IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IDENTIFICATION AND PREDICTION OF RECIPE USING INGREDIENTS' SNAPSHOT

Ananya B[1], B Skanda Ravi Rao[2], Meghana H S[3], Vaishnavi M S[4], Mrs.Malathi.P[5],Dr Vidyarani HJ[6]

[1234]Bachelor of Engineering , Information Science , Dr. Ambedkar instiitute of Technology, Bengaluru, Karnataka, India-560056

[56]Asst. Professor, Department of ISE, Dr. Ambedkar Institute of Technology, Bengaluru, Karnataka, India-560056

Abstract: This paper proposes a deep learning-based system for automatic recipe recommendation using a single image of the available ingredients. The system leverages the strengths of two powerful convolutional neural networks (CNNs): YOLOv5 and Inceptionv5.YOLOv5, a state-of-the-art object detection model, efficiently identifies and localizes individual ingredients within the image. This provides a crucial foundation for understanding the available ingredients.Following object detection, Inceptionv5, a robust image classification model, classifies the identified ingredients into predefined categories. This allows the system to categorize the ingredients (e.g., vegetables, fruits, spices) for further processing.

With the knowledge of the identified and classified ingredients, the system predicts potential recipes. This can be achieved through techniques like nearest neighbor search in a recipe database indexed based on ingredients. This approach offers several advantages: convenience by suggesting recipes based on existing ingredients, dietary management by recommending recipes that adhere to specific needs, and reduced food waste by suggesting recipes that utilize available ingredients.

Keywords- culinary research, deep learning, feature extraction, inception v5 architecture, yolov5.

I. INTRODUCTION

Automatic recipe recommendation has become a valuable tool for simplifying meal planning and potentially aiding dietary management. This paper proposes a novel deep learning approach that overcomes limitations of existing systems. Unlike traditional methods relying on user input or browsing history, our system utilizes a single image of the available ingredients to recommend recipes.

This approach leverages the strengths of two state-of-the-art convolutional neural networks (CNNs). YOLOv5, a highly efficient object detection model, efficiently identifies and locates individual ingredients within the image. This provides a crucial understanding of the available ingredients for recipe recommendation.

Following object detection, Inceptionv5, a robust image classification model, comes into play. Inceptionv5 classifies the identified ingredients into predefined categories (e.g., vegetables, fruits, spices). This categorization allows the system to process the ingredients more effectively. With the knowledge of the identified and classified ingredients, the system predicts potential recipes. This prediction can be achieved through techniques like nearest neighbor search in a recipe database indexed based on ingredients.

Our approach offers several advantages. Users can obtain recipe suggestions based on their existing ingredients, eliminating the need for manual searching. By identifying ingredients, the system can recommend recipes that adhere to specific dietary needs. Additionally, the system can suggest recipes that utilize available ingredients, potentially minimizing food waste.

II. LITERATURE SURVEY

The quest to identify and predict recipes from ingredient snapshots is an emerging field, capitalizing on advancements in deep learning and image processing. This survey offers insights into pivotal research works in this realm, focusing on methodologies, frameworks, and practical applications pertinent to the project.

Yunus and colleagues [8] present a pioneering framework for real-time estimation of food nutritional value, employing deep learning techniques. Their approach harnesses convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to analyze food images accurately, offering valuable insights into nutritional composition and health implications.

In a recent study, Chen et al. [9] explore multi-task and region-wise deep learning strategies for food ingredient recognition. Their research delves into the complexities of ingredient identification from diverse images, emphasizing the importance of contextual information and spatial relationships. By integrating these strategies, they achieve notable enhancements in recognition accuracy and robustness.

A.Disadvantages of Existing System:

1) ***Limited Accuracy in Ingredient Recognition:*** The existing systems may be showing limited accuracy in recognizing ingredients from snapshots. Traditional image processing techniques might struggle with variations in ingredient presentation and lighting conditions, leading to misidentification.

2) ***Slow Processing Speed*:** Some existing systems might experience slow processing speeds, particularly when large datasets of ingredient images are given. This can hinder real-time prediction, making the system impractical for use in dynamic environments such as cooking applications.

