Bangladeshi Street Food Calorie Estimation Using Improved YOLOv8 and Regression Model

Aparup Dhar[1]*, MD Tamim Hossain[1]*, Pritom Barua[1]
[1]*Department of Computer Science & Engineering, Premier University, Chittagong, Bangladesh*
iaparup@gmail.com, thossain3333@gmail.com, pritombarua71@gmail.com

***Abstract*—As obesity rates continue to increase, automated calorie tracking has become a vital tool for people seeking to maintain a healthy lifestyle or adhere to a diet plan. Although numerous research efforts have addressed this issue, existing approaches often face key limitations, such as providing only constant caloric output, struggling with multiple food recognition challenges, challenges in image scaling and normalization, and a predominant focus on Western cuisines. In this paper, we propose a tailored solution that specifically targets Bangladeshi street food. We first construct a diverse dataset of popular street foods found across Bangladesh. Then, we develop a refined calorie estimation system by modifying the state-of-the-art vision model YOLOv8. Our modified model achieves superior classification and segmentation results, with only a slight increase in computational complexity compared to the base variant. Coupled with a machine learning regression model, our system achieves an impressive 6.94 mean absolute error (MAE), 11.03 root mean squared error (RMSE), and a 96.0% R^2 score in calorie estimation, making it both highly effective and accurate for real-world food calorie calculations.**

***Index Terms*—Deep learning, Attention Module, Street Food, Calorie Estimation, Machine Learning, Regression**

I. Introduction

Obesity is a major public health problem affecting a large portion of the world's population. It is a disease that is characterized by the accumulation of excessive or abnormal amounts of body fat, causing possible negative effects on the health of an individual. At a global scale, obesity rates have doubled since the year 1990, while as of 2022, approximately 890 million adults can be classified as obese [1]. As such, obesity is now considered and regarded as a global epidemic that goes beyond traditional social and economic boundaries, as it disproportionately affects both lower and middle-income countries.

As a developing country, Bangladesh is currently facing this problem. The country was once synonymous with food insecurity, but now deals with a concurrent under-nutrition and over-nutrition situation. Reports from the Bangladesh Demographic and Health Survey (BDHS) found that 24% of women and 16% of men are overweight or obese, which is a staggering threefold increase since the year 2000 [2]. In urbanized areas, where the majority of people lead a sedentary lifestyle and regularly consume street food on a daily basis, obesity rates surpass 30% [3]. Processed snacks, fried foods, and sugary drinks now account for 40% of urban calorie consumption [4].

Effective obesity prevention requires accurate calorie monitoring, helping users to keep track of their food intake. The Centers for Disease Control and Prevention (CDC) recommends keeping track of food and beverage intake in order to ensure the amount of calories does not exceed the body's daily needs [5].

A lot of progress has been made on research relating to food calorie estimation. With the advancements in computer vision and deep learning technologies, novel methods have been applied in order to achieve accurate calorie measurements. Although many solutions exist, most are focused on Western cuisine, only providing a constant caloric value for food items without considering actual portions of the food. As of now, no work specifically addresses this problem in the context of Bangladeshi food, specifically street food, which is consumed by a major part of the working-class population.

In this paper, we introduce a solution to accurately estimate the calorie content of Bangladeshi street food from a singular image, which considers the actual food portion and weight by using a coin as a reference object to determine the real-world dimensions of food. Additionally, we modify the base YoloV8 state-of-the-art object detection and segmentation model by introducing Coordinate Convolution and Convolutional Block Attention Module to achieve improved performance on our dataset, specifically constructed for this research based on the most popular street foods of Bangladesh.

II. Literature Review

Recent years have seen extensive research on food calorie estimation, with many approaches leveraging convolutional neural networks (CNNs) for food detection and classification. Jiang et al. [6] proposed a deep CNN using a Region Proposal Network from Faster R-CNN and VGGNet for feature extraction and classification. Although efficient due to a regression module for localization, their assumption of a constant 400g weight per item undermines accuracy due to variable portion sizes.

Mask R-CNN [7] has also been widely adopted for object detection and segmentation in calorie estimation, with studies [8], [9] combining segmentation with linear regression for calorie prediction. However, these systems often depend on strict image capture protocols (e.g., fixed height), limiting usability.

To address scaling issues, Akpro et al. [10] introduced chopsticks as a reference object to estimate food volume.

