Deep Learning-based Ingredient Detection for a Recipe Recommendation System

Jan Ed Paulo A, Figueroa
Technological Institute of the Philippines
Quezon City, Philippines
qjepafigueroa@tip.edu.ph

Vhan Randolp S. Peña
Technological Institute of the Philippines
Quezon City, Philippines
qvrspena@tip.edu.ph

Richard Roman
Technological Institute of the Philippines
Quezon City, Philippines
rrichard.cpe@tip.edu.ph

***Abstract*—This paper aims to design an innovative recipe recommendation system that leverages deep learning techniques to detect ingredients from household food inventories, thereby optimizing food usage and minimizing waste. The system utilizes convulsionals neural networks (CNNs), specifically employing YOLO models versions 6, 8, and 9 for accurate and efficient ingredient detection. The YOLOv9 model achieved the best results with 91% precision and recall, indicating high accuracy in predicting positive labels and identifying most of the true positive labels backing with a mean-average precision (maP) of 95% demonstrating a good ratio to the object detection tasks. On the other hand, YOLOv6 and YOLOv8 models achieved around 81% in precision, 77% in recall, and 81% in maP demonstrating a more unreliable model for ingredient detection. By recognizing ingredients in real-time through image processing, the system recommends recipes that maximize the use of available ingredients, encouraging efficient meal planning and reducing the frequency of unused food items. The performance of the system is evaluated through experiments and evaluation results of the accuracy in ingredient detection and its efficiency in food inventory optimization.**

***Keywords*—convolutional neural networks, deep learning, object detection, YOLO models, recipe recommendation system,**

1. INTRODUCTION

Household food waste is the most damaging in terms of environmental and economic impact. Many efforts have been made to quantify and analyze the reasons for and problems associated with household food waste generation which has led to the development of both technical solutions and behavioral interventions (including education and awareness) to try and reduce its generation [1]. Despite food insecurity, approximately one-third of all food produced for human consumption is reported to be wasted. Plate waste is defined as edible portions of food which are left on the dining table or on the plates after the family has finished eating and are usually given to household pets or discarded [2]. The Sustainable Development Goal (SDG) 12.3 specifically targeted and aimed to halve plate waste and decrease food loss by 2030.

According to the World Wildlife Fund–Philippines (2020), it is estimated that around 2,175 tons of food scraps in greater Metro Manila alone were thrown in the garbage every day in the year 2020, while around 308,000 tons around the country are being considered as food waste [3]. Despite food insecurity, approximately one-third of all food produced for human consumption is reported to be wasted. It is believed that it is due to the lack of planning of the consumers in order to relieve the overstock of the food. In the Philippines, plate waste is closely linked to hunger incidence and threatened food security. Given that plate waste is mostly generated at home, the typical Filipino family generates 66.8g of plate waste each day which is 5.0g more than in 2015 [2].

Food security is one of the global concerns given its economic implications that almost reach a trillion dollars all over the world, its environmental impacts with its greenhouse gasses, and social impacts such as food insecurity and hunger. Therefore, there are many researches occurring in this field in order to attempt to alleviate the progress of this problem also to give solutions for people who are indecisive on which dish they'll prepare which is the aim of Team of Rodiguez with their conquest of a Recipe Recommendation System through Classification of Food Ingredients [4]. The research 'A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients' was proposed in order to plan the dishes to be prepared in the following days [5].

Deep Learning algorithms and Computer vision are being utilized in many fields including the Food Recommendation System. Therefore, this paper aims to produce a recipe recommendation system that aims to set out a plan of meals within a specified amount of time in order to optimize the food inventory. This research is only delimited to the counting of the specific ingredients and doesn't involve any weight estimations which are helpful for meat products. In order to achieve this, we will use several object detection models for ingredient detection for an accurate inventory of food stocks.