3) ***Dependency on Manual Annotation:*** Certain existing systems may rely heavily on manual annotation of ingredient images, which can be can consume more time and be labor-intensive. This manual effort can limit scalability and hinder the system's ability to adapt to new ingredients or recipes efficiently.

4) ***Limited Model Flexibility:*** Existing systems might be built on rigid architectures that lack flexibility in adapting to different types of ingredients or recipes. This can result in suboptimal performance when faced with variations in ingredient composition or dish categories.

B.Advantages of Proposed System:

1) ***Enhanced Accuracy with YOLOv5 and InceptionV5:*** The proposed system leverages state-of-the-art deep learning models such as YOLOv5 and InceptionV5, which offer superior accuracy in object detection and image classification tasks. By utilizing these advanced models, the system is capable of attaining superior accuracy in ingredient recognition and recipe prediction.

2) ***Improved Processing Speed :*** YOLOv5 and InceptionV5are optimized for fast and efficient processing, enabling the proposed system to deliver real-time prediction capabilities. This ensures quick identification and prediction of recipes from ingredient snapshots, enhancing user experience and usability.

3) ***Automated Annotation and Training:*** The proposed system can automate the process of ingredient annotation and model training, reducing the need for manual intervention. By leveraging transfer learning techniques with pre-trained models like InceptionV5, the system can adapt quickly to new ingredients or recipes without extensive manual effort.

4) ***Flexibility and Scalability:*** With YOLOv5 andInceptionV5 as its backbone, the proposed system offers flexibility and scalability to accommodate a wide range of ingredients and recipe categories. These models exhibit a high degree of adaptability and can generalize well to various object detection and image classification tasks, guaranteeing reliable performance in diverse culinary settings.

III. METHODOLOGY

A. MODULES:

1) ***Data Cleaning***: Besides unraveling the hidden patterns and insights, Data Exploration allows one to take up initial steps in building a distinctly accurate model.

It also supports superior and well-curated analysis and improved business intelligence for processing and decision making. Although our core_data_recpie.csv dataset includes 1000 Multi-cuisine recipes, but still a lot of them cannot be used due to lack of appropriate quality. Many community cooking blogs contain multiple recipes of a single food dish that are largely unstructured. As part of the Data Cleaning pipeline, we have identified few images and textual information and cleaned them as follow: Instructions: Manual Investigation of several records in the given dataset, proved some users have given URLs or emoticons and other special symbols in the cooking instructions text. Ingredients: The ingredient list makes up most of the unstructured portion of the dataset. It consists of roughly around 10,000 unique Ingredients mainly because they also contain pronouns and adjectives (e.g., Turkey Black Bean Burgers instead of only Bean Burgers). Removal or replacement of Special symbols (such as @, -, *,) with blank spaces.

2) ***Image Scrapping*****:** Image scraper used certain requests library to extract food images from the specified websites. The request libraries involve Beautiful Soup and pandas to export scraped data (i.e. image URLs) and present output data into our core_data_recipe.csv file. We assigned an appropriate web driver to pick the URL from which we scraped image links and created a list data structure for storing.

3) ***Image Resolution*****:** For training an accurate model, at least four food images of the same dish with adequate resolutions are necessary. Recipes without corresponding images were also retained as they can be matched with the ones with images. We have resized the food images to 63*63*3 pixels for our model input.

4) ***Data preprocessing*****:** Importing Libraries: We begin with importing the libraries that would be required to perform certain tasks in the code.

a) ***Splitting the Dataset*****:** The entire dataset of 1000 is split into Training Set, and Validation Set, and Test Set of similar distribution. A 60- 20-20 split was performed on the given dataset into three categories as follows:

• Training Set: This contains 60% of the entire dataset, which was used to train the CNN model.

• Test Set: This contains 20% of the entire dataset, which was used in finding the accuracy and loss of the model.

• Validation Set: This contains 20% of the entire data, which was used to fine-tune the hyper parameters to find the model learning rate.

Split count of Food Images:

• Count of Training images: 600

• Count of Test images: 200

• Count of Validation images: 200

b) ***Input pipeline*****:** The input stream is nothing but the dataset start_data_recipe which loads the training samples in the form of a tuple (n, img x(ing), x(inst)) for example.

