JOURNAL OF
EMERGING INVESTIGATORS

Article

Building deep neural networks to detect candy from photos and estimate nutrient portfolio

Louwyn Mingze An[1], Roy Junda Zhu[1], Ruopeng An[2]
[1] John Burroughs School, St. Louis, Missouri
[2] Brown School, Washington University in St. Louis, St. Louis, Missouri

SUMMARY

Approximately one third of American youth consume candy on a given day. Consuming excess candy contributes to added sugar intake and may lead to tooth decay and other health concerns. Diet-tracking apps may inform and help regulate candy consumption but depend on the availability of annotated candy image data and predictive models. A recent review documented differences in daily intakes of calories and macronutrients between app predictions and ground truths ranging from 1.4%–10.4%. Transfer learning-based deep neural networks could outperform those benchmarks and reduce the error margin. We built a dataset of 1,008 images comprising nine common candy types and developed four neural network models to detect candy pieces. The best-performing model achieved a mean average precision of 0.8736 for localizing candy pieces of different types in the validation dataset and an accuracy of 99.8% for predicting the quantity and types of multiple candy pieces in the test dataset. By combining candy type-specific nutritional information obtained from the nutrition facts label, the model accurately estimated (within an error margin of 0.5%) the aggregate nutrient portfolios, including total calories, total fat, saturated fat, cholesterol, sodium, carbohydrates, total sugars, and added sugars, of all candy pieces shown in an image. This study demonstrates the feasibility of automating candy calorie/nutrition counting using photos, which may facilitate the development of diet-tracking apps to provide real-time, accurate nutritional information to inform candy consumption.

INTRODUCTION

Approximately one-third of youth aged 2–18 years in the United States consume candy on a given day (1). On a day of candy consumption, consumers reported an average of 40 g (176 kcal) of candy intake (1). Although candy contributes a relatively modest share of energy, added sugars, and saturated fat to the total diet of most consumers, binge candy intake by some individuals and during holidays (e.g., Halloween) has led to dental and other health concerns (1–3). The 2020–2025 Dietary Guidelines for Americans recommends the consumption of added sugars be less than 10% of daily total energy intake (4). Candy is dense in added sugars but low in nutritional value (i.e., "empty calories") (5). Limiting candy consumption may help cut the daily intake of added sugars to meet the recommendation of dietary guidelines (6).

Real-time, accurate nutrition information helps inform dietary decision-making (7). Many candy pieces are typically contained in a single package on which the nutrition facts label is printed. For instance, a Nestlé "Party Pack" includes 60 pieces of each of the four candy types: Butterfinger, Crunch, Baby Ruth, and 100 Grand. The nutrition facts label provides each candy type's per-serving nutrient portfolio (e.g., total fat, saturated fat, cholesterol, sodium, and total sugar). This presentation impedes candy consumers' access to and assessment of relevant nutritional information for three reasons. First, two or more candy pieces make up a serving, so some "mental division" must be performed to convert the per-serving nutrient portfolio to per piece. Second, the nutrition facts label is generally not printed on the package of individual candy pieces, making it inaccessible during consumption. Finally, individuals often consume two or more candy pieces of the same or different types on a single occasion, increasing the cognitive load of calculating the total nutrient intake from candy even when per-serving nutrient portfolios are available (1).

Smartphone diet-tracking apps may help people monitor daily eating patterns, control weight, and manage chronic conditions (8, 9). A recent review found that diet-tracking apps scored well in usability, incorporated behavioral change constructs, and accurately coded daily energy and nutrient intakes (8). Diet-tracking apps have also been shown to lead to positive changes in dietary behavior (10). Most diet-tracking apps adopt deep learning as their core technology (11). Deep learning is based on artificial neural networks, in which multiple ("deep") layers of processing are used to extract progressively higher-level features from data (12). The layered representation enables modeling highly complex, dynamic patterns, which finds its utility in analyzing "big data", or data massive in scale and challenging to process (e.g., image, video, audio, and text) (13). Recent work has found that deep learning-based approaches, including image classification and object detection, may improve dietary assessment by optimizing efficiency and addressing systematic and random errors in self-reported nutrient intake (14, 15). As evidenced by the exponential growth of diet-tracking apps, nutrition experts and the general public increasingly demand food and diet monitoring automation (8, 9). This demand is best served through data collection and model building.