* These authors contributed equally to this work

Despite being practical, this limits applicability to East Asian cuisine and requires manual food classification. SCalE [11] estimated volume using distance-based calculations, pixel-to-centimeter ratios, and K-means clustering for segmentation. However, the system's reliance on user-specific measurements (e.g., arm length) and camera angles reduces flexibility and introduces potential errors.

A novel method by [12] utilized a mobile device's microphone and speaker with MLS signals for non-invasive vertical distance measurement. This two-image (top and side view) approach performed well but was sensitive to noise and required food to be in containers.

Steinbrener et al. [13] proposed using monocular video and smartphone inertial data for metric volume estimation, combining Mask R-CNN outputs with LSTM-based processing. While achieving competitive accuracy, the method's reliance on video input and lack of real-world robustness evaluation are noted limitations.

Finally, NIRSCam [14] introduced a mobile near-infrared sensing system for calorie estimation, outperforming image-based methods in accuracy and environmental robustness. It uses commercial LEDs for NIRS and three ML models to estimate macronutrients. However, the need for external hardware and high computational complexity may hinder widespread adoption.

III. Methodology

The proposed food calorie estimation system comprises three main components. First is the classification and segmentation module, which detects and identifies food items while segmenting them from the background. Next, the scaling algorithm adjusts the food item dimensions based on a reference object to determine real-world measurements. Finally, the calorie estimation model employs machine learning regression techniques to predict calorie values using the extracted features. Each component is detailed below.

A. Classification & Segmentation Model

The proposed system takes a single RGB image as input and performs classification and instance segmentation on each detected food item. Instance segmentation generates binary masks, which are later used for feature extraction. For this task, we build upon the state-of-the-art YOLOv8 model [15], an evolution of the widely-used YOLOv5 architecture.

YOLOv8 introduces several key improvements over its predecessor, most notably anchor-free detection, which enables direct object localization, leading to faster training, more precise detection, and better performance. It also features an enhanced backbone based on a refined CSPNet architecture [16], improving feature extraction capabilities. As a single-stage object detection and segmentation model, YOLOv8 performs predictions in one pass, making it more efficient and less computationally intensive than two-stage models like Mask R-CNN.

Studies [17], [18] have shown that YOLOv8 outperforms Mask R-CNN in terms of segmentation precision, recall, and inference speed—making it ideal for real-time applications. In our implementation, we enhance the base YOLOv8 by replacing the standard "C2f" module in the head with a custom "C2f_CD" module, incorporating Coordinate Convolutional layers [19] and a Convolutional Block Attention Module (CBAM) [20]. These additions aim to improve the detection of irregularly shaped food items and those with context-dependent appearances. Further architectural details are discussed in the following section.

B. Scaling Algorithm

Accurate food calorie estimation requires determining real-world food dimensions, but images captured from different angles and distances can distort these measurements. To address this, our system uses a reference object of known size to compute a scaling factor that converts image measurements into real-world units. We propose a standard coin as the reference, chosen over alternatives like teaspoons, photographic scales, or hands, due to its accessibility, consistency, and ease of use. In our implementation, a Bangladeshi 5 Taka coin (25.5 mm diameter) is detected in the image, its pixel size measured, and a scaling factor derived. This factor is then applied to segmented food items to obtain real-world dimensions, with the calculation method detailed in the next section.

$$\text{Height of Food Item} = F_h \text{ px}$$
$$\text{Width of Food Item} = F_w \text{ px}$$
$$\text{Diameter of Ref Coin} = D_c = 25.5 \text{ mm}$$

Scale factors for height and width are given by:

$$\text{Scale Factor (Height)}, S_h = \frac{D_c}{F_h} \tag{1}$$

$$\text{Scale Factor (Width)}, S_w = \frac{D_c}{F_w} \tag{2}$$

Thus, the overall scale factor is:

$$S_f = \frac{S_h + S_w}{2} \tag{3}$$

C. Estimation / Regression Model

The final step in the system is calorie estimation, which is performed using a machine learning regression model. Once the binary mask of the food item is obtained from the segmentation model, we extract key features—height, width, area, perimeter, and class label—using OpenCV's contour feature functions [21]. These features are then scaled using the previously calculated scaling factor to reflect real-world dimensions. The scaled features are organized into a structured data frame and passed into the regression model, which outputs a continuous value representing the estimated calorie content of the food item.