2. Methods

A. *Object Detection Dataset*

Data required for this research were images of raw ingredients of a common part of Filipino cuisine. The image datasets used for this research were carefully picked and web-scraped from different web platforms such as Kaggle, OpenCv, RoboFlow, Adobe Stock, and others. The following platforms have different qualities in order for the model to broaden its scope of understanding for object detection. We have gathered 250 images per class which have 200 images for train data and 50 for test data.

TABLE I: *List of Classes in Dataset*

Label	***Class***
0	*Beef*
1	*Bitter-Gourd*
2	*Bottle-Gourd*
3	*Broccoli*
4	*Cabbage*
5	*Carrots*
6	*Cauliflower*
7	*Chicken*
8	*Egg*
9	*Eggplant*
10	*Galunggong*
11	*Garlic*
12	*Ginger*
13	*Milkfish*
14	*Onion*
15	*Papaya*
16	*Pork*
17	*Potato*
18	*Sayote*
19	*Tilapia*
20	*Tomato*

The dataset images also were manually annotated with bounding boxes and class labels using annotation tools such as LabelImg or Roboflow to create the labeled datasets necessary for training the models.

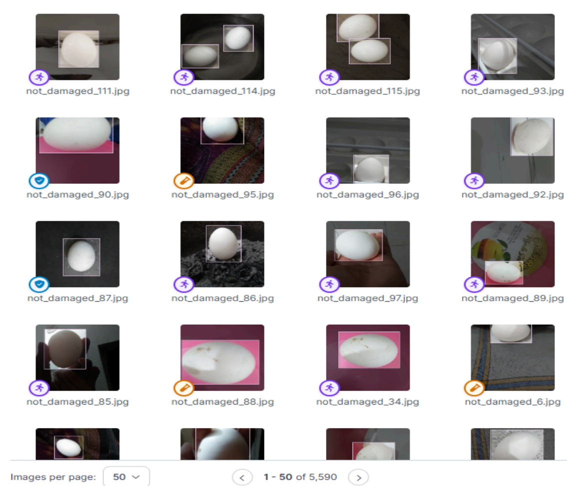


FIGURE I: Annotating of image datasets

B. *Recipe Recommendation System Dataset*

The dataset used for the recipe recommendation system is outsourced from GitHub by Shaan Subbaiah [10] on project, allrecipes-scraper. This dataset consists of 35,516 entries of 47 columns which includes the following classes:

1. Name: The name of the dish that the users are about to cook
2. Url: The source of the recipe from allrecipes.com

3. Category: The type of dishes about to serve such as drinks, bread, pasta, main dishes, etc.
4. Author: The name of the creator of the dish
5. Summary: A brief description about the dish
6. Ratings: A score related to the dish that includes the rating, rating_count, and review_count.
7. Ingredients: List of ingredients necessary for the dish
8. Directions: The set of cooking steps
9. Time: The duration of each process such as preparation time, cook time, and total time.
10. Servings: The number of people dedicated for the ratio of the dish
11. Yield: The final output of the dish
12. Nutrients: Includes the nutritional content such as calories, fat, carbohydrates, etc.

C. Object Detection Data Preprocessing

Data preprocessing is a crucial part of computer vision applications which involves simplification of the image dataset via resizing, interpolation, and augmentation. The preprocessing stage removes the complexity of the image which may confuse the model as it was recognized as one of the most significant features of the class. Tensorflow and RoboFlow were used for the preprocessing stage of this research. The model is chosen because of the performance it serves while training the custom dataset.

Figure II: Image Augmentation

D. Recipe Recommendation System Data Preprocessing

Performing data preprocessing aims to improve the data quality by filtering the dataset and collecting the necessary data relevant to the objective of the project. The following actions were done in order to preprocess the data.

1. Conversion of fractions in unicode format to decimal form for uniform presentation, data integrity, and normalization of data formats.
2. Extraction of crucial information from ingredients column to the another column which allows easier retrieval of core data for the recommendation system.