Since artificial intelligence (AI) models are trained using examples (i.e., data) rather than based on "hard-coded" rules, data availability and quality play an indispensable role in AI model development and applications (16). Several food image datasets have been assembled to enable many deep-learning applications in human dietetics, such as Fruits 360 (about 90,000 images of 131 fruits and vegetables) and Food-101 (about 100,000 images of 101 foods) (17, 18). However, systematically collected and annotated image data concerning commonly consumed candy types remains nonexistent. A recent review reported differences in daily intakes of calories and macronutrients (i.e., carbohydrate, protein, and total fat) between diet-tracking app predictions and ground truths ranging from 1.4%–10.4%, serving as the benchmark for AI-powered energy and nutrient auto-counting (8).

Given the discrepancy between actual and AI-predicted nutritional profiles, we sought to reduce the gap by applying state-of-the-art AI models to candy images. In particular, this study aimed to test the hypothesis that fine-tuning pre-trained deep neural network models to perform object detection tasks would enable accurate estimation of the number and types of candy pieces in a photo and their total nutritional portfolio (19).

Our study contributes to scientific literature in two aspects. First, we built and open-sourced a dataset containing images of nine popular candy types. Each image included four candy types and was annotated using bounding boxes to mark their positions, making the dataset suitable for training multilabel classification or object detection models. Second, we fine-tuned four neural network models of different architectures to detect and localize candy types in an image. We adopted several evidence-based techniques to improve model performances, including transfer learning, data augmentation, normalization, and learning rate optimization. The YOLOv5 model achieved the highest prediction accuracy among the four models on the validation and test datasets. Combined with candy-specific nutritional information, the YOLOv5 model predicted the total nutritional value (e.g., total calories, added sugars, cholesterol, and saturated fat) in an image containing multiple candy pieces of different types within an error margin of 0.5% compared to the ground truths.

In sum, the candy image dataset and deep neural network models developed from this study may advance the dietary tracking of candy consumption. Our study holds the potential to stimulate AI model deployment to promote informed, responsible candy intake as part of a healthy diet.

RESULTS

Dataset

We built and annotated a dataset of 1,008 images, which included nine candy types—Butterfinger, Crunch, Baby Ruth, 100 Grand, Snickers, Twix, 3 Musketeers, Milky Way, and Milky Way Midnight. Each image contained four candy pieces, each of a different type, surrounded by a rectangular bounding box labeled by type (**Figure 1**).

Figure 1: Sample annotated images. Individual candies are annotated with labels and rectangular bounding boxes.

Model Selection

We randomly divided the dataset into training (806 images), validation (101 images), and test sets (101 images). We hand-picked and fine-tuned four pre-trained models in *IceVision* with distinct architectures (Faster R-CNN, RetinaNet, YOLOv5, and EfficientDet) on the training set and evaluated their performances on the validation set using mean average precision (mAP) (**Table 1**). The YOLOv5 model achieved the highest mAP score of 0.8736, followed by EfficientDet (0.8377), RetinaNet (0.7547), and Faster R-CNN (0.6182). Thus, we chose YOLOv5 as the final model, given its top performance in predicting the candy types (i.e., labels) and corresponding bounding box locations. The model is also light in size (27.1 MB).

Model	Highest mAP Score	Number of Epochs
Faster R-CNN	0.6182	18
RetinaNet	0.7547	19
YOLOv5	0.8736	19
EfficientDet	0.8377	17

Table 1: Model performances on the validation set measured by mAP. The mAP score denotes the highest mAP score a model achieved during the fine-tuning process, and the number of epochs is the number of iterations over the entire dataset at which the highest mAP score was obtained. The YOLOv5 model was chosen as the final model, given its highest mAP score over all four models.

