JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 8, No. 4. September 2026 DOI: https://doi.org/10.34288/jri.v8i4.653

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

APPLICATION OF DEEP LEARNING FOR NUTRITIONAL SEGMENTATION AND INFORMATION EXTRACTION IN THE HEALTHIER APPLICATION

Tata Aditya Pamungkas[-1*], Nuke Lu'lu Ul Chusna[-2], Harry Dwiyana Kartika[-3]

Informatics Engineering
Universitas Krisnadwipayana, Jakarta, Indonesia
tataadityapa@gmail.com[-1*)], nukelchusna@unkris.ac.id[-2], harry_d@unkris.ac.id[-3]

(*) Corresponding Author

Abstract

Low nutritional literacy and the rising burden of non-communicable diseases (NCDs) motivate Healthier, an offline application integrating food-image segmentation, nutrition-label extraction, and a deterministic Clinical Rules Engine with Human-in-the-Loop (HITL) verification. The system uses YOLO11s-seg to identify five *Isi Piringku* groups, PaddleOCR (PP-OCRv4) with grid-based parsing, and SAFE/WARNING/DANGER feedback. Trained on 11,684 images and validated on 2,017, the segmentation model achieved mask mAP50 of 71.66% and mAP50-95 of 56.91%. External testing on 5,013 images yielded 69.15% on UECFoodPIXCOMPLETE, 63.66% on FoodSeg103, and 17.06% on a fruit-only dataset, indicating domain shift. On 131 Indonesian nutrition-label images, extraction achieved 42.65% field-level accuracy, compared with 35.11% for raw PaddleOCR and 14.89% for Tesseract. Grid reconstruction improved accuracy by 13.05 percentage points, while semantic filtering reduced performance. The rules engine passed all 66 checks, but extraction errors altered 15.0% of 786 clinical decisions, increasing to 21.4% for images containing nutrition tables. Additionally, 18 of 25 ground-truth DANGER decisions became SAFE because missing fields defaulted to zero, supporting the need for human verification. This study presents a reproducible system integration and decision mechanism with cross-source evaluation and an open evaluation protocol.

Keywords: Instance Segmentation; OCR; YOLO11s-seg; PaddleOCR; Nutrition Literacy

Abstrak

Rendahnya literasi gizi dan meningkatnya beban penyakit tidak menular (PTM) mendorong pengembangan Healthier, sebuah aplikasi offline yang mengintegrasikan segmentasi citra makanan, ekstraksi label gizi, dan Clinical Rules Engine deterministik dengan verifikasi Human-in-the-Loop (HITL). Sistem ini menggunakan YOLO11s-seg untuk mengidentifikasi lima kelompok Isi Piringku, PaddleOCR (PP-OCRv4) dengan parser berbasis grid untuk membaca tabel gizi, serta menghasilkan umpan balik SAFE/WARNING/DANGER. Model segmentasi dilatih menggunakan 11.684 citra dan divalidasi dengan 2.017 citra, menghasilkan mask mAP50 sebesar 71,66% dan mAP50-95 sebesar 56,91%. Pengujian eksternal terhadap 5.013 citra menghasilkan performa 69,15% pada UECFoodPIXCOMPLETE, 63,66% pada FoodSeg103, dan 17,06% pada dataset buah di lingkungan nyata, yang menunjukkan adanya pergeseran domain. Pada 131 foto label gizi produk Indonesia, sistem ekstraksi mencapai akurasi tingkat kolom sebesar 42,65%, dibandingkan PaddleOCR tanpa pemrosesan tambahan (35,11%) dan Tesseract (14,89%). Rekonstruksi grid meningkatkan akurasi sebesar 13,05 poin persentase, sedangkan filter semantik justru menurunkan performa. Clinical Rules Engine berhasil memenuhi seluruh 66 pengujian spesifikasi, tetapi kesalahan ekstraksi mengubah 15,0% dari 786 keputusan klinis, meningkat menjadi 21,4% pada citra yang memuat tabel gizi. Selain itu, 18 dari 25 keputusan DANGER berdasarkan data aktual berubah menjadi SAFE akibat kolom yang tidak terbaca dianggap bernilai nol, sehingga verifikasi pengguna tetap diperlukan. Penelitian ini menghasilkan integrasi sistem dan mekanisme pengambilan keputusan yang dapat direproduksi, didukung evaluasi lintas sumber dan protokol pengujian terbuka.

Kata kunci: Segmentasi Instan; OCR; YOLO11s-seg; PaddleOCR; Literasi Gizi

INTRODUCTION

Non-communicable diseases (NCDs) are strongly associated with dietary patterns, and diabetes and hypertension remain among the largest contributors to the disease burden in Indonesia. The 2023 Indonesian Health Survey (SKI) reported a diabetes prevalence of 11.7%, an

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

increase from 10.9% in the 2018 Basic Health Research (Riskesdas), while hypertension prevalence decreased from 34.1% to 30.8% over the same period (Kementerian Kesehatan RI, 2023) Analyses of the 2023 SKI further link the increase in diabetes cases to modifiable lifestyle and metabolic factors such as central obesity and infrequent blood-pressure screening(S.K.M., M. Epid. & Sevtiyani, 2025). Because a substantial share of diagnosed individuals do not maintain regular treatment or follow-up (Kementerian Kesehatan RI, 2023), practical tools that support everyday self-management are needed alongside clinical care.

At the individual level, the ability to interpret nutrition information is limited. A systematic review of traffic-light nutrition labelling in Indonesia reported that consumer understanding of nutrition labels remains low, even though attitudes toward simplified label formats are generally positive (Shanti, 2025). New front-of-pack initiatives such as the Nutri-Level label have been introduced with the same objective of helping consumers choose healthier products (Kahendra et al., n.d.). In practice, however, conventional nutrition tables are difficult to process automatically because the text can be small, layouts vary across products, and reflections or curved packaging degrade image quality (Nagayi et al., 2025). For meals served on a plate there is no printed table at all, so dietary assessment must begin from visual food recognition.

Image-based dietary assessment is usually described as a pipeline of food image analysis, volume or portion estimation, and nutrient derivation(Wang et al., n.d.). Reviews of the field note that segmentation, recognition, and nutrient estimation are commonly developed as separate tasks, and that dataset diversity, portion estimation, and real-world image conditions remain persistent challenges (Rouhafzay et al., 2025; Sultana et al., 2023) Indonesian contributions remain scarce: most existing work focuses on classification of local dishes rather than instance segmentation of plated meals (Fauzan & Pamungkas, 2025; Rasyidi et al., 2024). For example, Rasyidi et al introduced a classification dataset of 24,427 images covering 160 Indonesian food types, but to the best of our knowledge there is still no public instance-segmentation benchmark for Indonesian plated meals. Generalisation across sources is a known concern as well: recent food-video segmentation benchmarks show that current methods suffer from limited data diversity and degrade under new viewpoints and domains (AlMughrabi et al., n.d.). Domain shift must therefore be measured explicitly rather than assumed away.

The label-reading problem presents a parallel challenge. A benchmark on real-world food packaging compared Tesseract, EasyOCR, PaddleOCR, and TrOCR, and found that dense tabular structures such as nutrition-facts panels are a major source of recognition errors (Nagayi et al., 2025). For this reason, preprocessing, layout reconstruction, and domain-specific correction are commonly required in Indonesian settings (Homepage et al., 2026; Saputra et al., 2024). PaddleOCR provides an Apache-licensed OCR toolkit that includes the PP-OCR model family and is practical for local deployment (Cui et al., 2025).