- Where n is a unique prescription number.
- img displays images of the foods in the dataset.
- x (eng) shows the list of contents and;
- x(inst) contains cooking instructions and times in text format.

The clean dataset contains 4 images of each food item. During training, an image was chanced upon from the remaining 3 images to be trained correctly. Less training leads to Overfitting (i.e. the model recognizes

images of food in the training data for example. Therefore the accuracy of the training will be higher than the accuracy of the so called test).

c) *Ingredients*: The components of the signal are specified using the word encoder and the word size. Dividing the alphabet into pieces called Tokens and dropping characters one at a time, including symbols and special characters. The special character '^' is used to separate the elements of the list in a data list. The components are placed randomly because the program does not change the learning model.

d) *Title and Cooking Instructions*: The title of the recipe is created using the main ingredients in the ingredients list. We combined the title with the cooking instructions. Cooking instructions also include the timeframe needed for completion of all instructions and the total time will be the sum of the times.

5) *Building YOLOV5*: We use version 5, released by Ultralytics in June 2020, which is currently the most advanced detection algorithm available. It is a new neural network (CNN) that detects objects accurately and in real time. This method uses a single neural network to process the entire image, then segments the image and estimates probability bounds for each segment.

These bonded boxes weigh and are expected. The method "looks only once" at the picture in the form of an expression after the single front has spread along the line. It automatically creates detected objects after a small delay (allowing the object detection algorithm to detect each object once).

