FITMO

Fitness in Motion

Meal Intelligence Food Photo Accuracy White Paper

A transparent benchmark of single-photo calorie, portion, macronutrient, and ingredient estimation

SEPTEMBER 2026 • VERSION 1.0

3,229	85 kcal	2.5 s
meals in the locked primary cohort	mean absolute calorie error	median analysis turnaround

Prepared for product, research, and consumer transparency

EXECUTIVE SUMMARY

What the evidence says

Primary decision

Fitmo ranks first in the valid exact-cohort product comparison. Across 3,229 real cafeteria meals, its calorie MAE was 84.6 kcal versus 149.6 kcal for FatSecret—a 43% reduction in average absolute error. The paired confidence interval supports a clear advantage.

The result is compelling because both systems are evaluated against the same dish IDs and weighed Nutrition5k ground truth. Fitmo also reports protein, carbohydrate, and fat directly; the public FatSecret benchmark does not provide those macronutrients, so that part of Fitmo's evaluation stands on its own rather than as a head-to-head claim.

The strongest public statement is therefore specific: Fitmo materially outperformed FatSecret for calorie and edible-mass estimation on this exact 3,229-meal cohort. Broader statements such as "best in market" are not justified because other commercial and academic systems use different datasets, subsets, inputs, or training conditions.

Decision at a glance

Question	Answer	Confidence
Best exact-cohort product result	Fitmo	High (Grade A)
Calorie advantage vs FatSecret	43% lower MAE	High; paired bootstrap
Portion advantage vs FatSecret	37% lower MAE	High
Macro coverage	Protein, carbohydrate, and fat reported	High for Fitmo; FatSecret NR
Appropriate use	Editable food-diary draft	Not medical measurement

STUDY DESIGN

A locked, reproducible real-meal benchmark

We evaluated Fitmo Meal Intelligence v1 on the exact 3,229 Nutrition5k dish IDs released with the independent Sterling et al. food-recognition benchmark. Each dish contains an overhead RGB image plus weighed ingredient-level ground truth. No meals were manually photographed, relabeled, cherry-picked, or removed after results were known.

Stage	Count	What it means
Published overhead-image pool	3,490	Nutrition5k dishes with an overhead RGB image
Published exclusions	261	Deprecated/plate-only records and other invalid cases in the released cohort construction
Locked evaluation cohort	3,229	Exact author-released dish-ID list
Fitmo completed	3229	100.0% completion
FatSecret completed	3216	99.6% valid calorie outputs

Inputs and outputs

Fitmo received one overhead color image per meal, normalized within the production image envelope. It returned an editable meal draft with food identity, edible mass, calories, protein, carbohydrate, and fat. The evaluation used direct Fitmo output—no database enrichment, manual correction, or post-hoc calibration.

Metrics and uncertainty

We prespecified mean absolute error (MAE), mean and median absolute percentage error (MAPE and MdAPE), root mean squared error (RMSE), concordance correlation coefficient (CCC), signed bias, and the share of meals within 10%, 20%, and 25% of reference. Ingredient agreement uses the same visible-ingredient fuzzy Jaccard method as the independent benchmark. Fitmo confidence intervals use 10,000 deterministic nonparametric bootstrap resamples; exact-cohort product differences are paired by dish ID.

Why MAE and MAPE both matter

MAE expresses average error in familiar units such as calories or grams. MAPE expresses error relative to meal size, but can become unstable for nutrients whose true amount is close to zero. Neither should be interpreted alone.

TIER 1

Exact-cohort product comparison

Figure 1. Same 3,229-meal source cohort. FatSecret valid-output n varies by metric.

Rank	System	Calorie MAE	Calorie MAPE	Within 25%	Mass MAE	Ingredient Jaccard	Grade
1	Fitmo	84.6 kcal	40.2%	43.3%	58.2 g	0.674	A
2	FatSecret	149.6 kcal	87.9%	28.5%	92.3 g	0.583	A

Paired result. Fitmo minus FatSecret calorie MAE was −64.9 kcal (95% CI −70.4 to −59.6; n=3216). A negative value favors Fitmo. For edible mass, the paired MAE difference was −34.3 g (95% CI −38.0 to −30.9; n=3216).