Visualization of Model Predictions

We presented side-by-side examples of the ground truth annotations (i.e., bounding boxes and labels) and corresponding YOLOv5 predictions on the test set (**Figure 2**). The locations and sizes of the predicted bounding boxes closely resembled the ground truths, and labels were correctly predicted with high confidence.

Combining Nut Nutrient Information with Model Predictions

The study goal was to train a model to identify the quantity and types of candy pieces in an image, which allowed the calculation of the aggregate nutrient portfolio, such as total calories and added sugars, rather than each candy piece's precise location (i.e., bounding box). Therefore, we counted the number and proportion of matches between the ground truth labels and YOLOv5-predicted labels in the test set. Specifically, we created two matrices, one for ground truth labels and the other for predicted labels. In either matrix, each row represents a single image in the text set, and the nine columns denote the respective candy types. Each cell value in the ground truth (predicted) matrix documented the ground truth (predicted) quantity of a specific candy type in an image. We compared the value of the corresponding cell between the two matrices and made a total of 909 cell-level comparisons (101 images x 9 candy types per image). A match was defined if two corresponding cells in the two matrices share the same value. There were 907 matches among the 909 comparisons, achieving a prediction accuracy of 99.8% by the YOLOv5 model.

We combined candy type-specific nutritional information (e.g., calories per serving) from the nutrition facts labels of the original candy packaging with the YOLOv5 predictions (i.e., predicted candy types and quantity in an image) to estimate the following eight aggregate nutrient portfolios: total calories, total fat, saturated fat, cholesterol, sodium, total carbohydrate, total sugars, and added sugars of all candy pieces shown in an image. We reported the discrepancies between YOLOv5 predicted and ground truth nutrient portfolios (**Table 2**). Across the 12 nutrients, the percentages of discrepancies ranged from 0% (cholesterol) to a maximum of 0.43% (total fat). Our final model shows significant improvement over the previous benchmark of an error margin of 1.4%–10.4% across nutrients (8).

DISCUSSION

Consuming excess candy contributes to daily added sugar intake and may lead to tooth decay and other health concerns (2, 3). Diet-tracking apps may inform and help regulate candy consumption but depend on the availability of annotated candy image data, which, to our knowledge, remains absent. This study built a dataset of over 1,000 images comprising nine common candy types. Each candy piece in an image was annotated using a rectangular bounding box. Based on the dataset, neural network models were developed to identify and localize candy pieces. The final selected model, YOLOv5, achieved an mAP score of 0.87 in identifying candy types and corresponding locations in images and predicted nutritional content with an error rate of less than 0.5%. Based on the COCO dataset, YOLOv5 specializes in object detection and has achieved state-of-the-art performance. These findings confirmed the hypothesis that fine-tuning pre-trained deep neural network models could enable accurate estimation of the number and types of candy pieces in a photo and their overall nutritional portfolio. The study demonstrated the feasibility of using photos to automate candy calorie/nutrition counting. We have made the dataset and the model open source to facilitate data sharing and knowledge dissemination.

Traditionally, deep neural network models required large-scale datasets and were expensive to train, taking days or weeks to train and fine-tune a model, often resulting in a large carbon footprint (20). However, leveraging transfer learning, image preprocessing such as data augmentation, and learning rate optimization techniques showcased that a near-perfect prediction accuracy could be achieved using a modestly-sized dataset. The final model is also light in size. Embedded in a mobile app, users may use the model in real-time to estimate energy and nutrient intakes from candy consumption.

Besides candy consumption, candy images collected from

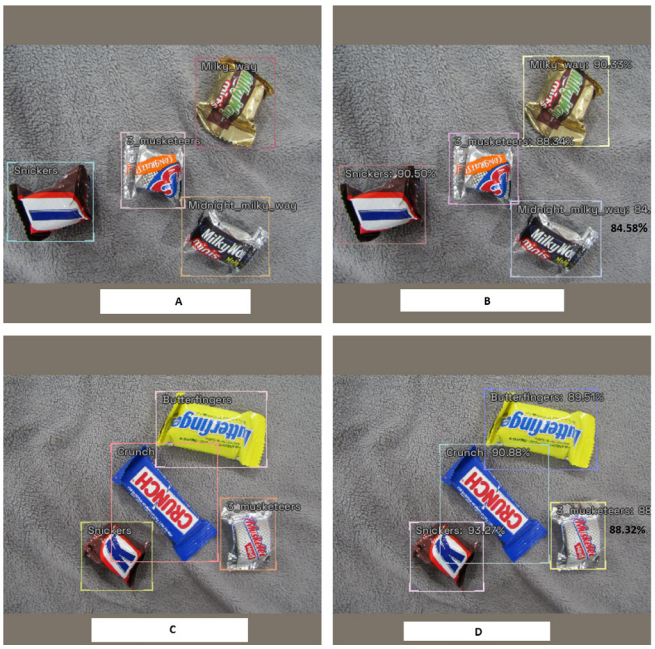

Figure 2: Examples of ground truth annotations and corresponding YOLOv5 model predictions. A) and **C)** Ground truth and **B)** and **D)** Model prediction annotations. "x%" denotes the YOLOv5 model predicted probability of a candy label. Multiple orientations and positions of the same candy pieces helped train the models to robustly recognize the candy objects from different angles and improve model generalizability to new, unseen data.

Model	Predicted Mean	Ground Truth Mean	Percentage of Discrepancy
Total energy (kcal)	239.77	238.83	0.40%
Total fat (g)	10.31	10.27	0.43%
Saturated fat (g)	6.06	6.04	0.42%
Sodium (mg)	108.50	108.12	0.35%
Cholesterol (mg)	3.80	3.80	0.00%
Total carbohydrate (g)	34.48	34.35	0.37%
Total sugars (g)	26.83	26.73	0.35%
Added sugars (g)	24.73	24.65	0.36%

Table 2: Percentages of discrepancies between model-predicted and ground truth image-level aggregate nutrient portfolios. The predicted mean denotes the average nutritional value calculated from the YOLOv5 model-predicted number and types of candy pieces in the test images (n = 909). The ground truth mean denotes the average nutritional value calculated from the actual number and types of candy pieces in the test images. The percentage of discrepancy was calculated by subtracting the model-predicted from the actual nutritional value, dividing it by the actual nutritional value, and then taking the average.

current and future projects can be easily merged with other food image databases, such as Food-101, to build neural network models that can detect a wide variety of food types (18). Such applications can be powerful tools to automate daily diet monitoring, substantially reducing the cognitive burden of those who need to monitor their caloric and nutrient intake for disease or weight management (8, 9). Data collected from the diet-tracking apps may also inform nutritionists and dietitians and facilitate their intervention design or evaluation.

Various psychosocial factors, such as motivation, desire, self-efficacy, attitudes, knowledge, and goal setting, may impact people's dietary behavior (21, 22). While real-time, accurate dietary intake measures are valuable, such information alone is insufficient to motivate behavioral change (23–25). Besides building more powerful and efficient AI models, developers should consider integrating appropriate theoretical constructs for health behavior change into those diet-tracking apps (26). In particular, they may target these psychosocial factors to help users adopt and sustain a healthier diet (10).

This study has several limitations. First, the dataset was modest in size and candy types. Substantially expanding the candy types in future efforts is necessary to enhance the dataset and model's generalizability and applicability. Second, the study serves as a proof-of-concept experiment to demonstrate the feasibility and potential usefulness of training a neural network model to detect candy pieces in a photo. However, many questions remain to be answered. Could the model achieve desirable prediction accuracy given user-provided candy images, which may deviate from those used to train the model (e.g., different resolution, angle, or lighting)? How can the model size be reduced without compromising its accuracy to meet real-world needs where low-end mobile devices have more constrained storage space? Those who do not have a smartphone but want to use the diet-tracking apps may upload images to the cloud for the neural network model to make inferences. Images are one of the many ways people may interact with AI systems, and other popular modes include but are not limited to voice, text, video, and sensors (27). Multimodal human-machine communication could improve model performance and reliability (28), but relevant experiments remain scarce in diet interventions. Moreover, the current project used about 25 samples per candy type, which may not be enough for generalizability. We plan to add more samples to the database to enhance data diversity in future endeavors. We compared our model prediction accuracy with some benchmarks achieved by previous AI-powered diet-tracking apps (8). However, since no previous work has estimated the nutrient portfolios of candy products and the benchmarks pertained to daily total calories and nutrient intakes, such comparison is far from ideal. We also hand-selected four popular deep neural network models with distinct architectures. Still, hundreds of other models exist, and some might outperform the YOLOv5 model. We could not expand the model list due to our budget and time constraints. Finally, the YOLOv5 model was chosen as the final model, given its highest mAP score over all four models. However, because each model was only fine-tuned once, no statistical testing was involved in comparing mAP scores across models. Although statistical testing could be desirable, it will involve heavy computation and can be economically and environmentally burdensome with a large carbon footprint.

In conclusion, limiting candy consumption may help cut daily intake of added sugars to meet the recommendation of dietary guidelines. We built and annotated a dataset containing 1,008 images of nine commonly consumed candy types. Neural network models were trained on the dataset to detect candy pieces of different types. The nutrient portfolios of candy pieces estimated by the final model were nearly identical to the ground truths in the test set. The dataset and model may facilitate the development of diet-tracking apps to provide real-time, accurate nutritional information to inform candy consumption.

MATERIALS AND METHODS

Data

We purchased two packs of candy from a local store—a Nestlé "party pack" containing 60 "fun size" pieces of four candy types (Butterfinger, Crunch, Baby Ruth, and 100 Grand) and a Mars pack containing 125 "minis party size" pieces of five candy types (Snickers, Twix, Musketeers, Milky Way, and Milky Way Midnight). Given the nine candy types, there are a total of 126 unique combinations of four candy types. For each combination, we took eight photos, each containing four candy pieces, one for each type. The candy pieces were placed on a grey rug to improve the identification of candy pieces. Before taking each photo, we swapped candy pieces and rearranged their positions and sides. A Canon PowerShot SX540 digital camera was used to take photos. The dataset contained 1,008 (126 × 8) images in total. All images in the dataset were annotated using VGG Image Annotator (VIA) (29). In an image, a rectangular bounding box was manually drawn around each candy piece with its type labeled. Following standard practice, the dataset was randomly split into three subsets: the training set (80% of the dataset or 806 images), the validation set (10%, 101 images), and the test set (10%, 101 images). This dataset is publicly available at: https://www.kaggle.com/datasets/ruopengan/candy-brands-for-object-detection.

Model

Neural network models were trained to detect and localize candy pieces and their positions given by the rectangular bounding boxes in an image using the *IceVision* module in

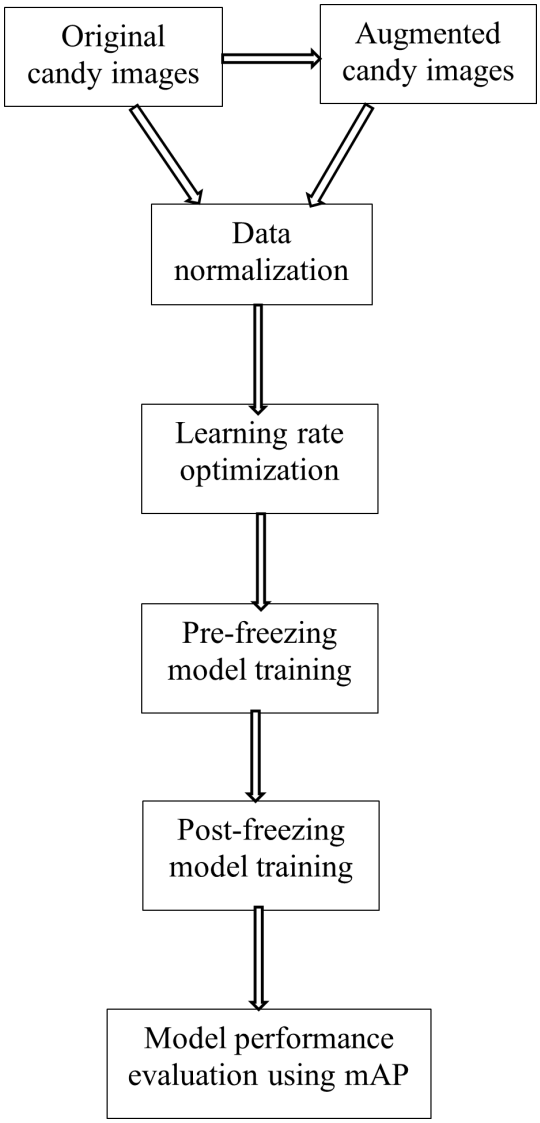

Figure 3: Model training flowchart. We hand-picked and fine-tuned four pre-trained models in IceVision with distinct architectures (Faster R-CNN, RetinaNet, YOLOv5, and EfficientDet) on the training set and evaluated their performances on the validation set using mean average precision (mAP). Each model was trained for a total of 30 epochs. The YOLOv5 model achieved the highest mAP score, followed by EfficientDet, RetinaNet, and Faster R-CNN. Thus, we chose YOLOv5 as the final model.

Python (**Figure 3**). *IceVision* is a computer vision framework offering a curated collection of many high-performance pre-trained models from Torchvision, MMDetection, YOLO, and EfficientDet. It orchestrates an end-to-end deep learning workflow to train models with robust libraries such as *PyTorch Lightning* and *fastai2*.

Given our modest sample size (1,008 images), the best practice called for transfer learning, which enables the knowledge gained while solving one problem to be applied to a different but related problem (30). For example, knowledge obtained while learning to recognize apples, stored as trainable weights in a neural network model, may be used when building a different model to identify avocados. We experimented with four pre-trained models—Faster R-CNN, RetinaNet, YOLOv5, and EfficientDet. These models have different architectures and were widely adopted in object detection tasks.

The original images taken by the camera had a height of 3,864 pixels and a width of 5,152 pixels. The images were first resized to 512 × 512 pixels through cropping (random cropping for the train set and center cropping for the validation set). After data augmentation (described below), the images were further resized to 384 × 384 pixels before being consumed by the models. *fastai2* supports this progressive resizing technique to enhance model performance.

Three evidence-based techniques were used to boost the model performance—data augmentation, normalization, and learning rate optimization. Data augmentation is a technique to increase the diversity of the training set by applying random but realistic transformations, effectively preventing model overfitting (i.e., model fitting exactly against its training data while losing generalizability to new, unseen data). Before being consumed by the model, images in the training set went through a data augmentation pipeline consisting of resizing, zooming, cropping, rotating, and contrast changing. Data normalization ensures that each input parameter (image pixel, in our case) shares a similar statistical distribution, which helps model convergence. Learning rate (LR) plays an essential role in model training. If the LR is too low, it will take a long time to train the model; if the LR is too high, the model may take too large steps and overshoot where the optimal model resides. The LR finder in *fastai2* implements the cyclical LR, enabling the LR to oscillate between reasonable boundary values during training (31). The optimal LRs identified by the LR finder were used for pre-freezing and post-freezing model training.

A two-step model training strategy was adopted. In the first step or pre-freezing phase, an object detection model head (tasked with candy piece localization and type classification) with randomly initiated weights was trained on top of the backbone pre-trained model with its layers, namely trainable weights, frozen. In the second step or post-freezing phase, all model layers were unfrozen and trained simultaneously. The use of data augmentation anticipated more iterations in model training, so pre-freezing and post-freezing training took 10 and 20 epochs, respectively. The pre-freezing model with the lowest validation loss was used to train the post-freezing model, and the post-freezing model with the lowest validation loss was retained as the final model. The final model was tested on the test set, with predictions compared to the ground-truth labels. mAP was used to evaluate model performances. The mAP compares the ground-truth bounding boxes to the model-predicted ones and returns a score ranging from 0 to 1, with a higher score denoting a greater precision level. All models were built in Google Colab Pro using Python 3.10 as the programming language. A Tesla V100 GPU accelerated the model training.

Received: October 10, 2022
Accepted: March 15, 2023
Published: September 13, 2023

REFERENCES

1. Duyff, R.L., et al. “Candy Consumption Patterns, Effects on Health, and Behavioral Strategies to Promote Moderation: Summary Report of a Roundtable

Discussion." *Advances in Nutrition*, vol. 6, no. 1, Jan. 2015, pp. 139S-46S. https://doi.org/10.3945/an.114.007302.
2. Porter, G.P. and Grills, N.J. "The Dark Side to Halloween: Marketing Unhealthy Products to Our Children." *The Medical Journal of Australia*, vol. 199, no. 8, Oct. 2013, pp. 528-29. https://doi.org/10.5694/mja13.10722.
3. Cortés, D.E., et al. "Factors Affecting Children's Oral Health: Perceptions Among Latino Parents." *Journal of Public Health Dentistry*, vol. 72, no. 1, Oct. 2012, pp. 82-9. https://doi.org/10.1111/j.1752-7325.2011.00287.x.
4. U.S. Department of Agriculture and U.S. Department of Health and Human Services. Dietary Guidelines for Americans, 2020-2025. 9th Edition. 2020.
5. Reedy, J. and Krebs-Smith, S.M. "Dietary Sources of Energy, Solid Fats, and Added Sugars Among Children and Adolescents in the United States." *Journal of the American Dietetic Association*, vol. 110, no. 10, Oct. 2010, pp. 1477-84. https://doi.org/10.1016/j.jada.2010.07.010.
6. Warshaw, H. and Edelman, S.V. "Practical Strategies to Help Reduce Added Sugars Consumption to Support Glycemic and Weight Management Goals." *Clinical Diabetes*, vol. 39, no. 1, Jan. 2021, pp. 45-56. https://doi.org/10.2337/cd20-0034.
7. Barnett, I. and Edwards, D. "Mobile Phones for Real-Time Nutrition Surveillance: Approaches, Challenges and Opportunities for Data Presentation And Dissemination." *IDS Evidence Report* 75, Brighton: IDS. opendocs.ids.ac.uk/opendocs/handle/20.500.12413/4020
8. Ferrara, G., et al. "A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates." *JMIR Mhealth Uhealth*, vol. 7, no. 5, May. 2019, pp. e9232. https://doi.org/10.2196/mhealth.9232.
9. Zecevic, M., et al. "User perspectives of diet-tracking apps: reviews content analysis and topic modeling." *Journal of Medical Internet Research*, vol. 23, no. 4, Apr. 2021, pp. e25160. https://doi.org/10.2196/25160.
10. West, J.H., et al. "Controlling your 'app'etite: how diet and nutrition-related mobile apps lead to behavior change." *JMIR Mhealth Uhealth*, vol. 5, no. 7, Jul. 2017, pp. e95. https://doi.org/10.2196/mhealth.7410.
11. Limketkai, B.N., et al. "The age of artificial intelligence: use of digital technology in clinical nutrition." *Current Surgery Reports*, vol. 9, no. 7, Jun. 2021, pp. 20.
12. Abiodun, O.I., et al. "State-of-the-art in artificial neural network applications: A survey." *Heliyon*, vol. 4, no. 11, Nov. 2018, pp. e00938. https://doi.org/10.1016/j.heliyon.2018.e00938.
13. LeCun, Y., et al. "Deep learning." *Nature*, vol. 521, no. 7553, 2015, pp. 436-44. https://doi.org/10.1038/nature14539.
14. Cote, M. and Lamarche, B. "Artificial intelligence in nutrition research: perspectives on current and future applications." *Applied Physiology Nutrition and Metabolism*, vol. 47, Sep. 2021, pp. 1-8. https://doi.org/10.1139/apnm-2021-0448.
15. Gemming, L. "Image-assisted dietary assessment: a systematic review of the evidence." *Journal of the Academy of Nutrition and Dietetics*, vol. 115, no. 1, Jan. 2015, pp. 64-77. https://doi.org/10.1016/j.jand.2014.09.015.
16. François, C. Deep Learning with Python, 2nd Ed. Manning Publications, 2021.
17. Oltean, M.H. Fruits 360 A dataset with 90380 images of 131 fruits and vegetables (Version 9). 23 September 2022, www.kaggle.com/moltean/fruits.
18. Becker, D. Food 101 Pictures of 101 types of food (Version 1). 23 September 2022, www.kaggle.com/dansbecker/food-101.
19. Karthi, M., et al. "Evolution of YOLO-V5 algorithm for object detection: automated detection of library books and performance validation of dataset." 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2021, pp. 1-6, https://doi.org/10.1109/ICSES52305.2021.9633834.
20. Anthony, L.F.W. "Carbontracker: Tracking and predicting the carbon footprint of training deep learning models." *arXiv Preprint*: 2007.03051, 2020. https://arxiv.org/abs/2007.03051
21. Ferranti, E.P., et al. "Psychosocial factors associated with diet quality in a working adult population." *Research in Nursing & Health*, vol. 36, no. 3, Jun. 2013, pp. 242-56. https://doi.org/10.1002/nur.21532.
22. Grossniklaus, D.A., et al. "Psychological factors are important correlates of dietary pattern in overweight adults." *Journal of Cardiovascular Nursing*, vol. 25, no. 6, Nov. 2010, pp. 450-60. https://doi.org/10.1097/JCN.0b013e3181d25433.
23. Baranowski, T., et al. "Psychosocial correlates of dietary intake: Advancing dietary intervention." *Annual Review of Nutrition*, vol. 19, Jul. 1999, pp. 17-40. https://doi.org/10.1146/annurev.nutr.19.1.17.
24. McClain A.D., et al. "Psychosocial correlates of eating behavior in children and adolescents: a review." *International Journal of Behavioral Nutrition and Physical Activity*, vol. 6, Aug. 2009, pp. 54. https://doi.org/10.1186/1479-5868-6-54.
25. Veloso, S.M., et al. "Psychosocial factors of different health behavior patterns in adolescents: association with overweight and weight control behaviors." *Journal of Obesity*, vol. 2012, Jul. 2012, pp. 852672. https://doi.org/10.1155/2012/852672.
26. West, J.H., et al. "Health behavior theories in diet apps." *Journal of Consumer Health on the Internet*, vol. 17, no. 1, Feb. 2013, pp. 10-24. https://doi.org/10.1080/15398285.2013.756343.
27. Bohr, A. and Memarzadeh, K. "The rise of artificial intelligence in healthcare applications." *Artificial Intelligence in Healthcare*, 2020, pp. 25-60. https://doi.org/10.1016/B978-0-12-818438-7.00002-2.
28. Garcia-Ceja, E., et al. "Mental health monitoring with multimodal sensing and machine learning: A survey." *Pervasive and Mobile Computing*, vol. 51, Dec. 2018, pp. 1-26. https://doi.org/10.1016/j.pmcj.2018.09.003.
29. Dutta, A. and Zisserman, A. "The VIA Annotation Software for Images, Audio and Video." Proceedings of the 27th ACM International Conference on Multimedia (MM '19), ACM, 2019, pp. 1-6. https://doi.org/10.1145/3343031.3350535.
30. Pan, S.J. and Yang, Q. "A survey on transfer learning." *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, Oct. 2010, pp. 1345-59. https://doi.org/10.1109/TKDE.2009.191.
31. Smith, L.N. "Cyclical learning rates for training neural

networks." 2015. Available at: https://arxiv.org/pdf/1506.01186.pdf.