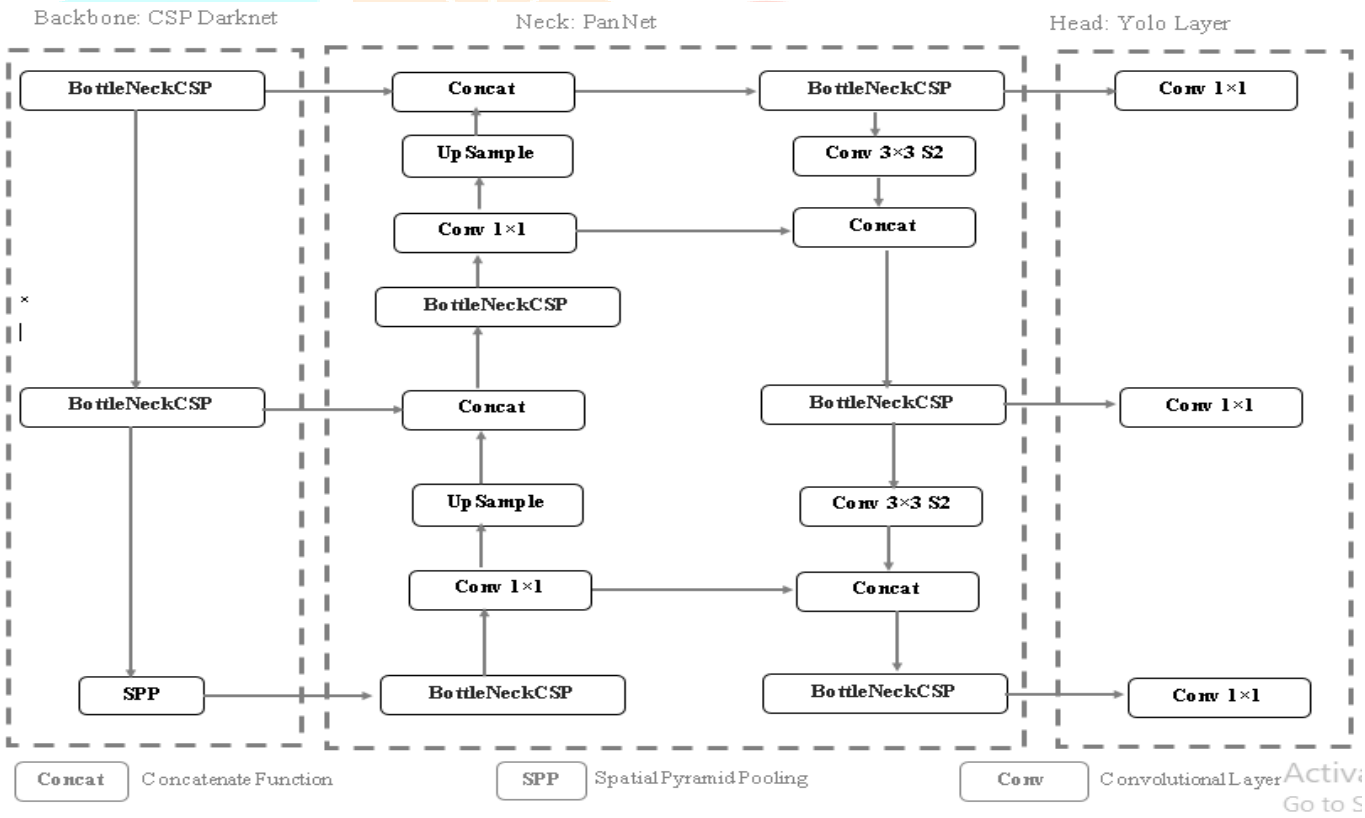

Fig 1: Block diagram of YOLOV5 Architecture

a) Backbone: The model backbone plays an important role in extracting essential components from the given image. In YOLO v5, CSP (Cross Stage Partial Networks) serves as the backbone, enabling the extraction of significant and pertinent characteristics from the input image.

b) Neck: The model neck is primarily responsible for generating feature pyramids. Feature pyramids are instrumental in helping models effectively generalize, particularly concerning object scaling. They facilitate the recognition of objects across various sizes and scales.

c) Head: The model head assumes the detection task. It utilizes anchor boxes to generate conclusive output vectors containing class probabilities, objectness scores, and bounding boxes.

***6) Cooking Instructions Decoder*:** The output of the Ingredients decoder serves as input for the cooking instruction decoder, facilitating the production of food recipes. This process is conditioned on both the embedded image (img) and the characteristics of the ingredients (x) retrieved from the Ingredients decoder. After processing, the output of the Instruction decoder undergoes further transformation through a linear layer, succeeded by the SoftMax activation function, to produce probabilities across the vocabulary of cooking instruction text.

B. *SYSTEM DESIGN*

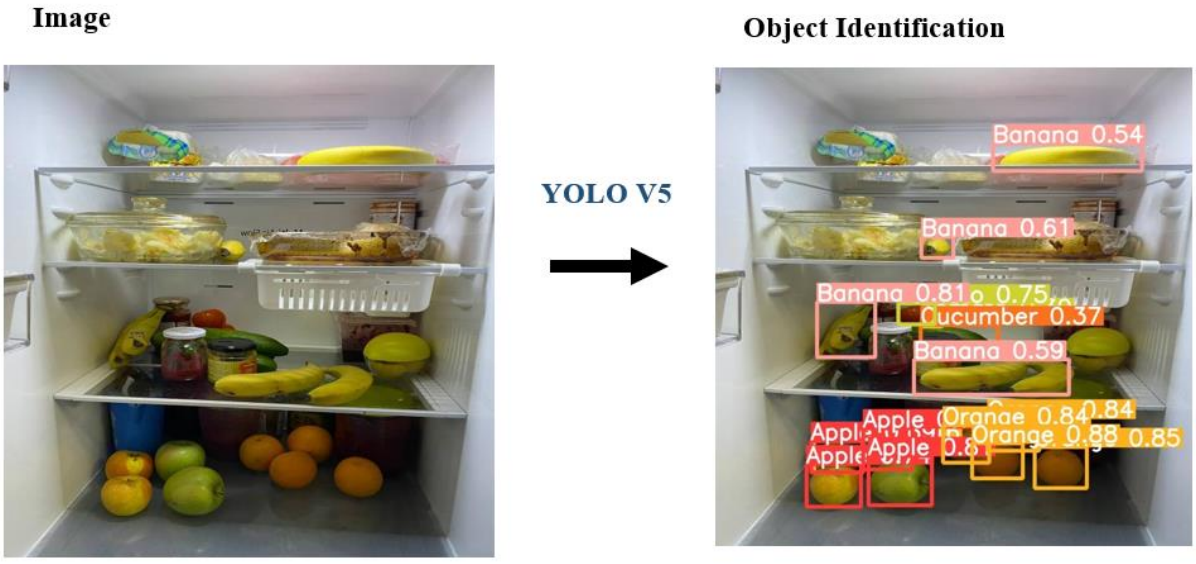

Fig 2: Object detection using YOLOV5

In the above diagram, we are passing fridge image as input system will apply modified CNN architecture to detect objects and generate the meal recipe using Inception v2 algorithm.