Fitmo accuracy profile

Outcome	MAE	MAPE (95% CI)	MdAPE	CCC	Bias
Calories	84.6 kcal	40.2% (38.4%–42.0%)	29.2%	0.735	−28.6 kcal
Edible mass	58.2 g	28.3% (27.4%–29.3%)	23.7%	0.740	−34.0 g

MACRONUTRIENTS

Beyond calories

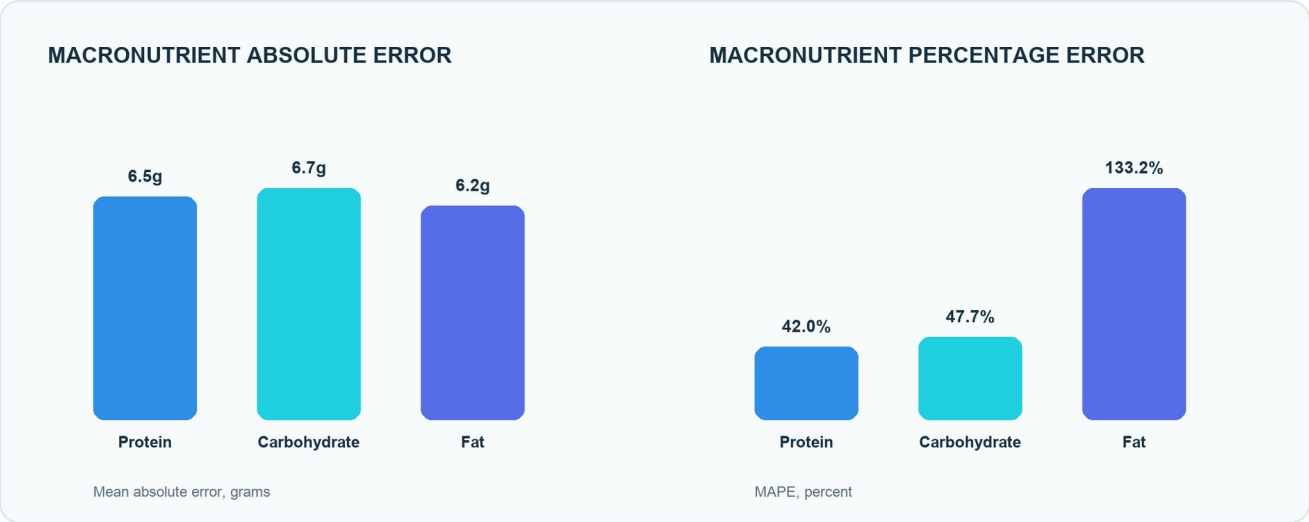

Figure 2. Fitmo macronutrient error across all completed Nutrition5k meals.

Nutrient	MAE	MAPE (95% CI)	MdAPE	CCC	Within 25%
Protein	6.5 g	42.0% (38.2%–47.0%)	29.1%	0.758	44.2%
Carbohydrate	6.7 g	47.7% (43.9%–52.1%)	29.7%	0.576	42.6%
Fat	6.2 g	133.2% (113.2%–155.2%)	42.0%	0.648	31.1%

Percentage errors are naturally harsher for nutrients present in small amounts. For example, a 4 g miss against a 2 g reference produces 200% error even though the absolute difference is small. Product copy should therefore emphasize estimated nutrition guidance, not exact measurement.

ROBUSTNESS

Performance changes with meal complexity

Figure 3. Prespecified ingredient-count strata; no post-hoc sample filtering.

Meal complexity	n	Calorie MAPE	Mass MAPE	Ingredient Jaccard
1 item	984	51.3%	34.4%	0.885
2–3 items	815	38.6%	26.0%	0.708
4–7 items	1066	35.8%	26.7%	0.546
8+ items	364	26.3%	21.7%	0.419

Complex plates remain the hardest operating condition because foods overlap and oils, sauces, and mixed-dish components may be visually hidden. Fitmo presents results as editable estimates so users can correct uncertain items.