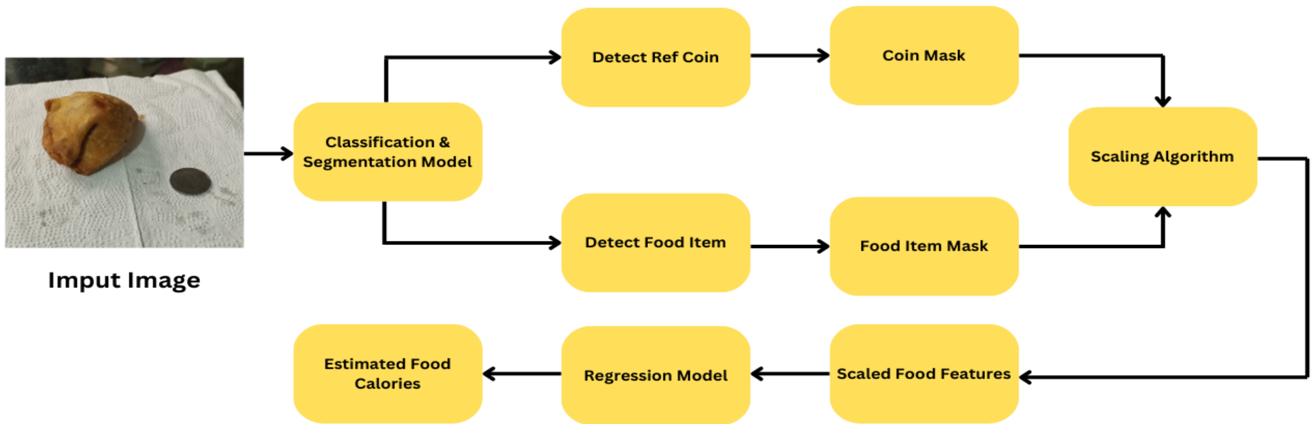


Fig. 1: Proposed System For Calorie Estimation

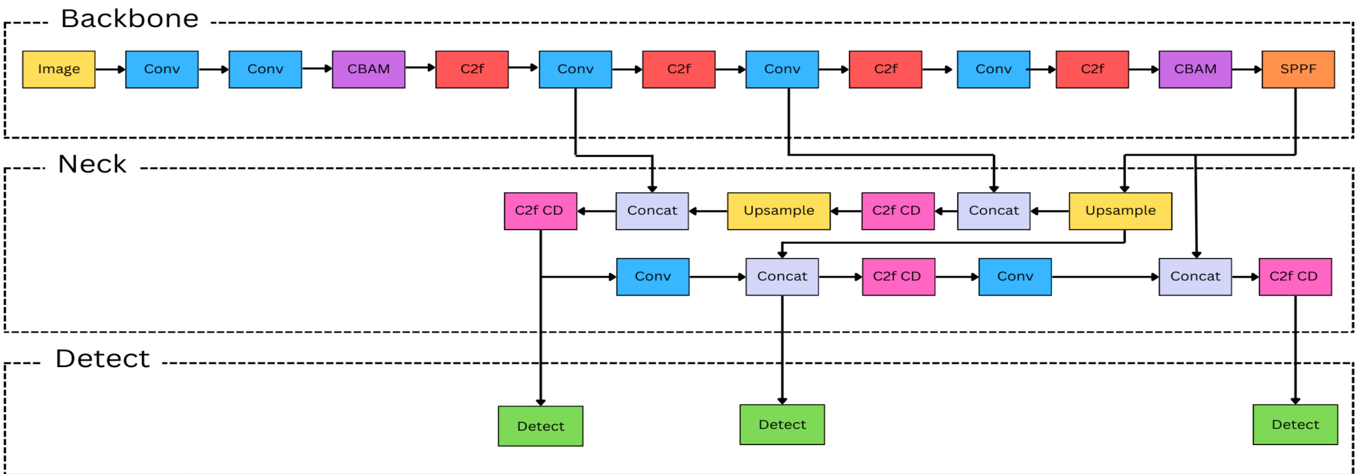


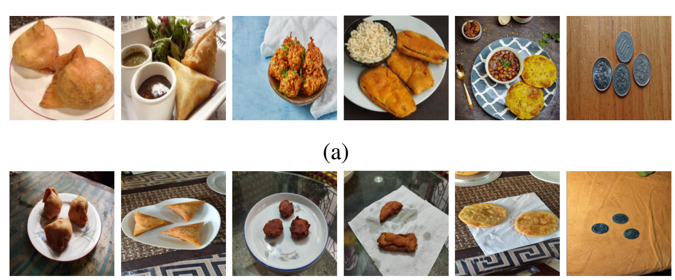
Fig. 2: Improved YoloV8 Model Architecture

IV. Dataset Description

For the proposed system, we require two datasets. One for training the classification, segmentation model, and a separate dataset with food items alongside a reference object coin for constructing our regression dataset. Both datasets are separately discussed in detail below.