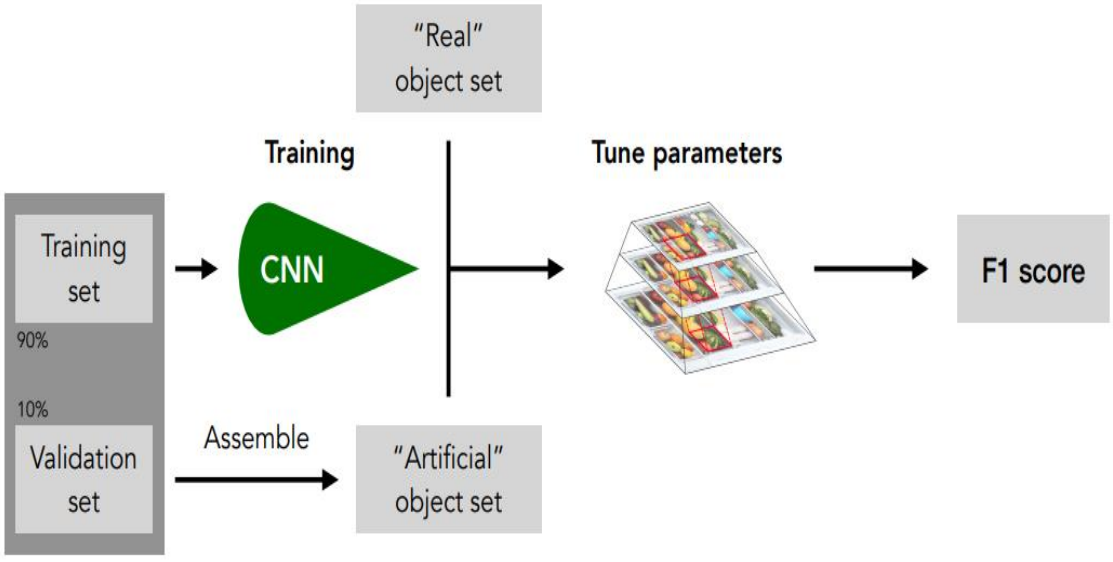

Fig. 3. Block Diagram of the Proposed Recipe Identification and Prediction System

C. CONCEPTS INVOLVED

1) *Convolution neural network*: In contrast to previous Neural Networks, Convolutional Neural Networks (CNNs) organize neurons in three dimensions: width, height, and depth. Each neuron in a CNN layer is connected only to a localized region (determined by a window size) of the preceding layer, rather than being fully connected to all neurons.

2) *Convolution layer*: In the convolution layer, a small window size, typically 5x5, is applied across the depth of the input matrix. This layer consists of learnable filters of the same window size. During each iteration, the window slides by a specified stride size, typically 1, and computes the dot product of the filter entries and input values at each position. This iterative process generates a 2-dimensional activation matrix, providing the response of that matrix at each spatial position. The network essentially learns to recognize filters that respond to particular visual characteristics, such as edges in different directions or areas with unique colors.

IV. RESULTS AND DISCUSSIONS

The study presents the performance achieved by the YOLOv5 model for ingredient detection and the deep learning model for recipe prediction. The discussion encompasses several key aspects:

- The influence of dataset size and diversity on the outcome of both models.
- The effectiveness of employing data augmentation techniques for ingredient detection within refrigerators.
- The balance between prediction accuracy and computational efficiency, particularly in real-time applications.

During the training phase, a food ingredient dataset comprising 1000 images is utilized. This dataset is divided into three segments: training, validation, and testing. The training subset consists of 600 images (60%), the validation subset contains 200 images (20%), and the testing subset also comprises 200 images (20%). The dataset encompasses 12 classes, including sprout, beef, chicken, egg, pork, garlic, onion, kimchi, potato, and spam. The training program is conducted using Python within the Jupyter environment.

The YOLOv5 model undergoes training on the food ingredient dataset for 40 epochs, with a completion time of 4 hours. Utilizing a powerful GPU could significantly reduce the training duration compared to using solely a CPU. Figure 4 illustrates the outcomes of training and validation for YOLOv5 on the food ingredient dataset. The upper row illustrates the training results, while the lower row depicts the validation outcomes. Each subgraph's horizontal axis represents the number of epochs, while the vertical axis illustrates box_loss, obj_loss, cls_loss, precision, recall, and mAP (mean average precision) sequentially.

The assessment of the model's effectiveness depends on several metrics, including precision and recall. A true positive denotes a correct detection by the model, a false positive refers to an incorrect detection, and a false negative indicates a missed detection. The model's effectiveness is assessed based on achieving high precision and recall, with a heuristic determination of the tradeoff among the two in the proposed application.

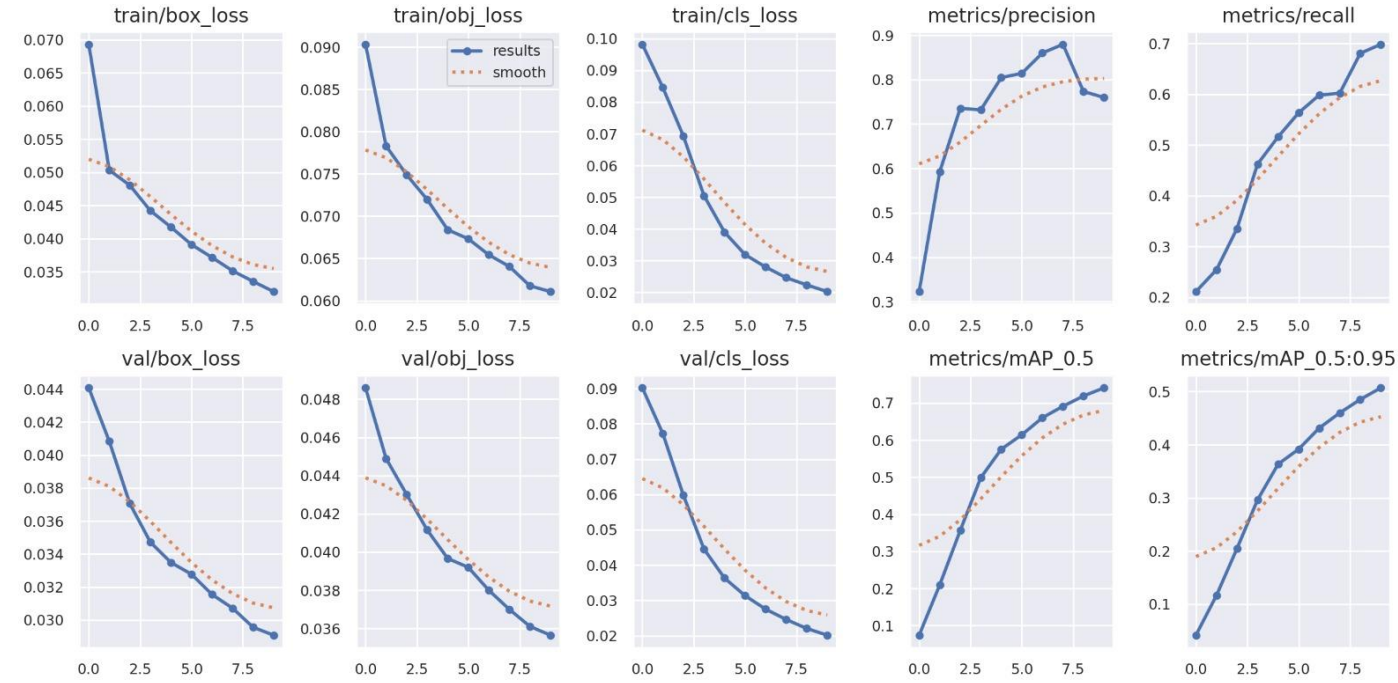

Fig 4: YOLOV5 results of training (upper row) and validation (lower row)

Figure 5 depicts the confusion matrix illustrating the recognition performance across 12 types of food ingredients. It is evident that eggs exhibit the highest detection accuracy, reaching 96%. Most other food ingredients achieve recognition accuracies well above 60%, with the exception of chicken, pork, and beef. These meat ingredients, often found in diverse shapes and packaging, tend to yield lower accuracy rates. To give a more precise evaluation of accuracy, the F-score is calculated as the harmonic mean of precision and recall. In summary, the YOLOv5 model attains an F-score of 0.97.