TIER 2

Published commercial and app context

The studies below are useful context, but not a single league table. They use different images, cohorts, app versions, image resolutions, and sometimes different outcomes. Rows are ordered by the reported calorie MAE when available; the comparability grade controls how strongly they should influence claims.

System / study	Dataset and n	Reported result	What can be concluded	Grade
Fitmo	Nutrition5k, n=3,229	Calorie MAE 84.6 kcal; macros reported	Primary current result	A
Foodvisor (DietAI24 study)	Nutrition5k subset, n=1,000	Energy MAE 168.5 kcal; P 12.9 g, C 21.9 g, F 8.24 g	Same dataset, different subset and input processing	B
Calorie Mama (DietAI24 study)	Nutrition5k subset, n=1,000	Energy MAE 966 kcal; P 16.5 g, C 30.4 g, F 14.5 g	Same dataset, different subset and app version	B
Foodvisor independent validation	30 prepared meals	Carbohydrate absolute error 13.9±12.4 g	Small, different protocol; no comparable calorie metric	C
Calorie Mama recognition study	185 web food images	Top-1 recognition 62.9%; top-5 87.6%	Identity only; no portion or nutrient sizing	C
Foodvisor recognition study	185 web food images	Top-1 recognition 46.2%; top-5 71.5%	Identity only; no portion or nutrient sizing	C

System / study	Dataset and n	Reported result	What can be concluded	Grade
SnapCalorie	Vendor FAQ; undisclosed evaluation detail	Claims ~15% mean calorie error	Vendor-reported; cannot be compared statistically	D

Claim discipline

Fitmo may accurately say it outperformed FatSecret on the exact 3,229-meal cohort. It should not say it has statistically beaten Foodvisor, Calorie Mama, or SnapCalorie until those systems are tested on the same locked images and scoring code.

TIER 3

Academic research reference points

Academic systems often train directly on Nutrition5k and evaluate on a held-out split from that same distribution. These results show what specialized research models can achieve, but they are not fair product comparisons with an unseen, general-purpose meal-photo workflow.

System	Dataset / setup	Reported result	Key limitation	Grade
DietAI24	Nutrition5k subset, n=1,000	Energy MAE 68.2 kcal; P 3.88 g, C 5.98 g, F 4.01 g	Study-specific workflow; different cohort	B
Nutrition5k RGB baseline	Official train/test split	Calorie MAE 70.6 kcal; PMAE 26.1%	Trained on Nutrition5k; research split	C
Swin-Nutrition	Nutrition5k train/test	PMAE: calories 15.3%, mass 12.5%, P 15.4%, C 20.8%, F 22.1%	Dataset-trained; PMAE is not per-meal MAPE	C
DPF-Nutrition	Nutrition5k train/test	Calories, mass, protein, carbohydrate, and fat reported	Dataset-trained; different model inputs	C
NutritionVerse-Direct	NutritionVerse-Real	MAE: calories 253.7 kcal; mass 98.4 g; P 22.1 g; C 24.2 g; F 14.3 g	Different dataset and split; dataset-trained	C
SNAQ	48 controlled meals with phone depth	Energy MAE 41.2 kcal; relative absolute error 12.7%	Depth-assisted, small controlled study	C

How to read the grades

Grade	Meaning
A	Exact same cohort with compatible per-dish outputs; paired comparison possible
B	Same public dataset but different subset, preprocessing, or evaluation protocol
C	Different dataset, controlled setting, or dataset-trained academic system
D	Vendor-only claim or insufficient methodological disclosure
NR	Outcome not reported

OPERATIONS

Fast enough for an editable diary workflow

Measure	Observed result	Product meaning
Completion	3229/3,229 (100.0%)	High benchmark reliability
Mean turnaround	2.58 s	Measured API round trip in the benchmark environment
Median turnaround	2.50 s	Typical response
95th percentile	3.77 s	Slow-tail planning value
New-study inference cost	$4.21	Operational estimate, not invoiced spend

Latency is environment-dependent and excludes user camera interaction, client upload overhead outside the timed request, and downstream UI rendering. The benchmark measures the inference response path rather than a complete device stopwatch journey.