A. Classification/Segmentation Dataset

Although many datasets exist for Western and Asian foods, to our knowledge, there is no dataset focused on Bangladeshi cuisine that includes instance segmentation masks alongside a reference object. To address this, we created a dataset based on popular Bangladeshi street foods [22], [23], comprising six classes: Singara, Somusa, Puri, Peaju, Beguni, and a reference object class for the coin. Images were collected both from the web and onsite, ensuring diversity. Web images offer a wide variety of environments and contexts, while physical data captures real-world challenges like poor lighting, occlusions, and unusual angles, improving the generalization ability of the segmentation and classification model.



(a)

(b)

Fig. 3: Sample Images From Dataset (a) Images Collected From Web (b) Images Collected Onsite

The base dataset contains 1,847 images, which were expanded through augmentation to 3,885 images. These are split

80% for training (3,146 images), 10% for validation (370 images), and 10% for testing (369 images). The total number of instances per class is shown in Table I, and sample images are presented in Fig. 3

TABLE I: Class Instance Distribution

Class	Train	Test	Valid
Singara	467	172	136
Somusa	567	185	173
Puri	415	149	144
Peaju	764	271	287
Beguni	603	207	203
Coin	469	162	161

B. Estimation / Regression Dataset

For the regression dataset, we prepared a separate set of images that were passed through the trained model to extract features like height, width, area, perimeter, and class. The dependent variable, calories, is calculated by multiplying the food's weight (measured in grams with a digital scale) by its caloric density. The calorie calculation formula is shown below.

$$C = W \times D \tag{4}$$

where:

$C =$ Calories

$W =$ Weight of food (grams)

$D =$ Caloric density of food (kcal per gram)

The caloric density of every food item used in the research is determined according to an online nutritional database [24]. The specific values are shown in Table II below. Caloric density has been considered for standard variants of each food item. For example, in the case of Singara, commonly used ingredients such as potato and peas have been considered as a baseline. Alternative variants using meat or other ingredients have not been considered in this analysis.

TABLE II: Calorie Density Values By Class

Class	Calorie Density
Singara	2.61
Somusa	2.11
Puri	2.44
Peaju	1.18
Beguni	1.48

Each food item was placed beside the reference coin and photographed from 10 different angles to capture visual variations. Although the appearance changes, the calorie content remains constant. Fig. 4 illustrates example images of a single food item from multiple views.

After passing each image through our scaling algorithm, we obtain the features, which are appended with caloric values in a comma-separated values (CSV) file. After preprocessing, we obtain our regression dataset comprising 644 records. The distribution of classes in the regression dataset is shown below in Table III.

Fig. 4: Example of Multiple Images Taken From Different Angles of a Singular Food Item

TABLE III: Sample Image Distribution

Class	Records
Singara	128
Somusa	131
Puri	116
Peaju	133
Beguni	136

V. Preprocessing Steps

Image preprocessing enhances quality and standardizes input for deep learning. To improve generalization and mitigate overfitting, training data is augmented using horizontal flips and 90° rotations, increasing the number of training samples from 1,847 to 3,146 images. Additionally, all images are resized to 640×640 pixels to ensure a consistent input shape, reduce computational load, and avoid feature distortion.

For regression modeling, the data undergoes further preprocessing. Categorical class names are converted into a numeric format using one-hot encoding, where each class is represented as a binary vector. Feature scaling is applied through min-max normalization to ensure that features with larger magnitudes, such as area or perimeter, do not dominate the learning process. Finally, Z-score normalization is used to eliminate outliers from mispredictions or segmentation errors, with a threshold set at 2.

VI. Performance Metrics

For Bangladeshi street food calorie estimation, the performance metrics are grouped into two categories: those evaluating classification and segmentation, and those evaluating regression-based estimation.

A. Classification & Segmentation Metrics

For classification and segmentation evaluation, we employ Precision, Recall, and mean Average Precision (mAP) to measure detection and segmentation accuracy across different thresholds. In addition, we report GFLOPs to quantify the computational cost of the models.

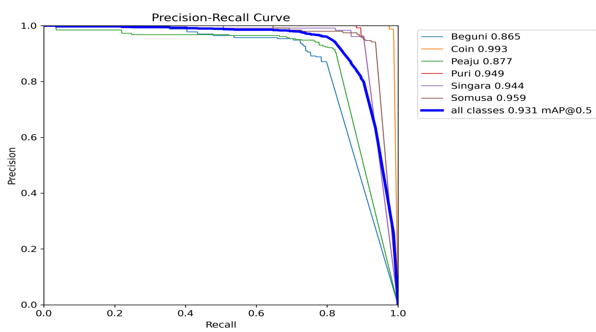
B. Regression Metrics

For calorie estimation, we use Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to measure prediction accuracy, while the R^2 score is used to assess the proportion of variance explained by the regression model.

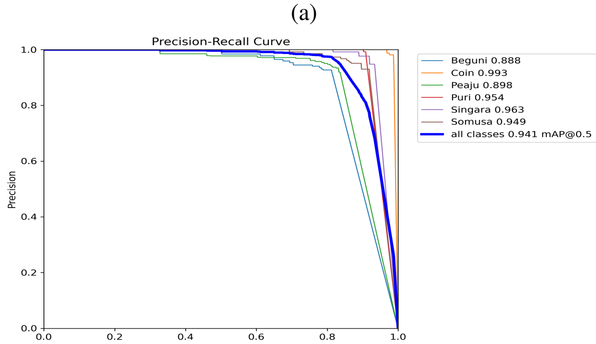
VII. Results

Table IV reveals a clear gap between Mask R-CNN and YoloV8, with the former trailing notably in mAP50-90. Adding CBAM or CoordConv alone to YoloV8n yields mixed results—CBAM slightly boosts box precision (+0.2%) but hurts mAP50-90, while CoordConv underperforms across all metrics. However, combining both leads to consistent gains: +0.9% precision, +0.8% recall, +1.0% mAP@50, and +0.6% mAP@50-95 over the base model. Mask metrics also improve, with +0.6% precision, +0.7% recall, +0.9% mAP@50, and a modest +0.1% mAP@50-95. Overall, the enhanced YoloV8n outperforms its base and other leading models in classifying and segmenting Bangladeshi street foods.

A. Bounding Box Precision Recall Curves



(a)

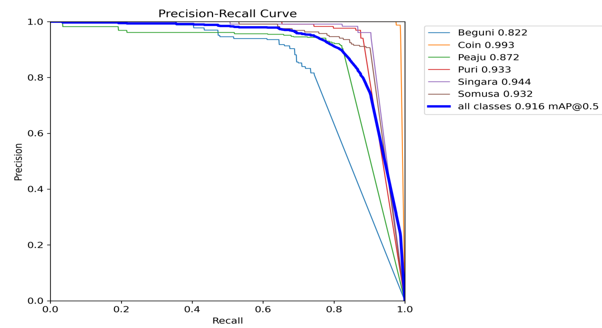


(b)

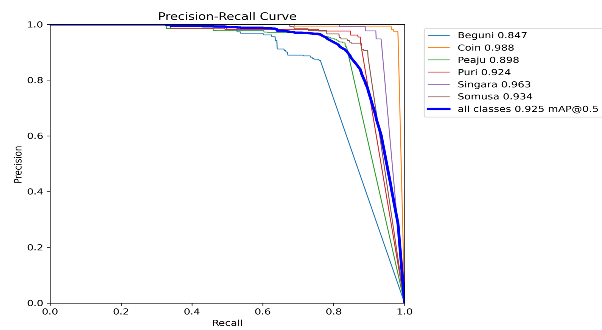
Fig. 5: Bounding Box Precision Recall Curve (a) Improved YoloV8n (b) Base YoloV8n

The Precision-Recall (PR) curves in Fig. 5 compare Improved YOLOv8n (a) with Base YOLOv8n (b) for food classification. The improved model achieves a slightly higher mAP@0.5 of 94.1% versus 93.1%, indicating an overall boost in precision and recall. Class-wise gains are also evident: Beguni improves from 86.5% to 88.8%, and Peaju from 87.7% to 89.8%. While some classes like Coin retain consistent high performance (99.3%), others show minor shifts—Singara improves to 96.3% from 94.4%, while Somusa drops slightly to 94.9% from 95.9%.