Automated recognition still makes mistakes, so human verification matters. In a longitudinal food-logging study, image recognition still required a 20.7% log edit rate, showing that users remain an essential part of the loop (Bakar et al., n.d.). This motivates the Human-in-the-Loop (HITL) design adopted in Healthier, where users can correct segmentation and extracted values before they are used by the decision layer. Local inference is also attractive for health applications because on-device processing can reduce privacy exposure and dependence on continuous cloud connectivity (Heydari & Mahmoud, 2025).

The literature provides strong evidence for individual components, but relatively few studies integrate plate-based food analysis, packaged-label extraction, and deterministic guideline-based decision-making into a single offline workflow with explicit human verification. Healthier addresses this integration gap through three research questions: (RQ1) how well a segmentation model trained on non-Indonesian food datasets generalises to new and unseen sources, and which classes fail (RQ2) to what extent a layout-aware OCR pipeline improves nutrition-table extraction over a baseline parser, and what error modes remain; and (RQ3) how segmentation and label outputs can be translated into transparent, guideline-aligned feedback with explicit user verification.

This study makes three explicit contributions, each stated together with how it is demonstrated experimentally, so that the novelty rests on a verifiable integration and decision mechanism rather than on merely combining existing technologies. First, a leakage-controlled evaluation protocol for cross-source food segmentation: we define a data protocol with perceptual-hash deduplication, a fixed validation split, and three independent external test sets that are never used for training or model selection, and we report per-class generalisation across sources (evidenced by the internal and external results and

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

the per-class error analysis in the Results and Discussion).

Second, a reproducible, layout-aware OCR pipeline for nutrition tables we combine four preprocessing stages with a grid-based table parser and fuzzy field mapping, and evaluate it at the field level with image-level cluster-bootstrap confidence intervals, together with a per-stage ablation design. Third, a deterministic, guideline-grounded decision layer with HITL verification we translate verified segmentation and label values into SAFE/WARNING/DANGER feedback using explicit thresholds derived from PERKENI 2024 and hypertension guidance, while retaining an explicit user-correction step. The novelty is therefore in the integration mechanism, its validation protocol, and its transparent decision strategy, all of which are specified and testable, rather than in a claim that any single model is state of the art.

RESEARCH METHODS

This research adopts an iterative experimental methodology informed by CRISP-ML(Q), with separate evaluation of the segmentation, OCR, clinical-rule, and usability components. The design emphasises repeated validation and strict separation of data because variability in food images, datasets, ground truth, and reporting practices is a major barrier to reliable comparison between dietary-assessment systems (Sultana et al., 2023; Wang et al., n.d.).

Types of Research

This is applied, quantitative, and experimental research. It develops an engineering artefact (the Healthier pipeline) and evaluates it with objective metrics rather than subjective judgement. The study is descriptive for the system design and comparative for the component evaluation the design compares each component against a baseline or an alternative configuration. The experimental design is a controlled component-wise evaluation: the data split is fixed once and reused by every experiment, and each module (segmentation, OCR, clinical rules) is evaluated independently before being integrated.

Time and Place of Research

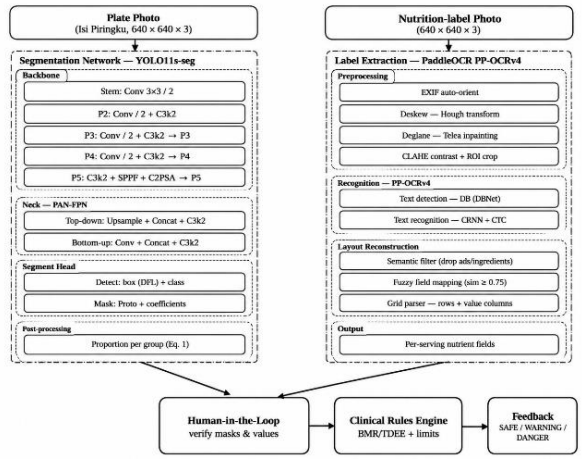
The research was carried out between June 2026 and November 2026 at the Informatics Engineering Study Program, Universitas Krisnadwipayana, Jakarta, Indonesia. Model training used a cloud Kaggle Tesla T4 GPU environment, while OCR experiments and interface checks ran locally on a laptop with an NVIDIA GeForce RTX 3050 Laptop GPU (6 GB) and a CPU PyTorch build, matching the intended offline deployment condition.

Research Target / Subject

This study covers plated-food images, Indonesian packaged-food nutrition labels, and users with diet-related NCD risk profiles. Segmentation was evaluated on three public datasets: UECFoodPIXCOMPLETE, FoodSeg103, and FruitSeg30. OCR evaluation covered five product categories. From 453 collected packaging photographs, 15 duplicates were removed, leaving 438 images, deterministically split into 131 test and 307 development images (seed 42, SHA-256 manifest lock). All test images were manually verified without viewing pipeline predictions. Nutrient values were transcribed from the Per Sajian column, excluding %AKG. Absent and unreadable fields were excluded from accuracy calculations, including 38 unreadable fields across 19 images. Verification corrected 177 cells across 36 images, while 95 remained unchanged. Evaluation was stratified by product category because brand identity was not separately annotated.

Procedure

The end-to-end procedure has two input paths that share one verification layer and one decision layer. A plate photograph is segmented into food instances and grouped into five Isi Piringku groups (Staple Food, Side Dishes, Vegetables, Fruits, Others); a packaging photograph is preprocessed, recognised, and parsed into nutrition fields. Both outputs pass through a HITL verification layer and are then converted by the deterministic Clinical Rules Engine into SAFE/WARNING/DANGER feedback.



The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

Figure 1 summarises the workflow

Figure 1 summarises the workflow of Healthier. A plate photograph is processed by a YOLO11s-seg network, which predicts food instances and their masks; the instances are grouped into the five *Isi Piringku* categories, and the proportion of each group is computed from polygon areas (Equation 1). A nutrition-label photograph is first preprocessed (EXIF orientation, deskewing, deglare, CLAHE, and ROI cropping) and then passed to PaddleOCR PP-OCRv4 for text detection and recognition. A layout-reconstruction stage, consisting of a semantic filter, fuzzy field mapping (similarity ≥ 0.75), and a grid parser, converts the recognised text into per-serving nutrient fields. Both outputs are verified by the user in the Human-in-the-Loop step and then evaluated by the Clinical Rules Engine, which uses BMR/TDEE-based limits to return SAFE, WARNING, or DANGER feedback.

Dataset Construction, Annotation, and Splits

Three public datasets were used: UECFoodPIXCOMPLETE (Okamoto et al., 2021), FoodSeg103 (Wu et al., 2021), and FruitSeg30 (Shamrat et al., 2024). FruitSeg30 contains 1,969 images in 30 fruit classes collected in natural and uncontrolled environments across Malaysia, Bangladesh, and Australia (Shamrat et al., 2024). The candidate corpus contained 19,087 images. Duplicate images were detected with a difference hash (dHash) computed on an 8×8 grayscale reduction of each image, and 373 duplicates were removed, leaving 18,714 unique images. Class labels from all sources were remapped into the five Isi Piringku groups using a fixed mapping table.

Public dataset annotations were retained without re-annotation, applying only deterministic label remapping through a frozen lookup table. A total of 104 FoodSeg103 classes and 102 UECFoodPIXCOMPLETE categories were mapped into five Isi Piringku groups. Notably, 25 FoodSeg103 classes (24.0%), including sweets and sauces, were assigned to Others, creating semantic heterogeneity that may explain its lower segmentation precision. Since remapping was performed by one author, inter-annotator agreement (Cohen's κ) was unavailable and is acknowledged as a limitation.

Official UEC and FoodSeg103 training portions were split 85:15 into 11,684 training and 2,017 validation images (seed 42, class-stratified, without overlap). External testing used three untouched sets: UEC test (948), FoodSeg103 validation (2,096), and FruitSeg30 (1,969), totalling 5,013 images. FruitSeg30 specifically evaluates domain shift because Fruits training samples originate exclusively from FoodSeg103. Class imbalance was assessed across all splits, with training instance and image-presence ratios of 3.04× and 4.12×, respectively. No oversampling or class-weighted loss was applied; imbalance was instead analysed alongside per-class AP. Tables 1 and 2 present split composition and per-class instance distribution.

Table 1. Dataset and split composition (images).

Split	Source	Images
Train	UECFoodPIXCOMPLETE	7,478
Train	FoodSeg103	4,206
Validation	UECFoodPIXCOMPLETE	1,285
Validation	FoodSeg103	732
External test	FoodSeg103 validation	2,096
External test	UECFoodPIXCOMPLETE test	948
External test	FruitSeg30	1,969
Total		18,714

Table 2. Per-class instance distribution per split.

Split	Staple	Side	Vegetables		Fruits	Others
Train (UECFoodPIXCOMPLETE)	4,188	3,614	1,466	0	4,996	
Train (FoodSeg103)	3,369	3,685	9,223	3,513	4,103	
Validation (UECFoodPIXCOMPLETE)	749	608	257	0	927	
Validation (FoodSeg103)	585	626	1,619	618	638	
External: FoodSeg103 validation	1,629	1,751	4,718	1,649	2,004	
External: UEC test	543	463	224	0	754	
External: FruitSeg30	0	0	0	1,993	0	

Segmentation Model and Training

The model is YOLO11s-seg (Khanam & Hussain, 2024). It was trained with Ultralytics 8.4.199 for at most 100 epochs with early-stopping patience 20; training stopped at epoch 67 and the best checkpoint was epoch 48. Optimisation used AdamW with an initial learning rate of 0.001, a final learning-rate factor of 0.01, weight decay 0.0005, and momentum 0.9, with batch size 16, image size 640, and seed 42 (deterministic). Spatial and photometric augmentation was applied (HSV jitter, translation, scaling, horizontal flipping, mosaic, and

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

random erasing). Training took 4.01 h. Table 3 lists the full configuration.

Table 3. Segmentation training configuration.

Parameter	Value
Model	YOLO11s-seg
Framework	Ultralytics 8.4.199
Torch / Python	2.10.0+cu128 / 3.12.13
Max epochs / patience	100 / 20
Stopped epoch / best epoch	67 / 48
Optimiser	AdamW
Initial learning rate	0.001
Final LR factor	0.01
Weight decay	0.0005
Momentum	0.9
Batch size	16
Image size	640
Seed	42
Augmentation	HSV, translate, scale, flip, mosaic, erasing
Training time	4.01 h (Tesla T4)

Plate Proportion Estimation

For each plate image, the model outputs one segmentation polygon per detected food instance. Instances are assigned to their Isi Piringku group, and the 2D visual area proportion of a group is computed as the ratio of the sum of its food-polygon areas to the total analysed food area on the plate (Equation 1).

$$\text{Proportion_g} = A_g / \Sigma_g A_g \quad (1)$$

Here, A_g represents the total polygon area of group g. Non-food objects, such as cutlery, plates, and hands, are excluded before calculating proportions. Equation 1 measures 2D visual area for qualitative Isi Piringku feedback, not food mass or caloric content. Camera perspective, occlusion, and food pile height introduce uncertainty when estimating 3D quantities from 2D images. Therefore, polygon areas are never converted into grams or kilocalories, and Clinical Rules Engine thresholds rely exclusively on verified nutrient values.

Nutrition-Label Extraction Pipeline

Packaging images undergo four preprocessing stages before recognition, including EXIF auto-orientation, deskewing using the Hough Transform (Heydari & Mahmoud, 2025), glare removal through Telea inpainting(Zhou et al., 2023) , CLAHE-based contrast normalisation, and region-of-interest (ROI) cropping around the nutrition table. The processed region is recognised with PaddleOCR (Cui et al., 2025) configured with the PP-OCRv4 recognition model (PP-OCRv4_mobile_det and en_PP-OCRv4_mobile_rec) and the SLANet_plus table-structure recognition model. An earlier draft of this manuscript referred to PP-OCRv5; that was a documentation error. The implementation pins the model version explicitly (NUTRILABEL_OCR_VERSION, default PP-OCRv4), and every result reported in this study was produced with PP-OCRv4.

Layout reconstruction then (i) groups text boxes into logical rows using Y-band grouping; (ii) detects the value columns (Per Sajian / Serving and %AKG / %DV) relative to the row-label column; and (iii) splits tall numeric boxes that PP-OCR merges across adjacent rows (for example, 1725 is split into 1, 7, 2, 5), thereby separating the per-serving and %AKG columns. A semantic filter removes non-nutritional text (promotional slogans, BPOM registration numbers, ingredient lists, and cooking instructions). Finally, fuzzy matching based on edit distance maps OCR tokens to standardised nutrition terms; the Levenshtein distance and similarity are given in Equations 2 to 4.

$$\text{lev}(a, b) = \max(|a|, |b|), \text{ if } \min(|a|, |b|) = 0 \quad (2)$$

$$\text{lev}(a, b) = \min(\text{lev}(a-1, b) + 1, \text{lev}(a, b-1) + 1, \text{lev}(a-1, b-1) + [a_i \neq b_j]) \quad (3)$$

$$Similarity(a, b) = \frac{1 - lev(a, b)}{max(|a|, |b|)} \quad (4)$$

A token is accepted as a match when Similarity ≥ 0.75

Clinical Rules Engine

Nutrition values that have passed HITL verification are evaluated by a deterministic rules engine that runs locally in the browser (no cloud inference). The engine computes Basal Metabolic Rate (BMR) with the Mifflin–St Jeor equations and Total Daily Energy Expenditure (TDEE) as BMR multiplied by an activity factor (Equations 5 to 7).

$$BMR_{male} = 10W + 6.25H - 5A + 5 \quad (5)$$

$$BMR_{female} = 10W + 6.25H - 5A - 161 \quad (6)$$

$$\text{TDEE} = \text{BMR} \times \text{AF} \quad (7)$$

where W is body weight (kg), H is height (cm), A is age (years), and AF is the activity factor (1.2, the sedentary value used by the application). Personal limits are then derived per profile (Table

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

4). For type-2 diabetes, the daily sugar limit follows PERKENI 2024 and is set dynamically at 5% of TDEE (0.05 × TDEE / 4 g); for non-diabetic profiles the default is 10% of TDEE (0.10 × TDEE / 4 g). Sodium limits are absolute values: 1,500 mg/day for hypertension and 2,300 mg/day otherwise. Saturated-fat limits are 10% of TDEE (0.10 × TDEE / 9 g), reduced to 7% (0.07 × TDEE / 9 g) for diabetes with dyslipidaemia; cholesterol is limited to 200 mg for diabetes with dyslipidaemia; and the protein limit for diabetic nephropathy is 0.8 g × body weight (Perkumpulan Endokrinologi Indonesia, n.d.) . All limits are implemented exactly as listed in Table 4.

Table 4. Daily nutrient limits implemented in the Clinical Rules Engine.

Limit	Value	Applies to
Snack calorie budget	0.20 × TDEE	All users
Sugar	0.10 × TDEE / 4 g	Default
Sugar	0.05 × TDEE / 4 g	Type-2 diabetes
Sodium	2,300 mg/day	Default
Sodium	1,500 mg/day	Hypertension
Sodium-density warning	800 mg/100 g	Default
Sodium-density warning	600 mg/100 g	Hypertension
Saturated fat	0.10 × TDEE / 9 g	Default
Saturated fat	0.07 × TDEE / 9 g	Diabetes with dyslipidaemia
Cholesterol	200 mg/day	Diabetes with dyslipidaemia
Protein	0.8 g × body weight	Diabetic nephropathy
Fibre target	25–38 g/day	Default
Fibre target	20–35 g/day	Type-2 diabetes
Carbohydrate target	45–65% of energy	Type-2 diabetes

The engine classifies accumulated intake into SAFE, WARNING, and DANGER (Table 5). A WARNING is raised when sugar or saturated fat exceeds 50% of the daily limit, when actual sodium exceeds 40% of the sodium limit, when protein exceeds 30% of the nephropathy quota, or when a product's sodium density exceeds 600 mg/100 g for hypertension (800 mg/100 g otherwise). DANGER is raised only when sugar exceeds 100% of the daily limit; sodium and saturated fat saturate at WARNING. These bands are the exact thresholds implemented in the browser reference implementation (frontend/js/clinical-rules.js) and verified against an independent specification harness (experiments/scripts/verify_clinical_rules.js); they are reported as functional rule verification, not clinical validation, because no independent registered-dietitian review was performed.

Table 5. Risk classification thresholds.

Status	Condition
SAFE	All nutrients below the warning thresholds
WARNING	Sugar or saturated fat > 50% of limit; sodium > 40% of limit; protein > 30% of nephropathy quota; sodium density > 600 mg/100 g (hypertension) or > 800 mg/100 g (other)
DANGER	Sugar > 100% of the daily limit

RESULTS AND DISCUSSION

This section reports the quantitative evaluation of the two recognition modules and their behaviour under domain shift, together with the component ablation, baseline, and integration analyses where results are available; the analyses that remain to be completed are marked explicitly.

Instance Segmentation Evaluation

On the 2,017-image internal validation split (best checkpoint, epoch 48), the YOLO11s-seg model reached a mask mAP50 of 71.66% and a mask mAP50-95 of 56.91%, with mask precision 72.95% and recall 65.50%. Box metrics were slightly higher: box mAP50 72.37%, box mAP50-95 60.11%, box precision 72.87%, and box recall 66.04% (Table 6). The gap between precision and recall is expected in instance segmentation of plated food, where overlapping and heterogeneous regions make some instances easier to localise than to recall fully; this recall gap is one reason the system retains an interactive correction step. Figures 2 - 4 show the training curves, the normalised confusion matrix, and qualitative ground truth versus prediction.

Table 6. Internal validation metrics at the best checkpoint.

Metric	Box	Mask
Precision	72.87	72.95
Recall	66.04	65.50
mAP50	72.37	71.66
mAP50-95	60.11	56.91

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i4.653

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

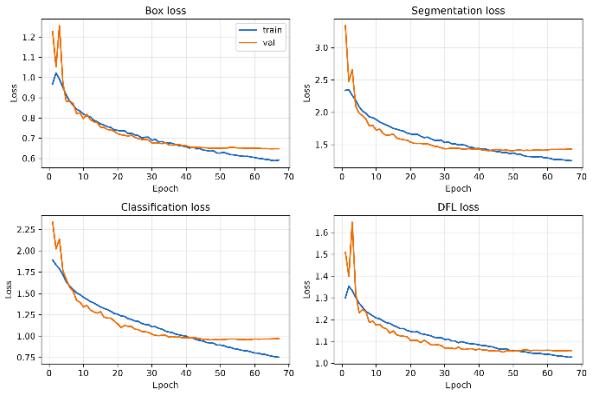


Figure 2. Training and validation loss curves for the YOLO11s-seg model.

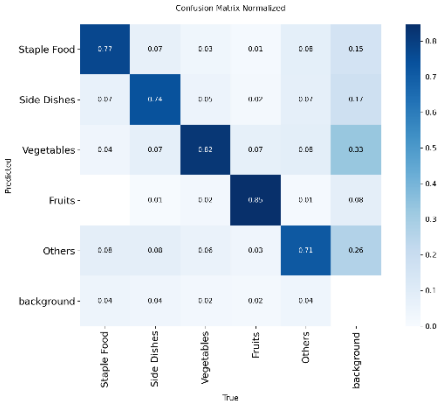


Figure 3. Normalised confusion matrix on the internal validation split.

Figures 2 and 3 show the evaluation results of the trained model: the loss curves in Figure 2 show a steady decrease in box, segmentation, classification, and DFL losses during training and validation, while the normalised confusion matrix in Figure 3 displays per-class prediction accuracy along its main diagonal. Figure 4 illustrates qualitative ground-truth masks versus model predictions on the internal validation split.

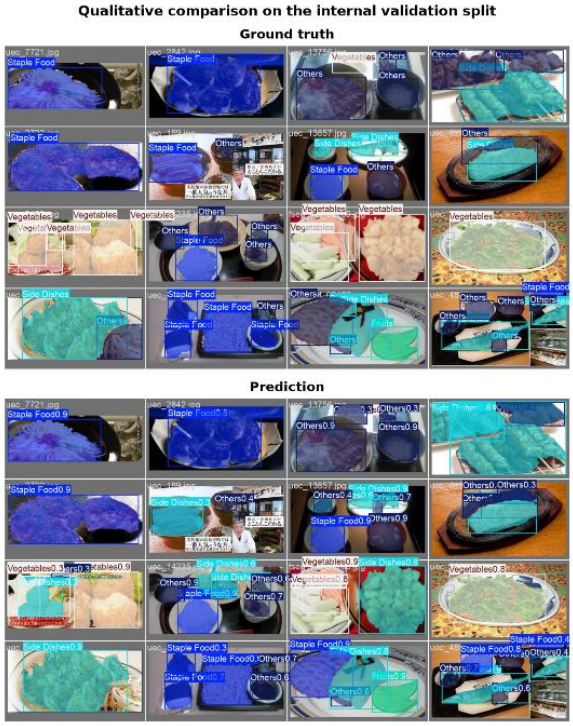


Figure 4. Qualitative comparison of ground truth

Generalisation to External Sources and Per-Class Error Analysis

Because the training data are not Indonesian-specific, external evaluation is the central generalisation test. The model was applied unchanged to the three held-out sources. Mask mAP50 was 69.15% on UECFoodPIXCOMPLETE test and 63.66% on the FoodSeg103 validation set a drop of only a few points relative to internal validation but fell to 17.06% on FruitSeg30 (Tables 7 and 8). Because the training and evaluation pipeline, classes, and metrics were identical across sources, the drop isolates a domain shift rather than a training error. Figures 5 to 7 show the external confusion matrices and Figure 8 shows the mask precision–recall and F1 curves.

