Evaluating Performance and Robustness of AI-based Calorie Estimation from Images

Franz Rott
University of Ulm
Ulm, Germany
franz.rott@uni-ulm.de

ABSTRACT

As a pressing health condition with serious personal and societal impairments, obesity is due to the imbalance between calories consumed and metabolized. To counter this imbalance, emerging areas of research focus on automatically estimating the caloric content of food and thereby supporting self-regulation. A promising approach bases its estimation on images of meals and predicts its calories using a trained artificial intelligence model. This study evaluates the robustness of such an AI-based calorie prediction model across varying photographic angles and image settings, including brightness, contrast, and saturation adjustments. For testing the models performance we create a custom dataset derived from the Nutrition5k dataset. It consists of overhead and side-angle images with systematic variations.

Contrary to previous studies we found significantly large prediction discrepancies compared to the ground truth. While image settings had little systematic impact on performance, we found that the estimation quality was further reduced when the images were taken from a side angle instead of overhead. These findings provide actionable insights for end users.

This study highlights the limitations of current visual AI models in predicting calories in meals and stresses the importance of testing these algorithms on a diverse data set. Future work should build on these findings to make the models more robust.

1 INTRODUCTION

According to an estimation by the WHO 16% of the global adult population was obese in 2022, a proportion which more than doubled since 1990 [14] and is likely to increase further [6]. Obesity is accompanied by an elevated risk for cardiovascular diseases, type 2 diabetes, and heightened mortality [14]. Especially during childhood it also harms psychosocial well-being and increases the likelihood of being bullied and stigmatized [14]. Besides grave personal impairments, obesity also has a negative impact on society. The yearly global costs of overweight and obesity are estimated to be US$ 3 trillion by 2030 and over US$ 18 trillion by 2060. Obesity results from an imbalance between energy intake and expenditure and is influenced by both psycho-social and genetic factors [14]. As a study by Zhou et al. [22] revealed, humans are inherently bad at estimating the calories contained in a meal. On average only for 25% of dishes, the estimation error was less than 20%. A possible solution is to replace the error-prone human estimation with an AI model. While most current approaches need the user to upload photos of their meals [15–17] first implementations use online footage from hat or body-worn cameras [2]

In both cases, a backend AI is used to estimate calories. Research is at an early stage, and little is known about the robustness of these approaches. We address this research gap by assessing both the model's performance and the influencing photographic factors.

2 RELATED WORK

2.1 State of the Art AI Models for Calorie Detection

In 2009 the Pittsburgh Fast-Food Image Dataset (PFID) was introduced containing 4,545 static images, 606 stereo image pairs, 303 360° videos, and 27 privacy-conscious videos capturing eating events [4]. The dataset consists of 101 foods from 11 popular fast food chains with 3 instances per class and 8 images per instance [4]. For half a decade it remained the only publicly available dataset for food classification. AI models trained on this data include Support Vector Machine (SVM) classifiers by Chen et al. [4] and Yang et al. [20] which depending on whether color histogram methods were used achieved a modest accuracy of 11% and 28%. Although marking an important step forward for automatic calorie detection, the dataset was limited in its scope (only containing fast food images) and created under artificial and laboratory conditions.

Both issues were addressed by Bossard et al. [3]. They created the Food-101 dataset containing 1,000 photos of each of the 101 most popular and consistently named dishes. It was used to train multiple AI models for image classification and calorie detection. Liu et al. [13] tested various deep learning architectures on food recognition, with the best performing being a CNN, reaching a top-1 accuracy of 77.4%. Ciocca et al. [5] chose a similar approach comparing different CNN architectures. Residual Networks (ResNet-50) performed best with a top-1 accuracy of 82.5%. VijayaKumari et al. [19] used Efficientnetb0, a transfer learning technique, to reach a categorization accuracy of 80%.

Other studies follow similar approaches partly with self-generated and not publically available datasets. Joutou and Yanai [11] used Multi-Kernel-Learning (MKL) to achieve a classification rate of 61.3% for 50 kinds of foods. Hoashi et al. [10] created the Food85 dataset and achieved a food classification accuracy of 62.5% by combining MKL and Bag-of-Features (BoF) Other food category recognition systems can be found in the review by Zhang et al. [21].

Although accurate food classification is important, some use cases including dietary assessment require further information like calorie content.Haris et al. [8] utilized deep learning techniques, especially CNN combined with ResNet50 for feature extraction, to analyze images, and identify food items with an accuracy of 75.8%. While the approach can also estimate portion sizes and predict the caloric content, no metrics were reported.

By now there exist similar open-source food datasets like the Open Images V6-Food Dataset [1], School Lunch Dataset, Vietnamese Food Dataset [2], or MAFood-121 Dataset [1]. Together with the Food-101 Dataset [3] they were used by Han et al. [7] to train YOLOv8, a food items detection model with an accuracy of 75.4%. Again the approach is generally capable of calorie

[1] https://www.kaggle.com/datasets/programmerrdai/open-images-v6
[2] https://universe.roboflow.com/nhh/vietnamese-food/dataset/1

prediction but without reporting performance. Liang and Li [12] propose a calorie estimation method that leverages Faster R-CNN for food. The system requires two images per food item (top and side views). It uses volume estimation combined with density and energy values to calculate calories with an estimation error generally below ±20%. Thames et al. [18] trained a computer vision algorithm for predicting calories and macronutrients of real-world dishes with an accuracy exceeding those of professional nutritionists.

To summarize there is an emerging field of research regarding image-based AI calorie detection. The main scientific focus lies on food classification. Currently, there is little evidence of how well the improving classification also translates to accurate calorie prediction. Studies that address this research question neglect to explore how photographic variables influence performance. We aim to close this research gap by focusing on calorie prediction accuracy under varying photographic conditions, such as angle, lighting, and contrast. Both factors are speculated to have an impact because of the presumed imbalance in the dataset. For our study, we focus on the state-of-the-art AI model trained by [8], because of its accessibility via its API. The model and the Food-101 dataset it was trained on are described in more detail in the following Sections.

2.2 Food-101 Dataset

The Food-101 Dataset was created by Bossard et al. [3] to make a publically available food recognition dataset that is less artificial, contains noise is generally less laboratory. This was implemented by downloading randomly chosen images of the 101 most popular and consistently named dishes from foodspotting.com. Each class consists of 750 train and 250 manually cleaned test images with a maximum side length of 512 pixels. The noise, especially false labeling and intense coloring was kept for the training data set to account for some real-world noise.

Although marking a clear improvement over former public datasets like the PFID Chen et al. [4], Food-101 still has some potential shortcomings. Since coming from a community-driven dish discovery website one must assume that most meals are photographed approximately from above and with "normal" settings to appeal to the audience. This might not properly represent the images used for calorie estimation, especially if they are captured by a body or hat-worn camera. Furthermore, 101 dishes represent only a fraction of the actual food variety.

2.3 Foodvisor AI and API

The Food-101 Dataset was utilized by Haris et al. [8] to train and and test the Foodvisor AI model, designed to recognize food items and estimate their caloric content. To augment the training dataset, preprocessing techniques such as random cropping and horizontal flipping were applied, ensuring the model's robustness against image orientation and composition variations. The model is built on a Convolutional Neural Network (CNN) architecture, paired with a residual network (ResNet50) He et al. [9] for feature extraction. ResNet50 incorporates residual mapping and skip connections, which help mitigate vanishing gradient problems while preserving learned features across layers. The extracted feature maps serve as input to the subsequent layers of the CNN, optimizing the model's ability to generalize across different food categories. The combined approach achieved a classification accuracy of 77.9%, with further reported metrics including a Recall of 77.9%, Precision of 77.7%, and F1 score of 77.6%.

2.4 Nutrition5k Dataset

The Nutrition5k dataset, introduced by Thames et al. [18], addresses the challenge of estimating the nutritional content of generic food items from visual data. It consists of video streams, depth images, and component weights of 5,000 diverse real-world food dishes captured from Google cafeterias custom scanning rig. The value for nutritional content and calorie amount are derived from the USDA Food and Nutrient Database.

The dataset was used to train a multitask learning model based on InceptionV2, which regressed calories, macronutrient weights, and total mass. The best-performing variation of this model achieved a mean average percentage error (MAPE) of 16.5%, outperforming professional nutritionists in portion size estimation. Despite limitations such as geographic biases, Nutrition5k provides a robust foundation for developing models that can predict caloric and nutritional content from food images.

3 EXPERIMENTAL APPARATUS

3.1 Dataset

For testing the capabilities of the Foodvisor-AI, we base our dataset on the Nutrition5k dataset Thames et al. [18]. For constructing our test dataset we extract the image taken from straight above and from a 45° angle for 100 randomly chosen dishes. For each of the 200 images, we create 7 variations using the ... Python library: normal, +20% brightness, -20% brightness, +20% contrast, -20% contrast, +20% saturation, -20% saturation. We assume the original image best reflects the average photographic setting. We focus on the above-mentioned image settings since they vary the most in everyday settings since they are directly influenced by changing light conditions, like time of day or indoor versus outdoor environments. The +/- 20% adjustments was chosen by the authors to simulate realistic variations in everyday photography, ensuring the model's robustness to common lighting and environmental conditions. An example of the variations of a dish can be found in Figure 1.

Our final dataset therefore consists of 1,400 images. We restrict it to that amount to keep the API costs reasonable. We go more into detail about how we utilize the API in the following Section.

3.2 Procedure

We build a connection to the Foodvisor AI via its API to individually prompt the images of our created dataset. We retrieve the amount of estimated calories and save the result for later analysis. Because of the two orthogonal dimensions *angle* with two values and *image settings* with seven, we have a total of 14 groups we can compare. We are especially interested in the accuracy within each group and the variability between groups. We aim to identify groups for which the prediction is rather inaccurate to drive the improvement of AI training (data) and derive actionable insights for real-world (life) usage of these models.

3.3 Measures

Descriptive Analysis: To assess the performance of calorie predictions, we evaluated descriptive metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the proportion of predictions with errors below 20% ($\text{Prop}_{<20\%}$). MAE measures the average magnitude of errors without considering

Figure 1: Variations in image conditions for two photographic angles (overhead and side). Each row corresponds to one photographic angle, while the columns represent different image settings, including RGB, adjusted brightness, saturation, contrast, and others. This visualization highlights the diversity in image conditions used in the study.

(a) Overhead Original **(b) Overhead Brightness+** **(c) Overhead Brightness-** **(d) Overhead Contrast+** **(e) Overhead Contrast-** **(f) Overhead Saturation+** **(g) Overhead Saturation-**

(h) Side Angle Original **(i) Side Angle Brightness+** **(j) Side Angle Brightness-** **(k) Side Angle Contrast+** **(l) Side Angle Contrast-** **(m) Side Angle Saturation+** **(n) Side Angle Saturation-**

their direction:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

where y_i is the ground truth, $\hat{y}_i$ is the predicted value, and n is the number of predictions. MAPE provides a percentage-based measure of error relative to the ground truth:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

Additionally, $\text{Prop}_{<20\%}$ quantifies the proportion of predictions with a relative error below 20%:

$$\text{Prop}_{<20\%} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right| < 0.2\right)$$

where $\mathbb{1}$ is the indicator function. We also evaluated signed error metrics, including Mean Error (ME) and Mean Relative Error (MRE), which reflect whether predictions systematically overestimate or underestimate the ground truth:

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i), \quad \text{MRE} = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\frac{1}{n}\sum_{i=1}^{n} y_i} \times 100\%$$

Inference Statistics: To determine whether the mean absolute difference ($\text{AbsDiff} = |\hat{y}_i - y_i|$) is significantly different from zero, we employed a one-sample t-test on absolute differences. The hypotheses are:

$$H_0 : \mu_{\text{AbsDiff}} = 0 \quad \text{(mean absolute difference is zero)}$$
$$H_1 : \mu_{\text{AbsDiff}} \neq 0 \quad \text{(mean absolute difference is not zero)}$$

The test statistic is calculated as:

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

where $\overline{x}$ is the sample mean of the absolute differences, $\mu_0 = 0$ is the hypothesized mean, s is the sample standard deviation, and n is the sample size. This test evaluates the magnitude of prediction errors independent of their direction by focusing on absolute differences. Each condition is tested against the ground truth, the base image settings condition of the same photographic angle, and the same image setting under a different photographic angle. For instance, images taken from above with reduced brightness are compared to the ground truth, the original overhead images, and the side angle images with reduced brightness. For all statistical test a standard significance level of p < 0.05 is chosen.

4 RESULTS

4.1 Descriptive Analysis

We evaluated the performance of calorie predictions across different photographic angles and image settings. To this end we use the descriptive metrics described in Section 3.3

MAPE Analysis. As shown in Figure 2, images captured from overhead exhibited lower MAPE across all settings compared to side-angle images. Under normal settings, the MAPE for overhead images was 94%, while side-angle images had a MAPE of 113%. Decreasing brightness improved overhead predictions, reducing MAPE to 91%, but increasing it elevated the MAPE to 95%. The opposite trend was observed for side-angle images, with brightness decreases leading to a MAPE of 116% and increases improving MAPE to 108%. Higher contrast yielded better results with an MAPE of 89% for overhead and 109% for side angle images, whereas lower contrast decreased it (97% overhead, 125% side angle). Both adjusted saturation conditions led to a lower MAPE in the overhead condition (86% lower saturation, 90% higher saturation), which was only the case for higher saturation in the side angle condition (109%, 129% for lower saturation).

Prop $< 20\%$ Analysis. The proportion of predictions with a relative error below 20% (Figure 3) was mostly higher for images taken from a side angle compared to overhead images. In the side angle condition, the proportion varied between 8% and 12% and between 9% and 14% for the side angle images.

MRE Analysis. The Mean Relative Error (MRE) results, which can be found in Figure 4 highlight a notable difference between the two photographic angles. Accross image settings, the overhead condition displays a more negative value, with the normal setting leading to a MRE of -29% (the same as for increased contrast). Only increased saturation performs even worse (-31%). Lower brightness (-24%), higher brightness (-26%), lower contrast (-26%), and lower saturation (-28%) yields better results in the overhead condition. The side angle condition portraits a different trend. Only lower brightness displays the same low absolute MRE of 5% as the normal setting (-5%). Higher brightness (-8%), lower contrast (-10%), higher contrast (-17%), lower saturation (-10%), and higher saturation (-13%) show a decreased performance.

A detailed analysis of Mean Absolute Error (MAE) and Mean Error (ME), can be found in Appendix (Figure 5 and Figure 6).

Figure 2: Mean Absolute Percentage Error (MAPE) across different photographic angles and image settings. Bars are grouped by image settings, with blue representing overhead and orange representing side angle conditions.

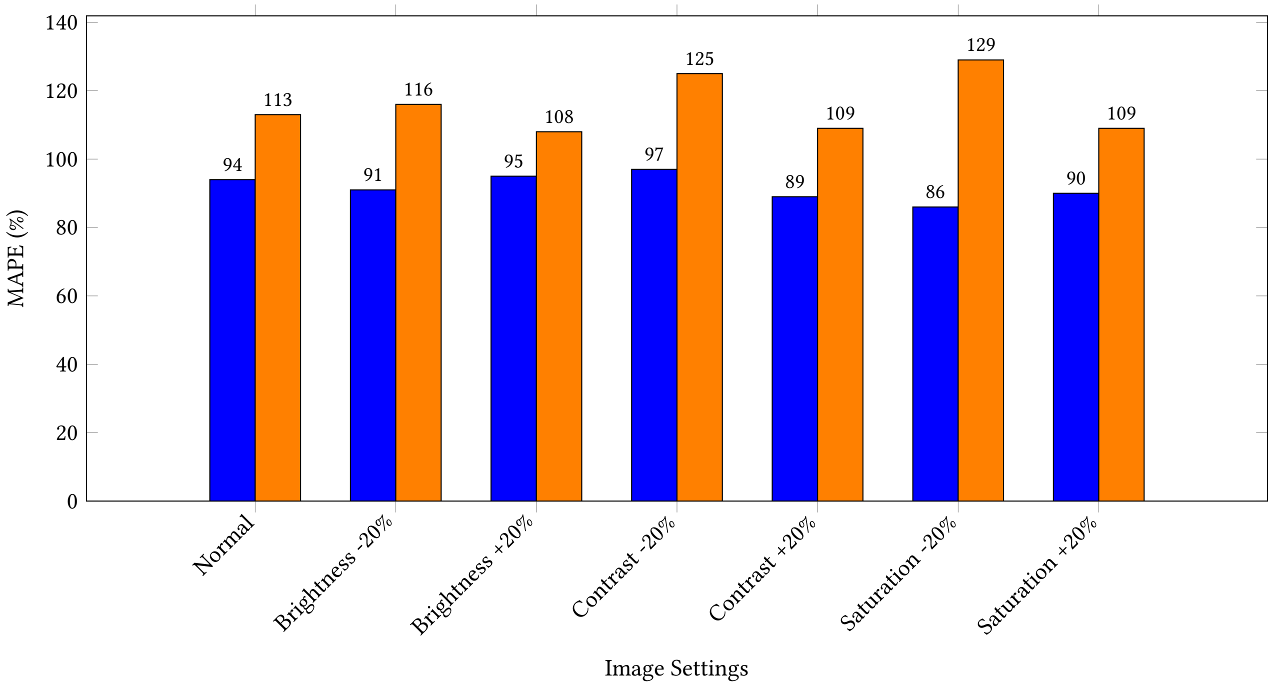
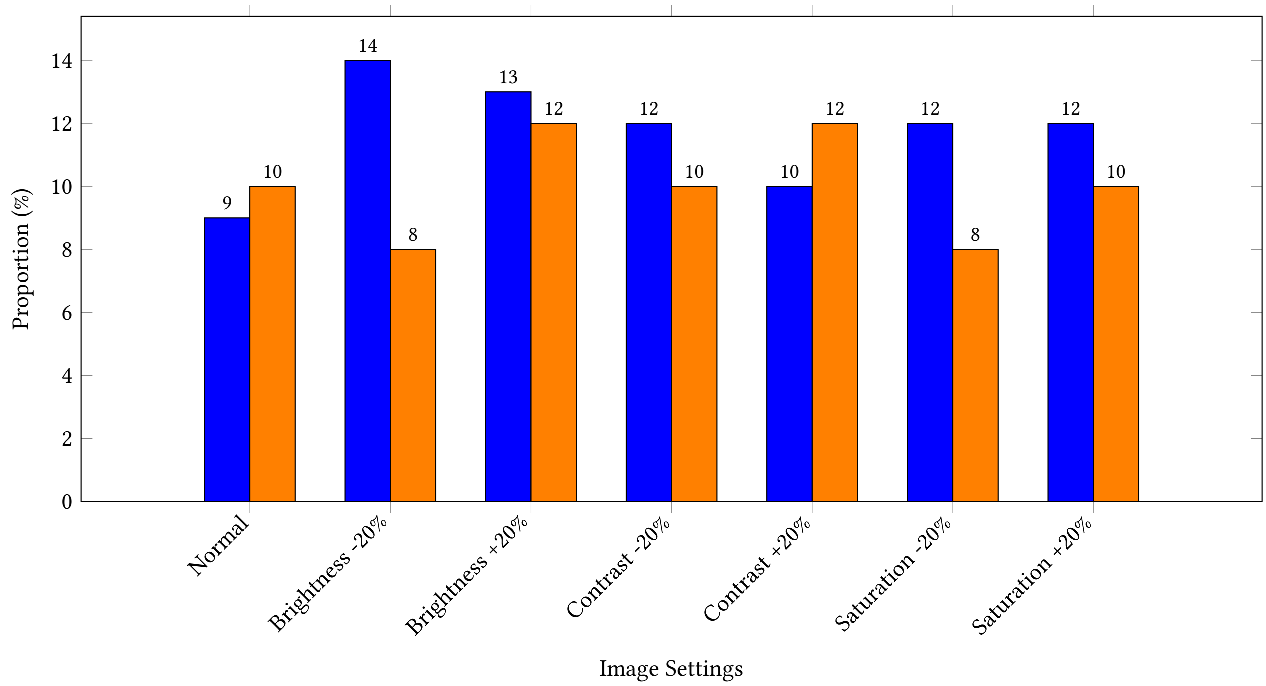


Figure 3: Proportion of Predictions with Relative Error Below 20% (Prop $< 20\%$) across different photographic angles and image settings.

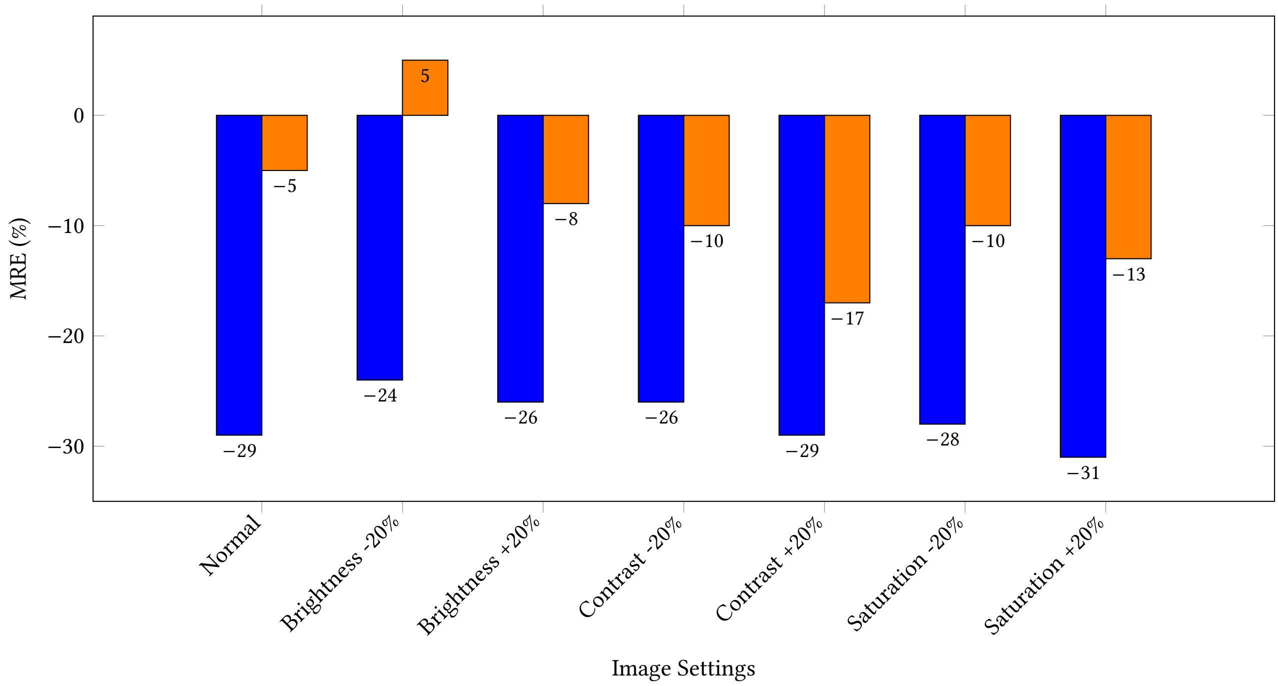


4.2 Inference Statistics

To assess whether the descriptive differences presented in 4.1 are significant we apply one-sample t-tests on the mean absolute differences (MAD) for each condition. We compare each of the 14 conditions against a) the ground truth, b) counterpart, and c) base condition. For example for the overhead condition with reduced brightness the counterpart would be side angle with reduced brightness and the base condition would be overhead without adjusted image settings.

The results, presented in Table 1, show significant deviations from the ground truth as well as their counterpart and base case for all conditions ($p < 0.0001$). Still the results differ in terms

Figure 4: Mean Relative Error (MRE) across different photographic angles and image settings.



of MAD. The largest MAD was observed when tested against the ground truth and higher for side angle images compared to overhead). Since this represents the MAE metric consult Section 4.1 for more details. The second largest MAD is observable in relation to the counterpart, with values ranging from 156 to 198. Comparing to the base condition yields the lowest MAD, especially for the overhead images. For this photographic angle the MAD varied between 82 and 129, depending on the image settings, while for side angle images the range was between 99 and 149.

5 DISCUSSION

5.1 Key Scientific Insights

The performance of the foodvisor AI is generally poor for the applied custom Nutrition5K data set.

As described in Section 4.1 MAPE was lower for the overhead condition, while side angle images performed better on MRE. The Prop < 20% on the other hand did not indicate a clear trend. This shows that images taken from overhead perform better in general but are more directionally biased. The vastly negative MRE indicates that the amount of calories is usually underestimated. While more prominent for overhead images this is also the case in the side angle condition. There are no clear identifiable trends regarding image settings in terms of performance. This is in accordance with the results reported in Section 4.2, namely that the MAD is larger between photographic angle conditions (counterpart) then within one angle but relative to the base settings (base condition).

To summarize, the foodvisor AI model performs poorly on this test data set and shows some robustness regarding adjusted image setting but little when photographic angles are varied.

5.2 Threat to Validity and Limitations

Several limitations must be considered when interpreting these results. First, the dataset was restricted to 100 arbitrarily chosen images per condition to contain API costs. Because of the limited pool of dishes this may limit the generalizability of the presented findings. We still regard our findings as sufficiently robust and verified that the 100 images cover a broad array of dishes.

While we attempted to simulate real-world conditions, the chosen variations in image settings (±20%) may not cover the full spectrum of variability and may be chosen sub-optimal. We regard this critique as negligible since these factors are continuous and allow infinitely many combinations, a complexity that cannot be adequately covered in controlled experiments. Another image settings related limitation comes with the assumption that the original image is similar across dishes and represents an average setting. Even in case these assumptions are flawed we find it highly unlikely that this impairs our results, since image settings was a minor influencing factor, compared to the overall poor performance and the photographic angle variation.

Additionally, the reliance on a single AI model and training set limits the scope of the study. Results may differ with alternative models or training datasets. The study also assumes accurate ground truth calorie values, which may introduce a bias if errors exist in the labeled dataset.

5.3 Impact and Future Work

With this study, we present first evidence that the current training and testing of visual AI-based calorie estimation models is flawed. Likely to the train and test dataset stemming from the same sources we regard the reported performance as overestimated. We caution users to rely solely on such AI applications, especially obese people since the models tend to underestimate the caloric content. When still applying these models the findings of this

Table 1: Mean Absolute Differences (MAD) for Calorie Predictions. Each condition is tested against the ground truth, its counterpart (same setting under a different photographic angle), and its base condition (unmodified images for the same angle). All results are significant ($p < 0.05$).

Condition	MAD vs. Ground Truth	MAD vs. Counterpart	MAD vs. Base Condition
Overhead (Normal)	**202**	**198**	-
Overhead (Brightness -20%)	**183**	**181**	**87**
Overhead (Brightness +20%)	**201**	**188**	**88**
Overhead (Contrast -20%)	**205**	**177**	**82**
Overhead (Contrast +20%)	**186**	**160**	**110**
Overhead (Saturation -20%)	**186**	**182**	**129**
Overhead (Saturation +20%)	**188**	**156**	**86**
Side Angle (Normal)	**226**	**198**	-
Side Angle (Brightness -20%)	**241**	**181**	**148**
Side Angle (Brightness +20%)	**218**	**188**	**99**
Side Angle (Contrast -20%)	**232**	**177**	**108**
Side Angle (Contrast +20%)	**188**	**160**	**149**
Side Angle (Saturation -20%)	**206**	**182**	**137**
Side Angle (Saturation +20%)	**201**	**156**	**100**

studies suggest to make sure the photo is taken from above with image settings within a broad range of reasonable values.

Future research should experiment with larger and more diverse datasets with a greater number of image setting combinations. Additional variations like occlusion and motion blur would be beneficial to reflect conditions encountered in wearable camera setups. Exploring other AI models' performance and robustness or engaging in model training would widen the scope of this study. Moreover, integrating more contextual information, such as depth images, could enhance prediction accuracy. Future research could pave the way for more reliable AI-based dietary assessment tools, contributing to combating obesity, elevating public health and fostering a more reasonable usage of such tools.

6 CONCLUSION

This study assesses the performance and robustness of a visual AI-based model when predicting calorie amount under different photographic angles and adjusted image settings. The results reveal an overall poor performance with little robustness regarding the angle from which images were taken. According to the found result, the image settings represent a subsidiary variable. This work provides a foundation for future research advancements in AI-driven calorie trackers. Building upon the results of this study and incorporating the suggested improvements, future research could make evaluative studies more refined and AI-tools more performant. Such a robust estimation model represents the prerequisite for using visual medical wearables to help with obesity and promote overall health.

REFERENCES

[1] Eduardo Aguilar, Marc Bolaños, and Petia Radeva. 2019. Regularized uncertainty-based multi-task learning model for food analysis. *Journal of Visual Communication and Image Representation* 60 (2019), 360–370.

[2] Shengjie Bi and David Kotz. 2022. Eating detection with a head-mounted video camera. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. IEEE, 60–66.

[3] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101–mining discriminative components with random forests. In *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part VI 13*. Springer, 446–461.

[4] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. 2009. PFID: Pittsburgh fast-food image dataset. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 289–292.

[5] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2018. CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding* 176 (2018), 70–77.

[6] Richard Dobbs, Corinne Sawers, Fraser Thompson, James Manyika, Jonathan R Woetzel, Peter Child, Sorcha McKenna, and Angela Spatharou. 2014. *Overcoming obesity: an initial economic analysis*. McKinsey global institute.

[7] Michelle Han, Junyao Chen, and Zhengyuan Zhou. 2024. NutrifyAI: An AI-Powered System for Real-Time Food Detection, Nutritional Analysis, and Personalized Meal Recommendations. arXiv:2408.10532 [cs.CV] https://arxiv.org/abs/2408.10532

[8] Mohammad Haris, Mohd Sarim, Vaibhav Mishra, and Garima Singh. 2024. FoodVisor: A Food Calorie Estimation System. In *2024 Second International Conference on Advances in Information Technology (ICAIT)*, Vol. 1. IEEE, 1–7.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[10] Hajime Hoashi, Taichi Joutou, and Keiji Yanai. 2010. Image recognition of 85 food categories by feature fusion. In *2010 IEEE International Symposium on Multimedia*. IEEE, 296–301.

[11] Taichi Joutou and Keiji Yanai. 2009. A food image recognition system with multiple kernel learning. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 285–288.

[12] Yanchao Liang and Jianhua Li. 2017. Computer vision-based food calorie estimation: dataset, method, and experiment. *arXiv preprint arXiv:1705.07632* (2017).

[13] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. 2016. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In *Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14*. Springer, 37–48.

[14] World Health Organization et al. 2024. Obesity and overweight. 2024: https://www. who. int/news-room/fact-sheets/detail/obesity-and-overweight. *Updated January* (2024).

[15] Ioannis Papathanail, Jana Brühlmann, Maria F Vasiloglou, Thomai Stathopoulou, Aristomenis K Exadaktylos, Zeno Stanga, Thomas Münzer, and Stavroula Mougiakakou. 2021. Evaluation of a novel artificial intelligence system to monitor and assess energy and macronutrient intake in hospitalised older patients. *Nutrients* 13, 12 (2021), 4539.

[16] G Ramkumar, CB Venkatramanan, V Manonmani, S Palanivel, Ata Kishore Kumar, et al. 2023. A Real-time Food Image Recognition System to Predict the Calories by Using Intelligent Deep Learning Strategy. In *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*. IEEE, 1–7.

[17] S Sapna, KM Yaswanth, Praveen Kumar, et al. 2023. Calorie Estimation of Food and Beverages using Deep Learning. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 324–329.

[18] Quin Thames, Arjun Karpur, Wade Norris, Fangting Xia, Liviu Panait, Tobias Weyand, and Jack Sim. 2021. Nutrition5k: Towards automatic nutritional understanding of generic food. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8903–8911.

[19] G VijayaKumari, Priyanka Vutkur, and P Vishwanath. 2022. Food classification using transfer learning technique. *Global transitions proceedings* 3, 1 (2022), 225–229.

[20] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. 2010. Food recognition using statistics of pairwise local features. In *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2249–2256.
[21] Yudong Zhang, Lijia Deng, Hengde Zhu, Wei Wang, Zeyu Ren, Qinghua Zhou, Siyuan Lu, Shiting Sun, Ziquan Zhu, Juan Manuel Gorriz, et al. 2023. Deep learning in food category recognition. *Information Fusion* 98 (2023), 101859.
[22] Jun Zhou, Dane Bell, Sabrina Nusrat, Melanie Hingle, Mihai Surdeanu, Stephen Kobourov, et al. 2018. Calorie estimation from pictures of food: Crowdsourcing study. *Interactive journal of medical research* 7, 2 (2018), e9359.

A DESCRIPTIVE METRIC VISUALIZATIONS

Figure 5: Mean Absolute Error (MAE) across different photographic angles and image settings. Bars are grouped by image settings, with blue representing overhead and orange representing side angle conditions.

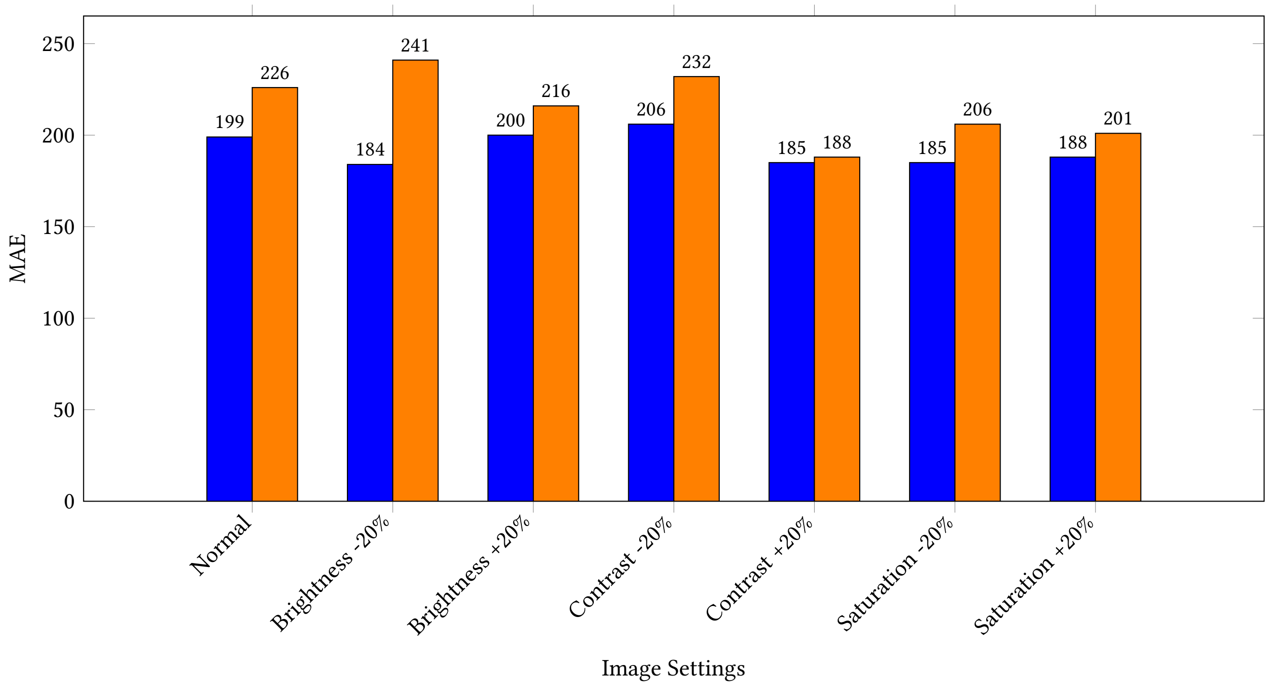


Figure 6: Mean Error (ME) across different photographic angles and image settings.