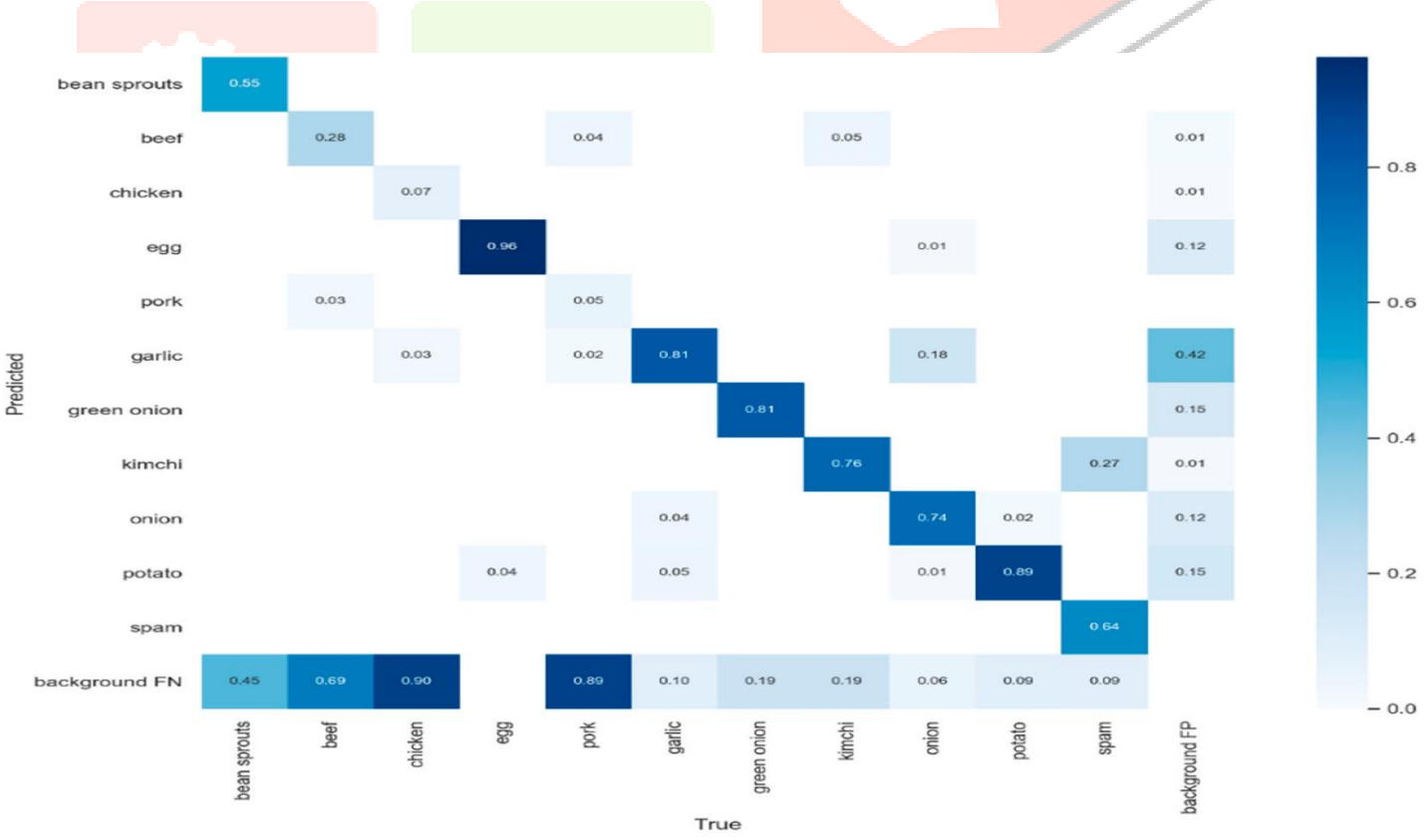

Fig 5: Recognition accuracy and confusion matrix of 12 food ingredients.

The confusion matrix analysis further highlights the model's accuracy, showing minimal misclassifications and robust detection across various food ingredients. This high precision and recall underscore the efficacy of YOLO v5 in accurately identifying food items, making it a powerful tool for real-time nutritional estimation and personalized recipe suggestions. The results demonstrate the ability of deep learning in enhancing food recognition technologies.

V. CONCLUSION

In this research, we successfully developed a robust system for the forecasting of recipes from ingredient snapshots using a union of YOLO V5 for object detection and Inception V5 for image classification. The integration of these advanced deep learning models enabled precise detection and recognition of various ingredients within a given image. Our results demonstrate that the hybrid approach of using YOLO V5's superior detection capabilities, paired with Inception V5's high classification accuracy, provides a comprehensive solution to the complex task of recipe identification.

The system was evaluated using a diverse dataset, showing promising performance metrics in both detection accuracy and classification precision. Our approach identifies individual ingredients and accurately predicts the corresponding recipes, highlighting the potential for applications in smart kitchens, dietary management, and culinary education.

Future work will focus on expanding the dataset to include a broader range of ingredients and recipes, enhancing the model's robustness. Additionally, integrating natural language processing to interpret and generate textual recipes from identified ingredients could further improve the system's usability.

Overall, this project lays a solid foundation for the development of intelligent culinary assistants and demonstrates the feasibility and effectiveness of applying state-of-the-art deep learning techniques to food-related applications.

VI. ACKNOWLEDGMENT

We are particularly grateful to Professor Malathi.P for her insightful feedback and guidance during the development of this project. We also extend our thanks to Dr. Ambedkar Institute of Technology for providing the computational resources and research ecosystem that made this project possible. Finally, we appreciate the helpful discussions and suggestions from our colleagues, which significantly contributed to the progress of this research.

VII. REFERENCES

[1] Md. Shafaat Jamil Rokon et al., "Food Recipe Recommendation Based on Ingredients Detection Using Deep Learning," arXiv preprint arXiv:2203.06721, 2022.

[2] M. Kumari and T. Singh, "Food Image to Cooking Instructions Conversion Through Compressed Embeddings Using Deep Learning,"

[3] L. Harnack, L. Steffen, D. Arnett, S. Gao, and R. Luepker, "Accuracy of estimation of large food portions," *Journal of the American Dietetic Association*, vol. 104, no. 5, pp. 804–806, 2004.

[4] R. M. Rahul Venkatesh Kumar, M. Anand Kumar, and K. P. Soman, "Cuisine Prediction based on Ingredients using Tree Boosting Algorithms," *Indian Journal of Science and Technology*, vol. 9, no. 45, DOI: 10.17485/ijst/2016/v9i45/106484, Dec. 2016.

[5] W. Min, S. Jiang, S. Wang, R. Xu, Y. Cao, L. Herranz, and Z. He, "A survey on context-aware mobile visual recognition," *Multimedia Systems*, vol. 23, no. 6, pp. 647–665, 2017.

[6] Y. Liu, J. Zhang, and Y. Li, "Deep Learning-based Target Detection and Recognition using YOLO V5," in Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), *2022*.

[7] M. Mirzaei, R. Sharma, and M. Jain, "Personal Protective Equipment Kit Detection using YOLO v5 and MobileNet SSD," in Proceedings of the 2022 International Conference on Smart City and Green Energy (ICSCGE), *2022*.

[8] R. Yunus, O. Arif, H. Afzal, M. F. Amjad, H. Abbas, H. N. Bokhari, S. T. Haider, N. Zafar, and R. Nawaz, “A framework to estimate the nutritional value of food in real time using deep learning techniques,” *IEEE Access*, vol. 7, pp. 2643–2652, 2019.

[9] J. Chen, B. Zhu, C. W. Ngo, T. S. Chua, and Y. G. Jiang, "A study of multi-task and region-wise deep learning for food ingredient recognition," *IEEE Transactions on Image Processing*, vol. 30, pp. 1514-1526, 2021.