Table 7. External-test instance-level metrics Box

Dataset	Box P	Box R	Box mAP 50	Box mAP 50-95
FoodSeg103 validation	66.97	59.36	64.57	52.08
UECFoodPIXCOMPLETE test	67.82	68.08	70.30	57.86
FruitSeg30	17.28	44.96	17.36	12.85

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i4.653

JURNAL RISET INFORMATIKA
Vol. 8, No. 4. September 2026

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Table 8 External-test instance-level metrics Mask

Dataset	Mask P	Mask R	Mask mAP 50	Mask mAP 50-95
FoodSeg103 validation	67.47	58.37	63.66	48.16
UECFoodPIXCOMPLETE test	67.42	67.44	69.15	54.30
FruitSeg30	17.01	44.25	17.06	13.55

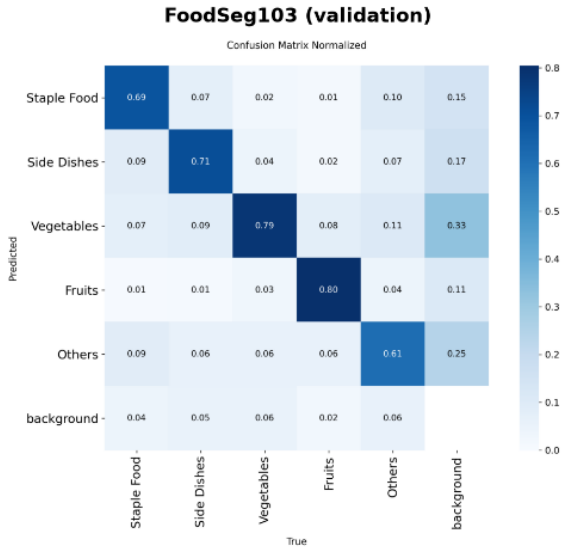


Figure 5. Normalised confusion matrices on the FoodSeg103

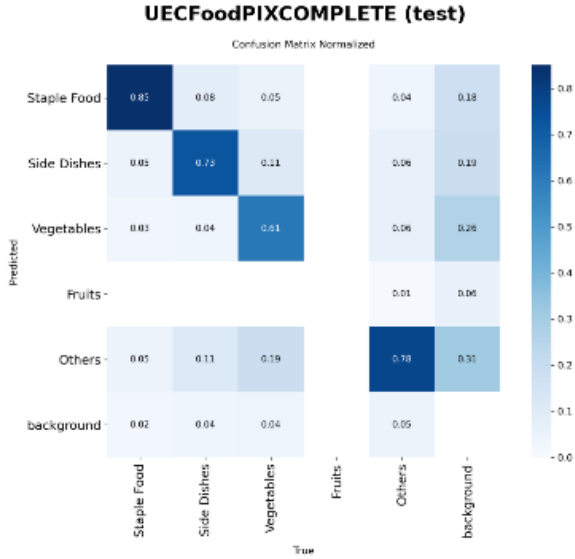


Figure 6. Normalised confusion matrices on the UECFoodPIXCOMPLETE.

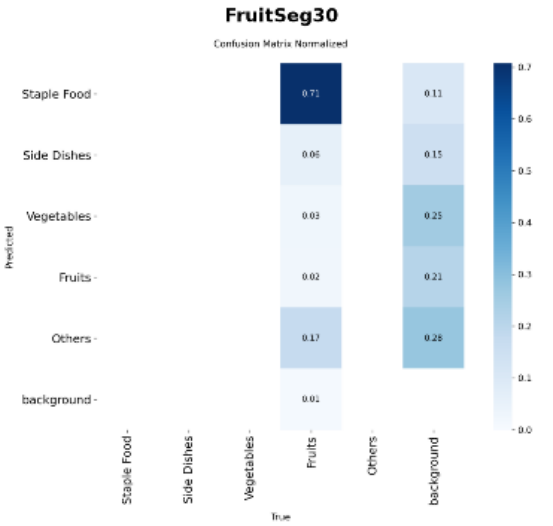
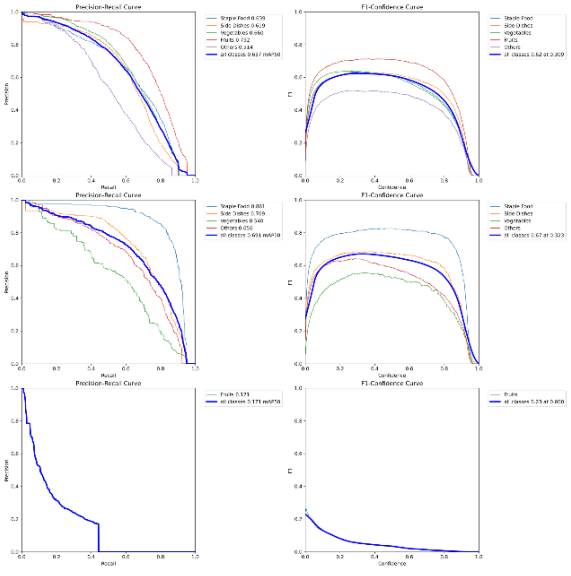


Figure 7. Normalised confusion matrices on the FruitSeg30

Figures 5–7 show the normalised confusion matrices on the three external sources. On FoodSeg103, the diagonal values were highest for Fruits (0.80) and Vegetables (0.79) and lowest for Others (0.61), and most errors were false positives from Vegetables and Others. On UECFoodPIXCOMPLETE, Staple Food was recognised best (0.83), Vegetables was weakest (0.61), and no Fruits class was present. On FruitSeg30, the model almost failed: only 0.02 of the fruit instances were labelled Fruits, while 0.71 were assigned to Staple Food and 0.17 to Others. This confirms the domain shift reported in Table 7.



The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i4.653

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Figure 8. Mask precision–recall and F1 curves for each external source.

Per-class average precision shows where the model succeeds and fails (Table 9). On UECFoodPIXCOMPLETE the strongest class was Staple Food (mask AP50 85.70%) and the weakest was Vegetables (55.42%); on FoodSeg103 the strongest was Fruits (74.96%) and the weakest was the heterogeneous Others class (51.24%). On FruitSeg30 the only class present is Fruits, which scored 17.29%. The pixel-level view sharpens the failure: on FruitSeg30 the Fruits class had pixel precision 97.17% but recall only 2.14% (F1 4.19%), meaning the few pixels it labelled as fruit were mostly correct, but it recovered almost none of the fruit and instead assigned fruit regions to other groups (Table 10, Figure 9).

Table 9 Segmentation performance

Dataset / Class	GT	M50	M95	B50	B95
UEC					
Staple	543	85.7	73.3	87.5	77.3
Side	463	70.7	53.5	71.5	58.8
Veg.	224	55.4	40.6	55.5	44.1
Others	754	65.2	50.6	65.9	52.2
FoodSeg103					
Staple	1629	63.9	49.9	65.6	55.0
Side	1751	61.6	47.2	64.4	52.2
Veg.	4718	65.8	48.3	66.9	52.7
Fruits	1649	75.0	56.6	72.6	58.3
Others	2004	51.2	39.2	53.0	43.2
FruitSeg30					
Fruits	1993	17.3	13.9	17.7	13.3

M50/M95 = Mask AP50/AP50–95; B50/B95 = Box AP50/AP50–95.

Table 10 Pixel-level performance

Dataset / Class	Acc.	Prec.	Rec.	F1
FoodSeg103				
Staple	92.9	72.7	67.0	69.7
Side	93.5	75.1	77.9	76.5
Veg.	93.6	75.4	74.5	74.9
Fruits	98.0	70.9	66.7	68.8
Others	94.6	55.9	55.6	55.8
UEC				
Staple	93.3	81.6	82.0	81.8
Side	94.3	72.7	73.4	73.1
Veg.	96.0	64.3	61.5	62.9
Others	94.4	61.7	61.4	61.5
FruitSeg30				
Fruits	61.3	97.2	2.1	4.2

FoodSeg103 demonstrates balanced performance with >92% accuracy (best in Side Dishes and Vegetables). UEC remains stable above 93% (best in Staple). Conversely, FruitSeg30 experiences an anomaly with very high precision (97.2%) but very low recall and F1-Score (<5%), indicating that the model struggles significantly in recognizing that class comprehensively.

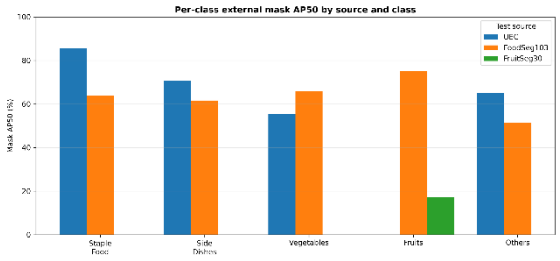


Figure 9. Per-class external mask AP50 by source and class.

The pattern is interpretable. UEC and FoodSeg103 contain plated meals and kitchen food photographs that share low-level statistics with the training pool, whereas FruitSeg30 contains isolated fruit in natural, uncontrolled environments (Shamrat et al., 2024) a different visual domain that the model was never trained on, especially because the Fruits class in training comes only from FoodSeg103. We therefore treat the FruitSeg30 result as an explicit limitation and as evidence for the generalisation and failure-analysis contribution, not as a headline strength.

Baseline Comparison

A full-budget comparison against an alternative segmentation architecture could not be completed: the main model was trained on a cloud T4 GPU, whereas the local environment available for this study had a CPU-only PyTorch build, so retraining an alternative architecture with the same 100-epoch budget and split was not feasible within the study window. A short five-epoch screening run with three alternative architectures is available from an earlier notebook: mask mAP50 was 54.67% for YOLOv8m-seg, 52.93% for YOLO11m-seg, and 53.45% for YOLO26m-seg (box mAP50 56.27%, 54.58%, and 54.89%, respectively). Because that run used a reduced training subset and only five epochs, these numbers are reported only as a screening pilot; they are not comparable to the 100-epoch YOLO11s-seg model (internal mask mAP50 71.66%) and must not be read as evidence that any architecture is superior. A matched full-budget baseline remains future work.

Nutrition-Label Extraction and Preprocessing Ablation

The OCR module was evaluated on the locked test set of 131 Indonesian packaged-food

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

label photographs spanning five product categories (Frozen Food 20, Makanan Kaleng 16, Makanan Ringan 53, Minuman Sachet 18, and Saos dan Perbumbuan 24) with human-verified ground truth. A field is counted as correct when the extracted numeric value is exact or within 5% of the verified per-serving ground-truth value, with 95% confidence intervals computed via a 2,000-iteration cluster bootstrap over images (seed 42) and paired comparisons evaluated using paired bootstrap intervals and McNemar's test. Of the 582 printed field annotations on the labels, 544 carry a legible numeric value and form the denominator for every configuration; 38 printed values were unreadable in the photographs (primarily trans fat, fibre, and cholesterol) and were excluded from all denominators rather than imputed. Sixty of the 131 photographs contain no nutrition table (composition-only or single-ingredient labels) and are retained to test whether a pipeline suppresses rather than hallucinates values.

Table 11 Pipeline Performance Comparison

Pipeline	Acc % (95% CI)	Ex	Wr	Ms
A. Tesseract	14.9 (10.3–20.2)	81	154	309
B. EasyOCR	19.7 (15.0–24.5)	107	210	226
C. Paddle + Line	35.1 (29.3–40.9)	191	159	193
D. Preproc + Layout	23.5 (17.6–29.9)	128	53	363
E. Full (Grid)	42.7 (36.2–49.0)	232	98	213

Ex. = Exact, Wr. = Wrong, Ms. = Missing

The full pipeline (E) reached 42.65% field-level accuracy (Table 11), 7.54 percentage points above raw PaddleOCR (C; paired bootstrap 95% CI [+2.54%, +12.80%], McNemar p = 0.0021) and 27.76 points above Tesseract (A; p < 0.001); EasyOCR (B) reached 19.67%. The bare layout parser without grid reconstruction (D) scored 23.53%, which is below raw PaddleOCR because without column-grid reconstruction it converts many candidate values into missing fields (363 missing). The parser only becomes beneficial once preprocessing and grid reconstruction are active, which is why the contribution is reported for the complete pipeline rather than for the parser in isolation. Figures 10 and 11 show the extraction pipeline and a qualitative extraction example.

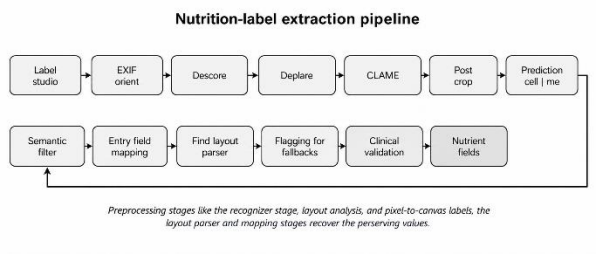


Figure 10. Nutrition-label extraction pipeline: preprocessing, recognition, and layout reconstruction.

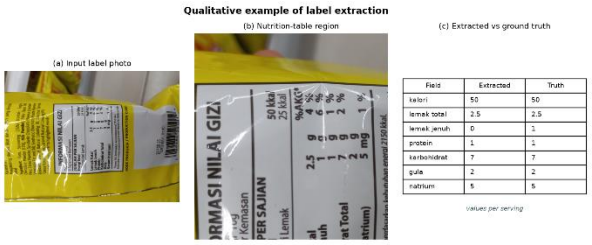


Figure 11. Qualitative example of label extraction.

Table 12 Cumulative preprocessing and post-processing ablation (n = 131 images, 544 fields).

Config	Acc % (95% CI)	Δ	Ex / Wr / Ms
Raw OCR	30.0% [23.7–36.6%]	-	163 / 52 / 328
+ deskew	28.7% [23.0–34.7%]	-1.29	156 / 58 / 330
+ deglare	28.7% [23.0–34.7%]	0.00	156 / 58 / 330
+ CLAHE	23.4% [17.9–29.1%]	-5.33	127 / 52 / 365
+ ROI	23.5% [17.6–29.9%]	+0.18	128 / 53 / 363
+ semantic	22.2% [16.8–27.9%]	-1.29	121 / 60 / 363
+ fuzzy match	27.2% [21.4–33.3%]	+4.96	148 / 78 / 318
+ fallback	30.3% [24.3–36.6%]	+3.12	165 / 92 / 287
+ validation	29.6% [23.7–35.7%]	-0.74	161 / 82 / 301

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

Full pipeline	42.7% [36.2–49.0%]	+13.05	232 / 98 / 213

The cumulative ablation results (Table 12) show that preprocessing stages do not consistently improve accuracy. Deskewing and deglare had negligible effects, while CLAHE and semantic filtering reduced performance. Fuzzy matching (+4.96 points) and energy/fat fallback (+3.12 points) provided the largest improvements before grid reconstruction. Grid reconstruction had the greatest impact, increasing accuracy from 29.60% to 42.65% (+13.05 points, 95% CI [+8.73%, +17.36%], McNemar p = 1.43e-13). LOO analysis confirmed fuzzy matching as beneficial (-10.29 points when removed), whereas removing semantic filtering improved accuracy to 46.14% (+3.49 points, p = 8.8e-4). Removing energy/fat fallback or clinical validation had no significant effect. Thus, semantic filtering was retained as a negative finding rather than tuned on the test set.