E. Design Architecture

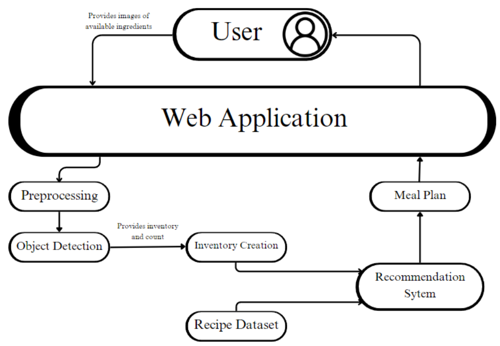


FIGURE III: Design architecture of the web application

This recommendation system utilizes Streamlit to produce a web application for the users. In order to use this application, it is necessary for the users to input images of their images into the application. Receiving the image data, it performs object detection that will produce an inventory of detected ingredients included in the ingredient classes we defined and the corresponding count of each ingredient. Then, the recommendation system will produce a meal plan within the given inventory.

F. Object Detection Models

1. *YOLOv6*

YOLOv6 is a state-of-the-art object detection model designed for a balance between speed and accuracy. It uses an efficient architecture for real-time applications, incorporating advanced techniques to enhance feature extraction while reducing computational complexity. YOLOv6 integrates innovative training strategies and optimization algorithms to improve detection performance across various scenarios.

2. *YOLOv8*

YOLOv8 is the latest version of the real-time object detection and image segmentation model, leveraging cutting-edge advancements in deep learning and computer vision. It offers unparalleled speed and accuracy, with a streamlined design suitable for various applications. YOLOv8 is easily adaptable to different hardware platforms, from edge devices to cloud APIs.

3. *YOLOv9*

YOLOv9 builds on its predecessors by incorporating improved backbone networks, spatial pyramid pooling, and attention mechanisms for better feature extraction and multi-scale object detection. It uses anchor-free detection and sophisticated loss functions like CIoU for improved accuracy and localization. YOLOv9 maintains the YOLO family's philosophy of high-speed, real-time object detection with enhanced precision and efficiency.

G. Metrics and Evaluation

The metrics to be used for the evaluation of the deep learning models ensure the integrity and accuracy of the ingredient detection for accurate inventory and recommendation. The following metrics contribute to the understanding of the accuracy of the object detection models.

1. **Confusion Matrix:** a performance measurement tool for machine learning classification tasks. It is a table that compares the predicted labels against the actual labels to visualize and evaluate the performance of an algorithm [6].

TABLE II: Confusion Matrix/ Contingency Table

	Actual Positive	**Actual Negative**
Predicted Positive	TP	FP
Predicted Negative	FN	TN

2. **Loss**: Training Loss metrics measure the disparity between the predicted and actual bounding boxes.

3. **Mean Average Precision (MAP)**: measures the average precision of the top-K recommended items, overall users. It takes the average of the precision values for each user, where precision is the proportion of relevant items among the top-K recommended items [7].

$$\frac{1}{n}\sum_{k=1}^{k=n} AP_K$$

n = number of class

APk = the average precision of class k

4. **Precision and recall**: provides a balanced understanding of a model's performance in object detection, highlighting both its accuracy in predictions and its ability to detect all relevant objects [8].

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

TP = True Positives

FP = False Positives

FN = False Negatives

G. Content-based Recommendation System

Content-based recommendation system refers to the usage of an algorithm that provides recommendation based on the content of the data. It finds the similarities between the features of recipe ingredients and generates recipe recommendations. [9][10] One of the advantages of using a content-based recommendation system is that it maintains an objective recommendation based on the data, recipe ingredients.

In order for the recommendation system to decide, a score system will be implemented which defines the consumption percentage of each recipe to the inventory.

$$score = \Sigma \frac{number\ of\ consumed\ ingredient}{total\ number\ of\ ingredient} \div number\ of\ classes$$

This score will be used as a reference for building a meal plan for the user ensuring that making full use of the inventory has been made as the first priority.

