall (AR). The model was trained and evaluated using three training–testing splits: 60:40, 70:30, and 80:20, to assess the impact of training data proportion on detection performance. During training, the model was configured with a fixed number of epochs and incorporated an early stopping mechanism with a defined patience value to prevent overfitting by terminating training when validation performance no longer improved. Hyperparameter settings were determined based on empirical configuration and prior studies, including learning rate, batch size, and optimizer selection, to ensure stable convergence during training. The SSD architecture utilized convolutional layers with standard filter configurations (e.g., 3×3 kernels) and progressively increasing feature depths across layers, enabling hierarchical feature extraction. Each detection layer generated multiple default bounding boxes with varying aspect ratios and scales to support multi-object detection.

These training configurations and architectural settings ensured that the evaluation results across different data splits reflected not only dataset variation but also a controlled and consistent model training process, allowing for a more reliable comparison of performance outcomes.

RESULTS AND DISCUSSION

The empirical results of the study derive from the application of the Single Shot Multibox Detector (SSD) to an image dataset of 1,800 photographs representing six common vegetables: cabbage, carrots, cucumbers, potatoes, tomatoes, and chayote. The system's performance was evaluated through functional testing, scenario-based testing, and controlled quantitative metrics, including Average Precision (AP), Average Recall (AR), and Mean Average Precision (mAP). The performance insights were obtained from three dataset splits: 60:40, 70:30, and 80:20. In addition, visual outputs were examined through a series of detection trials conducted under diverse environmental conditions. These trials demonstrated how well the SSD model generalized beyond the training data. Throughout this analysis, results are examined in relation to theoretical expectations from related studies, including those by (Chen & Ngo, 2016; Fuady *et al*., 2020; Tanujaya & Lina, 2023), enabling a grounded interpretation of the model's behavior.

The quantitative performance metrics showed that the 80:20 split delivered the strongest results, with AP reaching 88%, AR reaching 63%, and mAP reaching 88%. Lower proportions of training data resulted in reduced performance, particularly evident in the 60:40 split with approximately 69% AP and mAP. These results suggest a direct correlation between training data volume and model accuracy, consistent with prior findings that SSD benefits significantly from larger and more diverse training sets (Tanujaya & Lina, 2023). The functional and scenario-based testing validated the operational reliability of the system across devices and environmental variations, demonstrating that the SSD implementation could adapt effectively to user-generated images.

Functional Testing of the Application

Functional testing aimed to verify that the system performed as designed across all intended platforms and user workflows. The application successfully operated across multiple desktop browsers and smartphone browsers, confirming cross-platform compatibility. The system consistently accepted both uploaded images and direct smartphone camera inputs. Upon image submission, SSD processed the visual input to detect one or more ingredients and displayed the detection results along with the corresponding confidence levels. Once detection was completed, the application submitted ingredient names to HuggingChat and returned three recipes as intended.