HITL evaluation showed that 232 of 544 fields (42.6%) required no correction, while 423 edits were needed across 655 actionable cells, including 98 replacements, 213 missing-value entries, and 111 spurious-value removals. Overall, 100 of 131 images (76.3%) required corrections, with a median of three edits per image.

Clinical Rules Engine Verification and Error Propagation

Clinical Rules Engine verification achieved complete agreement across 36 daily-limit checks and 30 status decisions. Across 786 clinical decisions, the full pipeline preserved ground-truth status in 668 cases (85.0%), compared with 76.3% for raw PaddleOCR and 81.9% for the parser without grid. For images containing nutrition tables, full-pipeline agreement was 78.6%. However, 18 of 25 ground-truth DANGER decisions (72.0%) became false SAFE outcomes, while 55 of 117 WARNING decisions became SAFE due to missing fields defaulting to zero. These findings demonstrate that HITL verification remains essential to prevent clinically significant under-warning.

Table 13. End-to-end clinical decision propagation on the verified 131-image test set (786 total decisions; 426 on the 71 table-bearing images).

Pipeline	Unch. All	Unch. Table	DNG to SF	WRN to SF
Paddle OCR	600/786 (76.3%)	284/426 (66.7%)	9/25	47/117
No Grid	644/786 (81.9%)	308/426 (72.3%)	17/25	71/117
Full Pipeline	668/786 (85.0%)	335/426 (78.6%)	18/25	55/117

Unch = Unchanged Status, DNG = Danger, WRN = Warning

Usability Assessment

System usability was assessed with Nielsen's heuristic evaluation using four evaluators in defined roles (Table 14): two independent machine-learning engineers/data scientists (E1 and E2) who performed the severity-weighted heuristic review, one lead data scientist (E3) who conducted an asynchronous architecture review, and one self-evaluator (E4) who checked documentation–implementation consistency and was excluded from severity weighting. The evaluation found no severity-4 (catastrophic) issues, with all remaining findings rated severity 2–3 (minor to major). The two HITL controls the segmentation editor and the nutrient value verification form enabled users to recover from the recognition failures quantified above. Because quantitative severity weighting rests on the two independent evaluators (E1–E2), this evaluation is reported as an exploratory heuristic assessment rather than an end-user study.

Table 14 Usability evaluator Assessment

Eval	Role	Contribution
E1	Machine-learning engineer / data scientist	Independent heuristic review (severity-weighted)
E2	Machine-learning engineer / data scientist	Independent heuristic review (severity-weighted)
E3	Lead data scientist	Asynchronous architecture review
E4	Self-evaluator	Documentation–implementation consistency (excluded from severity weighting)

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

Edge Deployment Benchmark: Latency, Memory, and Artifact Size

To evaluate local offline deployment feasibility, component inference latency, peak memory footprint (resident set size, RSS), and on-disk model artifact sizes were measured on the local target CPU (Table 15). Because the local PyTorch installation is a CPU-only build (torch 2.15.0+cpu, torch.cuda.is_available() = False), Ultralytics segmentation could not be benchmarked on the local GPU without replacing the environment; therefore, only verified CPU measurements are reported, and no GPU latency claims are made.

Table 15. Local CPU inference latency, peak memory footprint, and on-disk artifact sizes.

Component	Med (s)	P95 (s)	RAM / Disk
YOLO11s-seg	0.13	0.19	770 MB / 21 MB
Full OCR Pipeline	20.58	45.29	3.8 GB / 21 MB
Rules Engine	<0.01	-	<0.1 MB
End-to-End Sum	20.71	-	3.8 GB / 41 MB

On CPU, plate segmentation is lightweight (median 0.1304 s, P95 0.1854 s, peak RSS 770.0 MB) and the Clinical Rules Engine is instantaneous (0.039 ms per call), with compact disk footprints (20.5 MB for YOLO11s-seg .pt weights and 20.7 MB across the three OCR models). However, the full nutrition-label OCR pipeline requires a median of 20.58 s per image (P95 45.29 s) and 3,760.8 MB peak RSS on CPU because it runs multi-stage preprocessing, two-pass detection/recognition, and SLANet_plus table-structure inference on high-resolution photos. Thus, while segmentation and rule evaluation are immediately practical on CPU, unaccelerated multi-stage OCR is the primary latency and memory bottleneck and is not yet real-time on CPU-only edge hardware.

Discussion

The findings demonstrate competitive in-domain segmentation but limited cross-domain generalisation, with mask mAP50 decreasing from 69.15% on UECFoodPIXCOMPLETE and 63.66% on FoodSeg103 to 17.06% on FruitSeg30(AlMughrabi et al., n.d.). For nutrition-label extraction, grid reconstruction was the largest contributor, improving field-level accuracy from 29.60% to 42.65% (+13.05 points), outperforming raw PaddleOCR, EasyOCR, and Tesseract. Conversely, semantic filtering reduced performance, with its removal increasing accuracy to 46.14% (+3.49 points, p = 8.8e-4), consistent with the importance of layout reconstruction in tabular OCR(Nagayi et al., 2025; Saputra et al., 2024) .

Several limitations remain. The OCR evaluation used 131 images across five product categories, excluded 38 unreadable fields, and lacked brand-level stratification. Segmentation was affected by class imbalance (3.04× instance and 4.12× image ratios), heterogeneous Others mapping (24.0% of FoodSeg103 classes), single-author remapping, and limited architecture screening. Plate analysis provides 2D visual proportions rather than quantitative food mass or calories due to perspective, occlusion, and pile-height uncertainty. Clinically, extraction errors altered 15.0% of 786 decisions (21.4% on table-bearing images), including 18 of 25 DANGER cases becoming false SAFE outcomes; 76.3% of images required human correction. On CPU, YOLO11s-seg required 0.130 s per image, whereas OCR required a median of 20.58 s and 3.76 GB RSS. Finally, clinical rules lacked registered-dietitian validation, usability assessment relied on a limited evaluator panel, and no public Indonesian plated-food segmentation benchmark was available.

CONCLUSIONS AND SUGGESTIONS

This chapter presents the primary conclusions drawn from the experimental evaluation and model performance analysis. Additionally, it outlines strategic recommendations and prospective directions for future research to address existing limitations and enhance overall system capability.

Conclusion

This study evaluated Healthier, an offline-oriented nutrition-assistance system integrating YOLO11s-seg, PaddleOCR (PP-OCRv4), a deterministic Clinical Rules Engine, and Human-in-the-Loop (HITL) verification. Segmentation achieved a validation mask mAP50 of 71.66% (mAP50-95: 56.91%), with external scores of 69.15% on UECFoodPIXCOMPLETE, 63.66% on FoodSeg103, and 17.06% on FruitSeg30, indicating domain shift. On 131 verified Indonesian nutrition-label images, extraction achieved 42.65% field-level accuracy, outperforming raw PaddleOCR (35.11%) and Tesseract (14.89%). Grid reconstruction improved accuracy by 13.05 percentage points, while removing the semantic filter increased it to 46.14% (+3.49 points). HITL evaluation showed that 76.3% of images required correction. The rules engine passed all 66 specification checks, but OCR

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

errors altered 15.0% of 786 clinical decisions (21.4% on table-bearing images), including 18 of 25 DANGER decisions becoming false SAFE outcomes due to missing values defaulting to zero. CPU execution required 0.130 seconds for segmentation and 20.58 seconds for OCR. Despite limitations in cross-domain generalisation, OCR accuracy, processing speed, and quantitative food estimation, Healthier demonstrates a transparent, guideline-aligned feedback workflow that combines deterministic rules with human verification to reduce the risk of misleading nutritional assessments.

Suggestion

Four directions are recommended for future work: (1) building a public Indonesian plated-meal segmentation benchmark with multi-annotator reliability (Cohen's κ) and training a matched 100-epoch alternative segmentation architecture on GPU; (2) improving cross-domain robustness and validating portion estimation against weighed food masses (grams) using depth sensing, fiducial markers, or multi-angle imaging; (3) augmenting the 131-image five-category OCR benchmark with verified brand metadata, annotating the 307-image development split for out-of-sample threshold tuning and semantic-filter redesign, and completing image-stage leave-one-out ablations; and (4) applying model quantisation and GPU/NPU acceleration to reduce the 20.58 s CPU OCR latency alongside prospective validation with end users and registered dietitians.

REFERENCES

AlMughrabi, A., Rivo, G., Jiménez-Farfán, C., Ha-roon, U., Al-Areqi, F., Jung, H., Busam, B., Marques, R., & Radeva SegMan M SegMan, P. M. (n.d.). *Graphical Abstract BenchSeg: A Large-Scale Dataset and Benchmark for Multi-View Food Video Segmentation*.

Bakar, L., Englhardt, Z., Srinivas, V., Narayanswamy, G., Nissanka, D., Patel, S., & Iyer, V. (n.d.). *SnappyMeal: Design and Longitudinal Evaluation of a Multimodal AI Food Logging Application* (Vol. 1).

Cui, C., Sun, T., Lin, M., Gao, T., Zhang, Y., Liu, J., Wang, X., Zhang, Z., Zhou, C., Liu, H., Zhang, Y., Lv, W., Huang, K., Zhang, Y., Zhang, J., Zhang, J., Liu, Y., Yu, D., & Ma, Y. (2025). *PaddleOCR 3.0 Technical Report*. http://arxiv.org/abs/2507.05595

Fauzan, A., & Pamungkas, H. (2025). *Regional Food Recognition with Nutritional Analysis: Attention Mechanism Implementation for Indonesian Cuisine* (pp. 186–200). https://doi.org/10.2991/978-94-6463-924-7_16

Heydari, S., & Mahmoud, Q. H. (2025). Tiny Machine Learning and On-Device Inference: A Survey of Applications, Challenges, and Future Directions. In *Sensors* (Vol. 25, Number 10). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/s25103191

Homepage, J., Aria Alireza, R., & Kurniawardhani, A. (2026). *MALCOM: Indonesian Journal of Machine Learning and Computer Science Development of an OCR System for Ingredient List Text Recognition on Food Labels Using CRAFT and CRNN Pengembangan Sistem OCR untuk Pengenalan Teks Daftar Bahan pada Label Makanan Berbasis CRAFT dan CRNN*. *6*, 2111–2121. https://doi.org/10.57152/malcom.v6i4.2707

Kahendra, F., Boni Lestari, P., Daffa Aushof, A., Nabila, A., Gunawan, T., Studi Administrasi Kesehatan, P., Studi Akupuntur dan Pengobatan Herbal, P., & Tinggi Ilmu Kesehatan Husada, S. R. (n.d.). *SerQua : Service Quality (Jurnal Pengabdian Masyarakat) Label Nutri-Level pada Kemasan Pangan sebagai Instrumen Kebijakan Kesehatan dalam Mendorong Perilaku Konsumsi Sehat Masyarakat* (Vol. 3, Number 2).

Khanam, R., & Hussain, M. (2024). YOLOv11: An overview of the key architectural enhancements. https://arxiv.org/abs/2410.17725

Kementerian Kesehatan RI. (2023). *Prevalence, Impact, and Efforts In Controlling Hypertension & Diabetes in Indonesia THE RESULT OF INDONESIAN HEALTH SURVEY 2023 Figure 1. Discrepancy in Treatment-Seeking Behavior for Hypertension Among Productive-Age and Elderly*

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

Individuals Diagnosed with Hypertension.

Nagayi, M., Khan, A., Frank, T., Swart, R., & Nyirenda, C. (2025). *Evaluating OCR performance on food packaging labels in South Africa*. http://arxiv.org/abs/2510.03570

Okamoto, K., Adachi, K., & Yanai, K. (2021). Region-Based Food Calorie Estimation for Multiple-Dish Meals. *CEA 2021 - Proceedings of the 13th International Workshop on Multimedia for Cooking and Eating Activities*, 17–24. https://doi.org/10.1145/3463947.3469236

PERKENI. (2024). Pedoman pengelolaan dan pencegahan diabetes melitus tipe 2 dewasa di Indonesia 2024. PB PERKENI.

Rasyidi, M. A., Mardhiyyah, Y. S., Nasution, Z., & Wijaya, C. H. (2024). Performance comparison of state-of-the-art deep learning model architectures in Indonesian food image classification. *Bulletin of Electrical Engineering and Informatics*, *13*(5), 3355–3368. https://doi.org/10.11591/eei.v13i5.7996

Rouhafzay, A., Rouhafzay, G., & Jbilou, J. (2025). Image-based food monitoring and dietary management for patients living with diabetes: a scoping review of calorie counting applications. In *Frontiers in Nutrition* (Vol. 12). Frontiers Media SA. https://doi.org/10.3389/fnut.2025.1501946

Saputra, T. M. E., Susanto, A., & Carmelita, B. J. (2024). Implementation of Tesseract OCR and Bounding Box for Text Extraction on Food Nutrition Labels. *Building of Informatics, Technology and Science (BITS)*, *6*(3), 1403–1412. https://doi.org/10.47065/bits.v6i3.6107

Shamrat, F. M. J. M., Shakil, R., Idris, M. Y. I., Akter, B., & Zhou, X. (2024). FruitSeg30_Segmentation dataset & mask annotations: A novel dataset for diverse fruit segmentation and classification. *Data in Brief*, *56*. https://doi.org/10.1016/j.dib.2024.110821

Shanti, K. M. (2025). Efektivitas Label Gizi Traffic Light terhadap Pemahaman, Persepsi, dan Sikap Konsumen di Indonesia: Systematic Literature Review. *Jurnal Gizi Kerja Dan Produktivitas*, *6*(1), 31. https://doi.org/10.62870/jgkp.v6i1.31589

S.K.M., M. Epid., M., & Sevtiyani, I. (2025). Factors Contributing to Rising Diabetes Cases in Indonesia: Insights from the 2023 Indonesian Health Survey. *Jurnal Epidemiologi Kesehatan Komunitas*, *10*(3), 57–69. https://doi.org/10.14710/jekk.v10i3.28184

Sultana, J., Ahmed, B. M., Masud, M. M., Huq, A. K. O., Ali, M. E., & Naznin, M. (2023). A Study on Food Value Estimation From Images: Taxonomies, Datasets, and Techniques. In *IEEE Access* (Vol. 11, pp. 45910–45935). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2023.3274475

Wang, W., Min, W., Li, T., Dong, X., Li, H., & Jiang, S. (n.d.). *A review on vision-based analysis for automatic dietary assessment*. Retrieved https://vipl.ict.ac.cn/homepage/minweiqing/Home.html

Wu, X., Fu, X., Liu, Y., Lim, E.-P., Hoi, S. C. H., & Sun, Q. (2021). *A Large-Scale Benchmark for Food Image Segmentation*. http://arxiv.org/abs/2105.05409

Zhou, Y., Zuo, S., Yang, Z., He, J., Shi, J., & Zhang, R. (2023). A Review of Document Image Enhancement Based on Document Degradation Problem. In *Applied Sciences (Switzerland)* (Vol. 13, Number 13). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/app131378555

The work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License