B. Mask Precision Recall Curves



(a)



(b)

Fig. 6: Mask Precision Recall Curve (a) Improved YoloV8n (b) Base YoloV8n

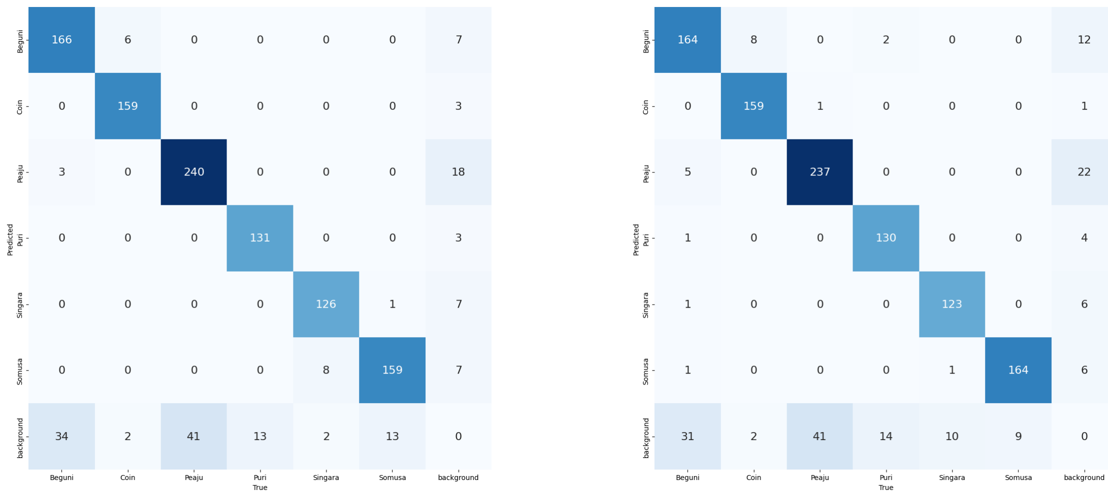
The mask Precision-Recall (PR) curves illustrate segmentation performance for both the improved YOLOv8n (a) and base YOLOv8n (b). The base model achieves a strong mAP@0.5 of 91.6%, with class-wise precision-recall scores like Beguni (82.2%), Coin (99.3%), Peaju (87.2%), Puri (93.3%), Singara (94.4%), and Somusa (93.2%) showing solid consistency. The improved YOLOv8n slightly outperforms it with a mAP of 92.5%, reflecting marginal but consistent gains across most categories: Beguni (84.7%), Coin (98.8%), Peaju (89.8%), Puri (92.4%), Singara (96.3%), and Somusa (93.4%). Overall, the improved model demonstrates more balanced, reliable segmentation across all classes.

C. Confusion Matrices Comparison

Comparison of the confusion matrices reveals that the improved YOLOv8n substantially reduces background hallucinations and false detections, indicating better discrimination between food items and irrelevant scene elements. It also improves intra-class separation, lowering confusion between visually similar items like Singara vs. Somusa or Beguni vs. Peaju. The improved model shows fewer misclassifications overall, evidenced by stronger diagonal dominance, confirming higher accuracy and robustness.

TABLE IV: Performance Metrics Comparison

Model	Bounding Box				Mask			
	P	**R**	**mAP50**	**mAP50-95**	**P**	**R**	**mAP50**	**mAP50-95**
Mask RCNN (ResNet50, C4)	90.1%	81.6%	90.1%	76.7%	89.4%	79.2%	89.3%	74.4%
Mask RCNN (ResNet50, DC5)	91.2%	82.4%	91.1%	78.2%	91.2%	81.9%	91.2%	77.7%
Mask RCNN (ResNet50, FPN)	90.5%	81.7%	90.5%	77.6%	90.8%	82.4%	90.8%	78.5%
YoloV8n	93.5%	89.2%	93.1%	86.2%	91.7%	87.4%	91.6%	81.1%
YoloV8n + CBAM	93.7%	88.9%	93.1%	85.8%	91.5%	86.8%	91.7%	80.6%
YoloV8n + CoodConv	93.4%	87.8%	92.5%	85.9%	91.7%	86.2%	91.3%	80.9%
YoloV8n + CBAM + CoodConv	**94.4%**	**90%**	**94.1%**	**86.8%**	**92.3%**	**88.1%**	**92.5%**	**81.2%**



(a) (b)

Fig. 7: Confusion Matrix Comparison(a) Improved YoloV8n (b) Base YoloV8n

D. Computational Complexity Comparison

Compared to Mask R-CNN variants, YOLOv8n models show much lower computational complexity, reflected in fewer GFLOPs, parameters, faster inference, and smaller model size.

TABLE V: Comparison of Models by Computational Complexity and Size

Model	GFLOPs	Params (M)	Time (ms)	Size (MB)
M-RCNN (R50, C4)	822	34	248.24	420.51
M-RCNN (R50, DC5)	344	166	149.83	2060.00
M-RCNN (R50, FPN)	447	44	58.16	526.23
YoloV8n (Base)	12.0	3.2	3.8	6.80
YoloV8n (Imp)	16.3	3.6	5.0	7.63

Mask R-CNN GFLOPs range from 344 to 822, with inference times up to 248 ms, making them unsuitable for real-time use. In contrast, base YOLOv8n has just 12 GFLOPs, 3.2M parameters, and a 3.8 ms inference time, making it far more efficient. Its model size is only 6.8 MB, compared to over 500 MB for Mask R-CNN. The improved YOLOv8n sees a slight rise in GFLOPs to 16.3 and parameters but still maintains fast inference around 5 ms, preserving real-time capability.

E. Rregression / Estimation Model Results

The performance evaluation of various machine learning models for calorie estimation shows that ensemble and non-linear models outperform traditional linear regression. Random Forest stands out with the lowest MAE of 6.94, RMSE of 11.03, and the highest R^2 of 96%, indicating accurate predictions with minimal errors.

TABLE VI: Performance Metrics of Machine Learning Models

Model	MAE	MSE	R^2 Score
Linear Regression	12.40	304.40	91.0%
KNN	7.62	140.50	95.0%
Decision Tree C	8.19	199.07	94.0%
Random Forest	**6.94**	**121.80**	**96.0%**
Gradient Boost	7.84	147.84	95.0%
AdaBoost	8.10	142.75	95.0%

This highlights the advantage of ensemble learning in capturing complex patterns and reducing overfitting. K-Nearest Neighbors, Gradient Boost, and AdaBoost also perform well, each achieving around 95% R^2, demonstrating their ability to model underlying data relationships. The Decision Tree performs moderately with 94% R^2 but has a higher RMSE (14.10), suggesting more susceptibility to overfitting. Linear

Regression fares worst, with the highest MAE (12.40) and RMSE (17.44), showing its limitations in modeling non-linear dependencies typical in calorie estimation. Overall, ensemble methods like Random Forest and boosting approaches are best suited for accurate calorie prediction, balancing bias and variance while effectively handling complex food feature interactions.

VIII. Conclusion and Future Work

This paper presents a novel approach for accurately estimating calories in Bangladeshi street food, addressing the lack of region-specific nutritional data. We created a diverse dataset of 3,885 images sourced both online and onsite, covering popular local street foods to ensure real-world relevance. Building on this, we developed an improved YoloV8 model that outperforms the base YoloV8 and other leading object detectors, delivering higher accuracy even in cluttered scenes typical of street food environments. To enhance calorie prediction, we employ a regression-based method that significantly lowers error margins, making the system not only accurate but also practical for real-world applications like mobile apps and automated dietary tracking.

Looking ahead, a major challenge is scaling food items without relying on a reference object like a coin or plate. Future work will explore techniques such as depth estimation from images to infer true food sizes without needing a reference. While our dataset covers many popular Bangladeshi street foods, it lacks numerous traditional and regional dishes. Expanding the dataset to include more cuisine-specific items will improve the system's versatility and reach. Additionally, optimizing the model for computational efficiency is vital to enable smooth performance on low-end mobile devices, increasing accessibility and impact, especially in areas with limited access to high-end technology.

References

[1] World Health Organization: WHO, "Obesity and overweight." https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight, 2024. Accessed: 2025-01-10.

[2] Bangladesh Demographic and Health Survey, "Bangladesh - demographic and health survey 2022." https://microdata.worldbank.org/index.php/catalog/6290text=The%202022%20Bangladesh%20Demographic%20and,and%20their%20families%20in%20Bangladesh, 2024. Accessed: 2025-02-10.

[3] National Institute of Population Research and Training (NIPORT), "Urban health survey report 2021." https://niport.gov.bd/site/publications/84bdcfef-8aff-468b-8226-ccbf1842e9e8/Urban-Health-Survey-Report-2021, 2023. Accessed: 2025-02-10.

[4] FAO, *Dhaka Urban Food System Atlas 2023*. Dhaka, Bangladesh: FAO, 2023.

[5] Centers for Disease Control and Prevention, "About healthy weight and growth." https://www.cdc.gov/healthy-weight-growth/about/index.html, 2024. Accessed: 2025-01-14.

[6] L. Jiang, B. Qiu, X. Liu, C. Huang, and K. Lin, "Deepfood: food image analysis and dietary assessment via deep model," *IEEE Access*, vol. 8, pp. 47477–47489, 2020.

[7] K. Dollár, R. Girshick, *et al.*, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[8] M.-L. Chiang, C.-A. Wu, J.-K. Feng, C.-Y. Fang, and S.-W. Chen, "Food calorie and nutrition analysis system based on mask r-cnn," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, pp. 1721–1728, IEEE, 2019.

[9] N. Aditama and R. Munir, "Indonesian street food calorie estimation using mask r-cnn and multiple linear regression," in *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, pp. 1–6, IEEE, 2022.

[10] E. A. H. Akpa, H. Suwa, Y. Arakawa, and K. Yasumoto, "Smartphone-based food weight and calorie estimation method for effective food journaling," *SICE Journal of Control, Measurement, and System Integration*, vol. 10, no. 5, pp. 360–369, 2017.

[11] N. Sadeq, F. R. Rahat, A. Rahman, S. I. Ahamed, and M. K. Hasan, "Smartphone-based calorie estimation from food image using distance information," in *2018 5th International Conference on Networking, Systems and Security (NSysS)*, pp. 1–8, IEEE, 2018.

[12] J. Gao, W. Tan, L. Ma, Y. Wang, and W. Tang, "Musefood: Multi-sensor-based food volume estimation on smartphones," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 899–906, IEEE, 2019.

[13] J. Steinbrener, V. Dimitrievska, F. Pittino, F. Starmans, R. Waldner, J. Holzbauer, and T. Arnold, "Learning metric volume estimation of fruits and vegetables from short monocular video sequences," *Heliyon*, vol. 9, no. 4, 2023.

[14] H. Hu, Q. Zhang, and Y. Chen, "Nirscam: A mobile near-infrared sensing system for food calorie estimation," *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 18934–18945, 2022.

[15] R. Varghese and M. Sambath, "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International conference on advances in data engineering and intelligent computing systems (ADICS)*, pp. 1–6, IEEE, 2024.

[16] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 390–391, 2020.

[17] R. Sapkota, D. Ahmed, and M. Karkee, "Comparing yolov8 and mask rcnn for object segmentation in complex orchard environments," *arXiv preprint arXiv:2312.07935*, 2023.

[18] J. C. Camacho and M. E. Morocho-Cayamcela, "Mask r-cnn and yolov8 comparison to perform tomato maturity recognition task," in *Conference on Information and Communication Technologies of Ecuador*, pp. 382–396, Springer, 2023.

[19] R. Liu, J. Lehman, P. Molino, F. Petroski Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," *Advances in neural information processing systems*, vol. 31, 2018.

[20] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

[21] OpenCV, "Contour features." https://docs.opencv.org/4.x/dd/d49/tutorial_py_contour_features.html. Accessed: 2024-08-20.

[22] S. R. Choudhury, A. A. Shamim, N. Shaheen, N. Bahar, S. Sarwar, M. M. Ashraf, M. A. Al Mamun, S. M. M. Sobhan, M. J. Abedin, K. Hossen, *et al.*, "Identification of frequently consumed commercially prepared ready-to-eat foods and beverages in bangladesh: Identification of frequently consumed rte foods," *Bioresearch Communications*, vol. 7, no. 2, pp. 1019–1030, 2021.

[23] N. Islam, M. S. Arefin, T. Nigar, S. N. Haque, K. I. Haq, M. T. A. Emran, and T. Nazrul, "Street food eating habits in bangladesh: a study on dhaka city," *Int. Journal of Management and Development Studies*, vol. 6, no. 9, pp. 49–57, 2017.

[24] Nutritionix, "Nutrition information and food database." https://www.nutritionix.com/.