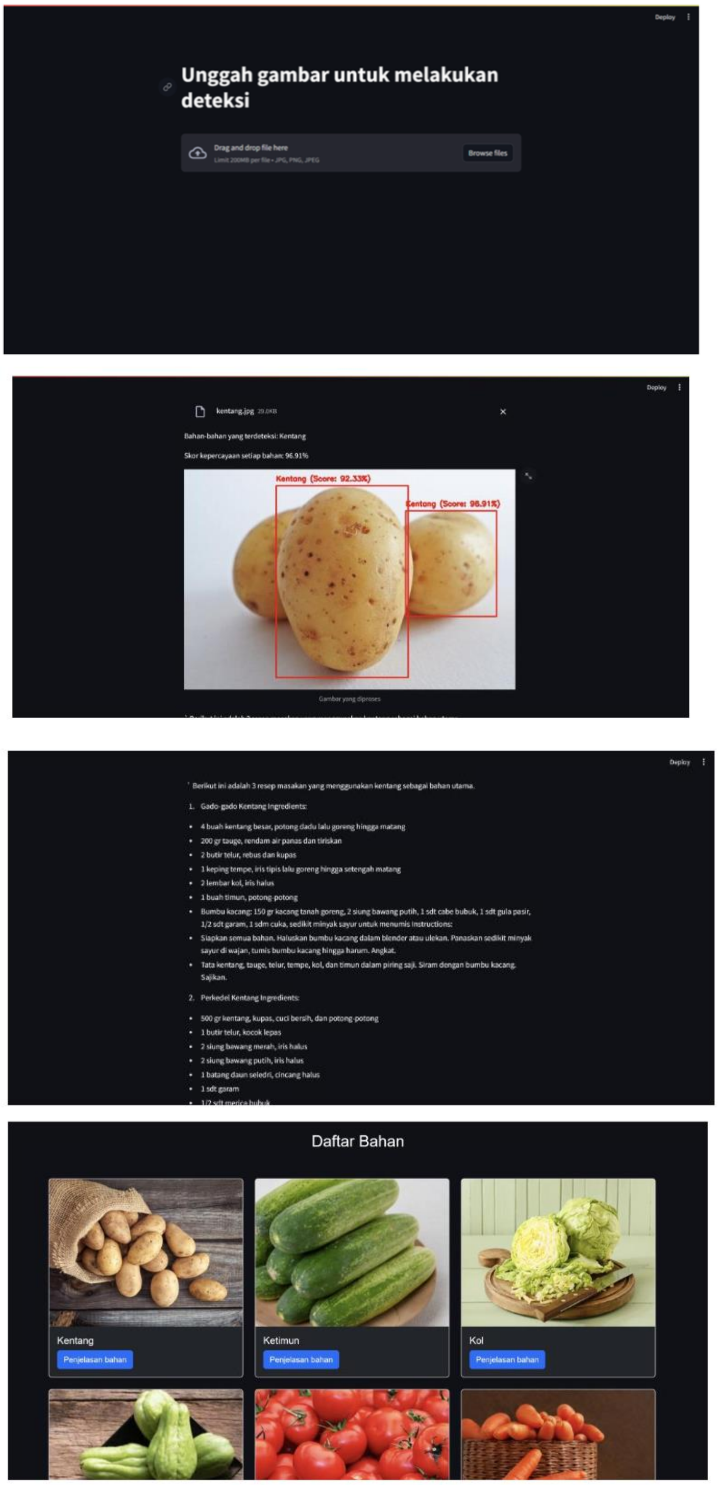

Figure 2. Application Displays

The successful performance across these functional components demonstrated that the integration between SSD and the recipe retrieval module operated effectively. This aligns with the conceptual framework introduced by (Chen & Ngo, 2016; Purwanto *et al*., 2022), where ingredient recognition is coupled with a semantic retrieval system to deliver recipe suggestions. In this study, HuggingChat (Greyling, 2023) served as the retrieval engine, and its integration worked reliably during testing. Output consistency supported the hypothesis that SSD could serve as an effective ingredient recognizer within a broader recipe suggestion ecosystem.

Scenario-Based Detection Analysis

Scenario-based testing provided insights into the model's robustness in real-world conditions. The system demonstrated reliable detection capabilities across various settings, including reduced lighting, multiple-object images, differing backgrounds, and increased camera-object distances. The figures presented in the original document illustrate these scenarios, reaffirming the SSD model's adaptability. For instance, Figure 2 shows the detection of tomatoes under dim lighting and depict the system's ability to distinguish ingredients against different background textures and colors. Figure 3 illustrates the detection of two objects in a single frame, showing the model's ability to process multiple bounding boxes simultaneously.

Figure 3. Tomatoes (dim lightning) and Tomatoes (with background) and another scenario

Figure 4. Two objects and another scenario

These findings align with SSD's theoretical strengths as a single-pass detector capable of handling multi-scale and multi-object scenarios efficiently, as outlined by (Juhartini *et al.*, 2025). The ability of SSD to detect ingredients under varied surfaces, lighting, and orientations reflects the effectiveness of the training dataset, which intentionally included diverse environmental conditions to improve generalization. This outcome is consistent with observations by (Fuady *et al.*, 2020; Tanujaya & Lina, 2023), who also reported strong SSD performance under varying conditions in unrelated object-detection tasks (Figure 4).

Quantitative Performance Evaluation

Quantitative evaluation provided the core metric-based evidence of SSD's effectiveness. The metrics—AP, AR, and mAP—were computed after training the model under three dataset splits. The 80:20 split yielded the strongest results, with AP at 88%, AR at 63%, and mAP at 88%. These values are consistent with SSD's expected behavior as a high-performing detector when equipped with a sufficient training dataset. By contrast, the 60:40 and 70:30 splits delivered lower performance, specifically around 69% AP and mAP for the former and around 75% for the latter (Table 2).

Table 2. The Performance Evaluation

Train–Test Split	Average Precision (AP)	Average Recall (AR)	Mean Average Precision (mAP)
60:40	69%	61	69%
70:30	75%	54	75%
80:20	88%	63%	88%

These results support the hypothesis that SSD can accurately detect food ingredients if provided with high-quality training data. The positive impact of expanding the training portion is consistent with the finding from related work (Tanujaya & Lina, 2023) that increasing epochs or dataset size can influence accuracy. Although the referenced study noted that excessive epochs lowered accuracy, the present findings affirm that dataset richness, rather than epoch expansion, contributes more substantially to model precision.

These performance metrics also reflect SSD's ranking in the literature as a detector with strong speed–accuracy balance compared to YOLO and CNN-based models (Tanujaya & Lina, 2023). While the study does not compare SSD directly with other algorithms, the achieved accuracy exceeding 75% across all conditions—and reaching 88% in the optimal split—falls within ranges reported by previous SSD applications in diverse domains. For example, (Fuady *et al*., 2020) achieved 92% accuracy in detecting objects within a navigation tool, while (Tanujaya & Lina, 2023) reported 86.08% accuracy in recognizing grocery items. Although the context of food-ingredient recognition presents unique challenges such as color similarity among vegetables, the model nonetheless attained competitive results.

Analysis of Model Behavior and Error Patterns

The results reveal several patterns in the model's detection behavior. The relatively high AP suggests that the model is effective at predicting correct ingredient classes when detections are made, but the lower AR value of 63% indicates that some ingredients were not consistently detected across all test images. This discrepancy may stem from the visual similarity among certain vegetables, such as cabbage and chayote or carrots and cucumbers when photographed under specific angles or light conditions.

Additionally, AR performance highlights the model's sensitivity to incomplete or partially occluded objects. Some test images, particularly those taken from longer distances or with complex backgrounds, displayed reduced confidence scores. This finding aligns with the theoretical challenges of object detection in cluttered scenes, where bounding-box localization may compete with background textures or lighting gradients. Nonetheless, the model's capacity to detect multiple objects simultaneously, as illustrated in Figure 2, demonstrates its capability to manage cluttered scenes effectively when object edges are sufficiently distinct.

The error distribution also corresponds with findings from previous SSD-based research in domains such as waste classification (Pernando *et al.*, 2023) and weed detection (Saputra *et al*., 2022), in which detection performance varied across object categories. In both referenced studies, intra-class similarity posed detection challenges. The current study's dataset includes closely related vegetable textures, indicating that SSD's performance could potentially be improved with additional preprocessing or the inclusion of enhanced edge-detection layers.

Discussion Linking Results to Research Hypothesis

The research posited that SSD could be implemented successfully to detect food ingredients and subsequently generate relevant recipe suggestions through integration with HuggingChat. The empirical findings support this hypothesis. The system's ability to accurately detect six vegetable classes with high AP and mAP values, coupled with its reliable retrieval of recipes, demonstrates that SSD is an effective tool for ingredient recognition in everyday cooking contexts. The multiple forms of testing further confirm that the model generalizes well across real-world scenarios.

Comparing the results to earlier studies strengthens this conclusion. The achieved AP of 88% exceeds or aligns with performance levels in existing SSD-based works (Fuady *et al.*, 2020; Tanujaya & Lina, 2023), while the inclusion of recipe retrieval aligns with the approach taken by (Chen & Ngo, 2016) The integration of object detection with open-source AI for recipe retrieval contributes to the emerging field of intelligent domestic assistance systems. Although AR performance reveals room for improvement, especially for visually similar ingredients, the overall results demonstrate that the system performs adequately for its intended purpose and is empirically supported by both functional outcomes and quantitative measures.

CONCLUSION

This study successfully developed an object-recognition system that can detect common food ingredients and retrieve suitable recipes to support users in making cooking decisions. Using the SSD model on a dataset of 1,800 images across six ingredient classes, the system achieved its best performance with an 80:20 training–testing split, reaching 88% Average Precision, 88% Mean Average Precision, and 63% Average Recall. The results show that SSD can reliably recognize ingredients under varied real-world conditions, while the integration with HuggingChat enabled the system to automatically generate relevant recipe recommendations from detected ingredient labels. Overall, the study demonstrates the practical potential of combining object detection and conversational AI in a unified application for everyday culinary assistance.

ACKNOWLEDGMENT

Thank you to Rector De La Salle Katolik University Manado and colleagues for supporting us.

REFERENCES

Alfin, M., Akbar, M. ., Manik, A. R., Purba, J., & Perdana, A. (2025). Pengembangan Aplikasi Rekomendasi Resep Makanan Berdasarkan Bahan Menggunakan Api. *JATI (Jurnal Mahasiswa Teknik Informatika), 9*(6), 10152–10160. https://doi.org/10.36040/jati.v9i6.16419

Asmawati, A., Kurniawan, Y., & Fadilah, R. . (2025). Efisiensi Pemilahan Sampah Daur Ulang Menggunakan Model CNN SSD-MobileNet V2. *ICIT Journal*, *11*(2), 197–208.

Ayyadevara, V. . (2019). *Neural Networks with Keras Cookbook*. Packt Publishing.

Batubara, N. ., Awangga, R. M., & Pane, S. F. (2020). *Perbandingan Faster R-CNN dengan SSD Mobilenet Untuk Mendeteksi Plat Nomor*. Kreatif Industri Nusantara.

Bouazizi, O., Azroumahli, C., El Mourabit, A., & Oussouaddi, M. (2024). Road Object Detection using SSD-MobileNet Algorithm: Case Study for Real-Time ADAS Applications. *Journal of Robotics and Control (JRC), 5*(2), 551–560. https://doi.org/10.18196/jrc.v5i2.21145

Chen, J., & Ngo, C. (2016). Deep-based Ingredient Recognition for Cooking Recipe Retrieval. *Proceedings of the 24th ACM International Conference on Multimedia*, 32–41. https://doi.org/10.1145/2964284.2964315

Dangeti, P. (2017). Statistics for Machine Learning. Birmingham: Packt Publishing.

Dharma, A.S., Pardosi, C.N.S., & Silaen, Z. P. (2025). Comparative Performance of Yolov8 and Ssd-mobilenet Algorithms for Road Damage Detection in Mobile Applications. *Sinkron*, *9*(3), 1159–1169. https://doi.org/10.33395/sinkron.v9i3.15008

Fuady, S., Nehru, N., & Anggraeni, G. (2020). Deteksi Objek Menggunakan Metode Single Shot Multibox Detector Pada Alat Bantu Tongkat Tunanetra Berbasis Kamera. *Journal of Electrical Power Control and Automation (JEPCA), 3*(2), 39–43. https://doi.org/10.33087/jepca.v3i2.38

Greyling, C. (2023). HuggingChat (online). https://cobusgreyling.medium.com/huggingchat-from-huggingface-is-an-open-source-ai-chat-interface-9be3b0e09b79

Habibi, R., & Aprilian, R. (2020). Tutorial dan penjelasan aplikasi e-office berbasis web menggunakan metode RAD. Jakarta: Kreatif Industri Nusantara.

Iskandar, L. ., & Nugra, R. . (2022). nnovations in Culinary Exploration: AI-Enhanced Food Recipe Search Techniques. *Advanced International Journal of Material Science and Engineering*, *7*(3), 12–20.

Juhartini, Dwinita Arwidiyarti, & Desmiwati. (2025). Single Shot Multibox Detector (SSD) in Object Detection: A Review. *IJACI : International Journal of Advanced Computing and Informatics*, *1*(2), 118–127. https://doi.org/10.71129/ijaci.v1i2.pp118-127

Kumar, V., Goel, V., Amoriya, A., & Kumar, A. (2023). Object Detection Using SSD. *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 743–748. DOI: 10.1109/ICAC3N60023.2023.10541704

Liu, D., Zuo, E., Wang, D., He, L., Dong, L., & Lu, X. (2025). Deep Learning in Food Image Recognition: A Comprehensive Review. *Applied Sciences*, *15*(14), 7626. https://doi.org/10.3390/app15147626

Maulidiansyah, M., & Abdillah, I. (2023). Klasifikasi Jenis Jagung Berdasarkan Bentuk Biji Menggunakan Metode You Only Look Once (YOLO). *JEECOM Journal of Electrical Engineering and Computer*, *5*(2), 215–226. https://doi.org/10.33650/jeecom.v5i2.6802

Mellyssa, W., Dewi, A., Misriana, M., Suryati, S., & Rachmawati, R. (2022). Pengaruh Algoritma Deep Learning dalam Meningkatkan Akurasi Sistem Pendeteksian Kondisi Jalan Raya. *Prosiding Seminar Nasional Politeknik Negeri Lhokseumawe*.

Millstein, F. (2018). Convolutional Neural Networks In Python: Beginner's Guide To Convolutional Neural Networks in Phyton. South Carilina: CreateSpace Independent Publishing.

Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2022). A Survey on Food Computing. *ACM Computing Surveys*, *55*(1), 1–37. https://doi.org/10.48550/arXiv.1808.07202

Negoro, W. A., Amalia, F., & Santoso, E. (2019). Pengembangan Aplikasi Resep Masakan dengan Rekomendasi berdasarkan Bahan-Bahan Makanan Berbasis Web. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(9), 9212–9221.

Pernando, P., Rafi Muttaqin, M., & Raymond Ramadhan, Y. (2023). Deteksi Jenis Sampah Secara Realtime Menggunakan Metode Single Shot Multibox Detector (SSD). *JATI (Jurnal Mahasiswa Teknik Informatika), 7*(3), 1890–1895. DOI: 10.36040/jati.v7i3.6976

Prasetyo, S. E., Atmaja, C., Ardian, M., Ardhiansyah, A., Sudarni, A. R., & Khaira, M. (2024). Benchmarking YOLOv3 and SSD: A Performance Comparison for Multi-Object Detection. *Edu Komputika Journal*, *11*(2), 136–146. https://doi.org/10.15294/edukom.v11i2.28005

Purwanto, H., Novriadi, A., & At Thariq, F. (2022). Application of Artificial Intelligence SSD MobileNet and Tiny YOLOv2 for Food Recipe Search. *International Journal of Engineering, Science and Information Technology*, *2*(3), 1–7. https://doi.org/10.52088/ijesty.v2i3.276

Puspitasari, F. ., & Pangaribuan, H. (2025). Deteksi Kalori pada Citra Makanan dengan Algoritma Single Shot Multibox Detector. *Comasie Journal*, *12*(2), 9–100.

Ramadhan, M. H. . (2021). Pendeteksian Jenis Sampah dengan Single Shot Multibox Detector secara Real Time [Skripsi]. Universitas Sriwijaya, Palembang.

Salsabila, Z. K., Prakisya, N. P. T., & Liantoni, F. (2025). Deep Learning Architectures for Waste Detection: A Systematic Literature Review. *Jurnal Media Informasi Teknologi*, *2*(1), 41–52. https://doi.org/10.69616/mit.v2i1.213

Saputra, A. A., Susilo, B., Yusa, M., & Nurjanah, U. (2022). Sistem Pendeteksi Genus Gulma Pada Tanaman Jagung Menggunakan Algoritme Single Shot Detector. *Rekursif: Jurnal Informatika*, *10*(1), 48–60. https://doi.org/10.33369/rekursif.v10i1.18634

Tanujaya, H., & Lina, L. (2023). Pengenalan Objek menggunakan Metode Single Shot Multibox Detector pada Bahan Sembako. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *11*(1), 1–6.

Younis, A., Shixin, L., Jn, S., & Hai, Z. (2020). Real-Time Object Detection Using Pre-Trained Deep Learning Models MobileNet-SSD. *Proceedings of 2020 6th International Conference on Computing and Data Engineering*, 44–48. https://doi.org/10.1145/3379247.3379264

Zhou, X., Wang, D., & Krähenbühl, P. (2021). Objects as points: CenterNet-based detection advancements. *IEEE Transactions on Paten Analysis and Machine Intelligence*, *4*(3), 1–15.