III. Results and Discussion

Using the custom ingredients dataset, the performance of YOLOv6, YOLOv8, and YOLOv9 models was evaluated for ingredient detection. The YOLOv6 and YOLOv8 models were trained for 80 epochs each. Due to the resource-intensive nature of the latest YOLOv9 version, it was trained for only 30 epochs, which proved to be sufficient.

TABLE III. Performance Metrics

	Precision	Recall	mAP50	mAP50-95
YOLOv6	**82.30%**	**76.09%**	**81.64%**	**71.95%**
YOLOv8	**81.25%**	**78.59%**	**82.63%**	**75.19%**
YOLOv9	**91.96%**	**91.53%**	**95.40%**	**77.92%**

The table shows the performance metrics for YOLOv6, YOLOv8, and YOLOv9 object detection models. YOLOv9 exhibits the highest precision at 91.96%, significantly outperforming YOLOv6 (82.30%) and YOLOv8 (81.25%). In terms of recall, YOLOv9 again leads with 91.53%, followed by YOLOv8 at 78.59% and YOLOv6 at 76.09%. When looking at mAP50, which measures Average Precision at an IoU threshold of 0.50, YOLOv9 achieves 95.40%, indicating a considerable improvement over YOLOv6 (81.64%) and YOLOv8 (82.63%). For mAP50-95, which averages precision across multiple IoU thresholds, YOLOv9 records 77.92%, surpassing YOLOv6 (71.95%) and YOLOv8 (75.19%).

A. *YOLOv6*

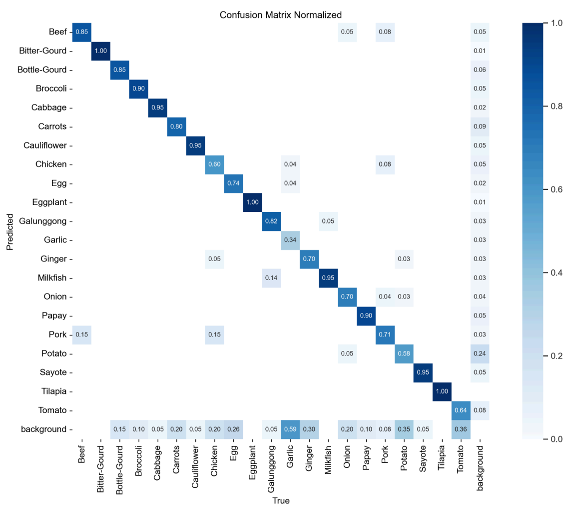


FIGURE IV: YOLOv6 Normalized Confusion Matrix

The confusion matrix indicates perfect accuracy in detecting bitter gourd and eggplant, and high accuracy for cabbage and cauliflower. However, there are significant misclassifications for chicken, garlic, and especially ginger. Misclassifications frequently occur between similar-looking ingredients, such as pork with beef and papaya, and ginger with garlic.

TABLE IV. YOLOv6 Metrics Performance

Metrics	Value
train/box_loss	36.35%
train/cls_loss	59.69%
train/dfl_loss	105.71%
val/box_loss	54.16%
val/cls_loss	84.55%
val/dfl_loss	120.03%
metrics/precision (B)	82.30%
metrics/recall(B)	76.09%
metrics/mAP50(B)	81.64%
metrics/mAP50-95(B)	71.95%

The graphs indicate that the object detection model shows significant improvement in performance over time. Both box and classification losses for training and validation steadily decrease, indicating enhanced accuracy in predicting bounding boxes and classifying objects. Precision and recall metrics stabilize above 0.8, reflecting high accuracy and the model's capability to identify true objects with low false positive and negative rates. The mean average precision (mAP) metrics, particularly mAP@50 and mAP@50-95, demonstrate strong and robust detection performance.

B. *YOLOv8*

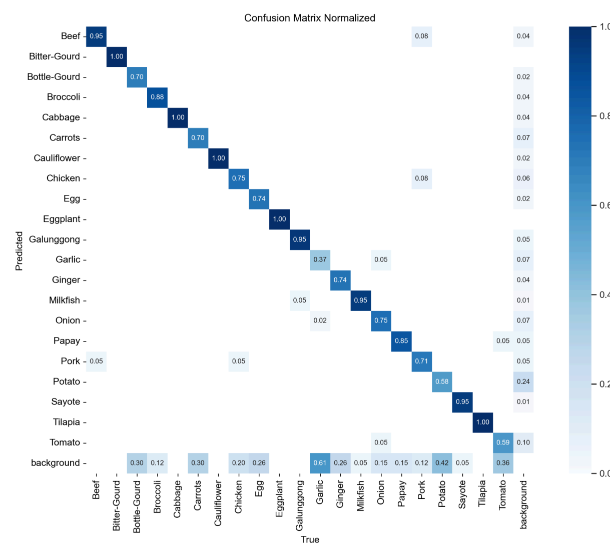


FIGURE VI: YOLOv8 Normalized Confusion Matrix

The confusion matrix for YOLOv8 shows varied performance across different categories. Bitter-gourd, Cabbage, Cauliflower, Eggplant, and Sayote achieve perfect classification accuracy. However, some categories like Bottle-Gourd and Carrots have lower accuracies of 70%. Notable misclassifications include Beef being confused with Pork and Milkfish, and Chicken with Cauliflower. Background images also show significant confusion with various categories such as Tomato and Tilapia.

TABLE V. YOLOv8 Metrics Performance

Metrics	Value
train/box_loss	24.60%
train/cls_loss	32.93%
train/dfl_loss	95.11%
val/box_loss	48.19%
val/cls_loss	84.37%
val/dfl_loss	113.20%
metrics/precision (B)	81.25%
metrics/recall(B)	78.59%
metrics/mAP50(B)	82.63%
metrics/mAP50-95(B)	75.19%

The metrics graph for YOLOv8 shows a consistent decrease in both training and validation losses, indicating effective learning. Precision and recall metrics steadily improve, approaching values near 0.8, demonstrating the model's increasing accuracy. The mAP metrics also stabilize at high values, reflecting strong performance in object detection and segmentation tasks which is a reasonable result but not good enough.

C. YOLOv9

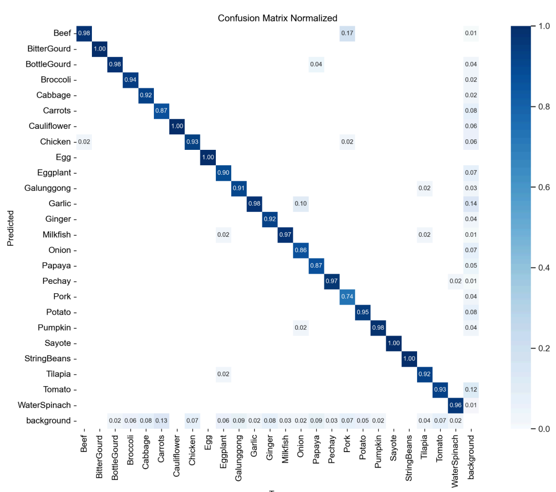


FIGURE VII: YOLOv9 Normalized Confusion Matrix

This figure shows the confusion matrix for YOLOv9 shows high accuracy for most categories, with many items having near-perfect precision and recall. Beef, BitterGourd, Pumpkin, and Sayote achieve 100% correct classifications. Other items like Chicken, Broccoli, and Cabbage also exhibit high accuracy, although there are some misclassifications, such as Carrots being confused with Broccoli and Galunggong with Milkfish. Overall, YOLOv9 performs robustly across various categories, maintaining high precision and recall

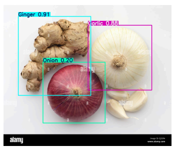
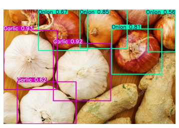
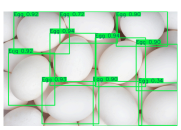
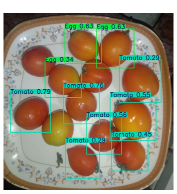
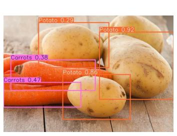
TABLE VI. YOLOv9 Metrics Performance

Metrics	Value
train/box_loss	38.49%
train/cls_loss	21.67%
train/dfl_loss	105.11%
val/box_loss	73.46%
val/cls_loss	42.69%
val/dfl_loss	129.33%
metrics/precision (B)	91.96%
metrics/recall(B)	91.53%
metrics/mAP50(B)	95.40%
metrics/mAP50-95(B)	77.92%

The YOLOv9 metrics show significant improvements in model performance, with loss values consistently decreasing and stabilizing. Training and validation box losses fall below 0.5, and classification losses settle around 0.5 for training and 0.75 for validation. Precision and recall exceed 0.9, indicating high accuracy. Mean average precision (mAP) metrics stabilize around 0.95 (mAP@50) and 0.78 (mAP@50-95), reflecting robust performance. Overall, YOLOv9 achieves high accuracy and precision in ingredient detection.

D. Recipe Recommendation System

TABLE VII: YOLOv9 Test Predictions

Images	True Label	Predicted Labels
	[11, 12, 14]	[11, 12, 14]
	[16, 16, 16]	[16]
	[7]	[7]
	[11, 11, 11, 12, 14, 14, 14, 14, 14]	[11, 11, 11, 14, 14, 14, 14]
	[8, 8, 8, 8, 8, 8, 8, 8, 8, 8]	[8, 8, 8, 8, 8, 8, 8, 8, 8, 8]
	[4]	[4]
	[20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20]	[8, 8, 8, 20, 20, 20, 20, 20, 20, 20]
	[5, 5, 17, 17, 17, 17]	[5, 5, 17, 17, 17]

Comparing the performance metrics of the three YOLO models, YOLOv9 produced the highest values proving its effectiveness in correctly detecting most of the ingredients. YOLOv9 is not an exception to the challenges of computer vision such as occlusion. In relevance to the results of the prediction, the model proved to be reliable to its high precision and recall along with the maP50.

Recommended Recipes

name	category	url
Vegan Carrot Soup	soups-stews-and-chili	https://www.allrecipes.com/recipe/70101/vega
Egg Pesto Breakfast Wrap	breakfast-and-brunch	https://www.allrecipes.com/recipe/182811/egg
Posole Soup	soups-stews-and-chili	https://www.allrecipes.com/recipe/17826/poso
Asian Twist Chicken Wings	appetizers-and-snacks	https://www.allrecipes.com/recipe/162598/asia
Garlic Steak with Garlic	meat-and-poultry	https://www.allrecipes.com/recipe/222330/garl
Zucchini Egg White Frittata	breakfast-and-brunch	https://www.allrecipes.com/recipe/232664/zuc
Sheet Pan Passover Dinner	everyday-cooking	https://www.allrecipes.com/recipe/283070/she
Rainbow Roasted Pepper Soup	soups-stews-and-chili	https://www.allrecipes.com/recipe/13263/rainb
Potato Salad Deviled Eggs	appetizers-and-snacks	https://www.allrecipes.com/recipe/214713/pot
Sun Dried Tomato Focaccia	bread	https://www.allrecipes.com/recipe/6692/sun-d

FIGURE VIII: Recipe Recommendation Output

The recipes shown in Figure VIII resulted from the usage of percentage consumption and knapsack solution. This approach ensures the maximum utilization of the ingredients by selecting recipes from having the highest percentage consumption after scaling to the new serving size inputted by the user.

IV. Conclusion

To conclude, this paper presents a comparison on different YOLO models, specifically version 6, 8 and 9 for an ingredient detection model for a recipe recommendation system. The models were trained from a custom dataset web scraped through different platforms such as Roboflow, Kaggle, etc. building an inventory that will be used for a base of recipe recommendation system. The models demonstrated consistent improvements in detecting and classifying various household ingredients, with significant decreases in loss values and increases in precision, recall, and mean average precision (mAP). These metrics indicate the models' effectiveness in real-time ingredient detection. Among the models, YOLOv9 performed the best, achieving 91% precision and recall along with the 95% mean average precision (mAP). These metrics are significant for the recipe recommendation system as its input for a more accurate count and detection of the available ingredients. The generated recipes are aligned with the user's available inventory and desired serving size optimizing the inventory usage.

For future work, the researchers plan to improve the detection models to achieve higher performance such as detection of canned goods, leafy vegetables, and other ingredients. Additionally, integration of user preferences and nutritional value into the recipe recommendation system will be done to provide a greater satisfaction to the user by optimizing both inventory, dietary needs, and satisfaction.

References

[1] E. Woolley, Z. Luo, A. Jellil, and A. Simeone, “A Data Driven Approach to Reducing Household Food Waste,” Sustainable Production and Consumption, Nov. 2021, doi: https://doi.org/10.1016/j.spc.2021.11.004.

[2] I. Angeles-Agdeppa, M. B. Toledo, and J. A. T. Zamora, “Does plate waste matter?: A two-stage cluster survey to assess the household plate waste in the Philippines,” BMC Public Health, vol. 23, no. 1, Jan. 2023, doi: https://doi.org/10.1186/s12889-022-14894-z.

[3] Rodrigues, M. S., Fidalgo, F., & Oliveira, Â. (2023). RecipeIS—Recipe Recommendation System Based on Recognition of Food Ingredients. Applied Sciences, 13(13), 7880. https://doi.org/10.3390/app13137880

[4] Yanai, K., Maruyama, T., & Kawano, Y. (2014). A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients. International Journal of Interactive Mobile Technologies (IJIM), 8(2), 28. https://doi.org/10.3991/ijim.v8i2.3623

[5] K. Joyce, P. Ortiz, N. Bautista, V. Dimailig, A. Christian, and D. Llamzon, “Recipe Recommendation System Using IoT-Based Food Inventory Management of Perishables for Household Food Waste Reduction,” Chemical Engineering Transactions, vol. 106, pp. 361–366, 2023, doi: https://doi.org/10.3303/CET23106061.

[6] M.S. Sandeep, Koravith Tiprak, Sakdirat Kaewunruen, Phoonsak Pheinsusom, and Withit Pansuk, “Shear strength prediction of reinforced concrete beams using machine learning,” vol. 47, pp. 1196–1211, Jan. 2023, doi: https://doi.org/10.1016/j.istruc.2022.11.140.

[7] N.A. Ahmed, Mean Average Precision (mAP): A Complete Guide. (n.d.).Kili-Website.https://kili-technology.com/data-labeling/machine-learning/mean-average-precision-map-a-complete-guide

[8] Huilgol, P. (2020, September 3). Precision vs Recall | Precision and Recall Machine Learning. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning

[9] Y. W. Syaifudin *et al.*, “An implementation of recipe recommendation system based on ingredients availability using content-based and collaborative filtering,” *2023 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE)*, Oct. 2023. doi:10.1109/iccteie60099.2023.10366631

[10] Shaansubbaiah, “Shaansubbaiah/allrecipes-scraper: 🥗 scrapy spider to scrape recipe and nutritional data from allrecipes.com,” GitHub, https://github.com/shaansubbaiah/allrecipes-scraper/tree/main .