LIMITATIONS

What this study does not prove

- Nutrition5k was collected in a small number of California cafeterias. It does not cover every cuisine, packaged food, beverage, restaurant, or home presentation.
- The overhead capture angle and visible hardware differ from ordinary handheld phone photos. Real-world accuracy can be lower when scale, lighting, occlusion, or framing is worse.
- Calories and macronutrients are estimated from visual evidence. Invisible oil, ingredients inside mixed dishes, and preparation methods create irreducible uncertainty.
- FatSecret's public result file does not include protein, carbohydrate, or fat, so no exact-cohort macro comparison is possible.
- Tier 2 and Tier 3 values are descriptive context only. Differences in cohort, training exposure, image preprocessing, and metric definitions prevent statistical ranking across those rows.
- This work evaluates estimation accuracy, not clinical outcomes. Results should not guide insulin dosing or replace a dietitian, laboratory measurement, or verified nutrition label.

Recommended next evidence step

Run a prospectively locked, diverse phone-photo study across home, restaurant, packaged, and culturally varied meals, using weighed ingredients and independent human review. Test competing apps on the identical images under a documented version and account configuration. That is the appropriate path to broader market-leading claims.

CONSUMER SUMMARY

A meal photo becomes a useful starting point

In plain language

Fitmo turned one meal photo into an editable nutrition estimate and, on 3,229 weighed real meals, averaged about 85 calories of absolute error. On the exact same meals, FatSecret averaged about 150 calories of error. Fitmo also estimated protein, carbohydrates, and fat in the same pass.

The practical benefit is speed: photograph a meal, review the foods and portions Fitmo recognized, make any needed edits, and save it to the diary. The output is guidance—not a promise that every hidden ingredient or serving size can be measured from a picture.

Approved evidence-led language

Use	Avoid
“Benchmarked on 3,229 weighed Nutrition5k meals.”	“Perfectly measures every meal.”
“Outperformed FatSecret for calorie estimation on the same 3,229-meal cohort.”	“The most accurate food app in the world.”
“Estimates calories, protein, carbohydrates, fat, and portions from a photo.”	“Medical-grade nutrition measurement.”
“Review and edit your result before saving.”	“No correction is ever needed.”

REFERENCES

Sources and reproducibility

1. Google Research. Nutrition5k dataset and CVPR 2021 paper. https://github.com/google-research-datasets/Nutrition5k
2. Sterling et al. Benchmarking Vision-Language Models for Automated Food Recognition (2026 preprint). https://www.biorxiv.org/content/10.64898/2026.07.26.740845v1
3. Sterling et al. Public benchmark code and released per-dish result files. https://github.com/charterer/vlm-food-benchmark
4. DietAI24: Bridging Vision and Nutrition (peer-reviewed full text). https://pmc.ncbi.nlm.nih.gov/articles/PMC12 589 391/
5. Swin-Nutrition: Predicting Food Nutrition Information Using Vision Transformer. https://pmc.ncbi.nlm.nih.gov/articles/PMC9 656 370/
6. DPF-Nutrition: Food Nutrition Estimation via Depth Prediction and Fusion. https://arxiv.org/abs/2310.11702
7. NutritionVerse dataset. https://arxiv.org/abs/2401.08598
8. NutritionVerse-Direct. https://arxiv.org/abs/2405.07814
9. SNAQ preclinical validation. https://pmc.ncbi.nlm.nih.gov/articles/PMC7 142 738/
10. SNAQ and Calorie Mama carbohydrate-estimation comparison. https://pmc.ncbi.nlm.nih.gov/articles/PMC11 571 748/
11. Independent Foodvisor carbohydrate-estimation evaluation. https://pmc.ncbi.nlm.nih.gov/articles/PMC8 266 981/
12. Image-recognition comparison of commercial food apps. https://pmc.ncbi.nlm.nih.gov/articles/PMC7 752 530/
13. SnapCalorie FAQ and vendor accuracy statement. https://www3.snapcalorie.com/faq.html
14. IEEE Professional Communication Society. Write Effective Reports. https://procomm.ieee.org/communication-resources-for-engineers/written-reports/write-effective-reports/
15. Nature Portfolio reporting standards and reproducibility guidance. https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards