2022 Iraqi International Conference on Communication & Information Technologies (IICCIT-2022) , Basrah University , Basrah , Iraq

Deep Transfer Learning and Data Augmentation for Food Image Classification

Doaa AL-RUBAYE
Dept. of Computer Engineering
Yildiz Technical University
Istanbul, Turkey
f0120095@std.yildiz.edu.tr

Serkan AYVAZ
Dept. of Computer Engineering
Yildiz Technical University
Istanbul, Turkey
sayvaz@yildiz.edu.tr

2022 Iraqi International Conference on Communication and Information Technologies (IICCIT) | 978-1-6654-7220-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/IICCIT55816.2022.10010432

***Abstract*— The problem of food image classification has become a prominent topic that attracts many researchers due to its multiple benefits and applications in various aspects of life, from health to marketing. Image classification applications rely heavily on recent advancements in computer vision-based object recognition. In this paper, several deep transfer learning methods were investigated for food image classification. Furthermore, we applied a data augmentation approach to expand the Food-101 dataset. The impact of applying data augmentation and transfer learning was evaluated using five different deep learning models including Mobilenet, EfficientNetB1, and ResNet. It was noted that the EfficientNetB1 classifier achieved the best results with a score of 96.13%. In addition, we found that our data augmentation process was able to improve model performance.**

Keywords—Deep Learning, Food-image, Data Augmentation, Convolutional Neural Network, Transfer learning, Mobilenet, ResNet50, Resnet50V, ResNet101, EfficientNetB1

I. INTRODUCTION

Obesity is a major cause of many chronic diseases, such as diabetes, heart disease, high blood pressure, and some types of cancer. Additionally, 39% of people age 18 or older were overweight, and 13% were reported obese in 2016 [1]. It is indicating that the problem of being overweight is imminent and widespread around the world.

The detection and classification of food images are essential for health applications. By knowing what kind of food, a particular person eats, it is possible to predict their lifestyle and eating habits, thus helping to warn him/her of potential consequences of this behavior in the future.

In addition, this subject can help the blind or visually impaired. By a specific application, they can learn about the types of food served, for example, in open buffet restaurants, just as important for travelers who enjoy discovering new places and foods, knowing the ingredients and names of dishes can help avoid potentially serious allergic reactions or food poisoning issues.

Moreover, food safety is another important aspect, as reported by [2] about 690 million people globally are undernourished. The stunting rate was 21.3% in 2019.

From the point of view of marketing and economics, the classification of food images has a great influence on the choice of right method of e-marketing. For instance, if a restaurant owner wants to advertise a particular food, it is important that this advertisement appears to those who are inclined towards that type of food.

On the other hand, it is important to know the types of foods and foods that are most consumed to make long-term plans regulating consumption processes according to the quantities grown.

The great development brought about by artificial intelligence has led to a huge leap in the world of technology and in various aspects of life, including the problem of classifying food images.

This paper aimed to explore the use of deep learning techniques for the food classification problem. Furthermore, the performance implications of data augmentation and transfer learning were evaluated in detail. For evaluations, the two datasets, Food-101, and Enhanced Food101 were utilized for training and testing of five deep learning models including MobileNet, Resnet50, Resnet50V2, Resnet101, and EfficientNetB1.

This paper used transfer learning which is characterized by saving resources by using a small number of samples to train the model (if our problem is similar to the one on which the model was trained) or reducing the time required to train (if not the same), since some general features of the images are similar, so models can take advantage of the initial value of weights.

The remainder of the paper is structured as follows. The review of relevant work is covered in section 2. Section 3 describes our methods and the datasets that we dealt with. The evaluation of the results obtained is provided in section 4. Finally, section 5 contains the conclusion and future work.

II. RELATED WORK

The recent advancements in deep learning technologies have begun to transform many visual data processing tasks including image recognition [3], autonomous vehicles [4], robotics[5], disease diagnosis [6], entertainment [7], security surveillance [8]. Image classification is a major problem in computer vision. More particularly, the problem of food image classification and recognition has drawn a lot of attention recently, because of its great importance in its direct relationship with human life in many aspects, such as the type of nutrition [9], food safety [8], health issues [10], etc.

The field of image classification has gained significant benefits from developments in deep learning algorithms. However, deep learning algorithms are heavily dependent on the size and quality of the datasets used in the typical training process. Due to differences in the available datasets, applications and studies of food image classification have

reported widely fluctuating results. Below is a brief description of some of the available datasets that are frequently used to train food classification models:

A. *Food-5K*

This dataset provided by Singla, Yuan, and Ebrahimi [11], consists of 2500 food images with selected images from Food-100, UEC-Food-101, and UEC-Food-256, which are image sets used for food recognition. In 2019, Sengur et al. [12] introduced a method in which food images were given pre-trained CNN models (AlexNet and VGG16). The final feature vector was generated by combining all the collected feature vectors. In the last step, an SVM classifier was employed to determine the class label of each image, 99% accuracy was achieved when using AlexNet fc6 and VGG16 fc6 with 8192 features. On the other hand, 97.90% was achieved when using AlexNet fc6, AlexNet fc7 with the same number of features.

In 2016, an accurate GoogLeNet model was published, and the results were successful, with an accuracy of 99.2% [11] while binary classification showed an accuracy of 98.7% in the Food-5K database [13]. On the other hand, in [14] the validation and evaluation datasets each reached 99.4% and 98%, respectively, using the radial basis function (RBF) with kernel-based SVM and ResNet-152.

B. *UEC-FOOD- 256*

Another related dataset was created by Yoshiyuki Kawano and Keiji Yanai [15] covering 2,500 images of the food samples and the same number of non-food images with a total of 5,000 images. The deep CNN Inception-V3 model was presented by Hassannejad et al. as a pre-trained model in 2016 [3]. In addition to the other datasets, they used UEC FOOD 256. For FOOD- 256, they scored 88.28 and 96.88 percent for the Top-1 and Top-5 accuracy, respectively, using a deep neural network (DNN) with two principal branches, the residual and slice networks [3]. Another food identification study was presented in 2018. The datasets used were UECFood100, UECFood256, and Food-101. The WISer design was able to outperform other existing architectures in terms of performance[16].

C. *FoodX-251*

This dataset was provided by Parneet Kaur, Parneet Kaur, Parneet Kaur, Serge Belongie, and Ajay Divakaran [17] and its content was from 251 micro-classes and divided into 118 thousand for training, 12 thousand for verification, and 28 thousand for image testing. The same paper proposed a transfer learning method based on ResNet-101 with fine-tuning only the last layer. They obtained Top-3 Error of %37. Then, they used the same model but with fine-tuning all the layers and received 17% as a Top-3 Error.

D. *UEC-FOOD-100*

Presented by Yuji Matsuda, Hajime Hwashi, and Keiji Yanai [18], the dataset contains a total of 100 classes and 14,361 sample images. The difference here is that each shot has a bounding box with a tag indicating where the food is in the image. In contrast to [16], which uses the WISeR architecture to merge features, collected from two network branches to propose a food identification approach utilizing DCNN architecture, a pre-trained model named ResNet50 has been tuned, yielding an accuracy rate of 39.75% [19]. The remaining learning branches create a deep hierarchy capable of capturing the culinary attributes of the most food groups. Food plates with vertical layers were captured by the slice convolution branch, and for the food100 dataset they obtained 41.72% Top-1 accuracy in slice@WISeR and 86.71% Top-1 in residual@WISeR. In addition to that, H. Hassannejad, et al. [3] proposed a model based on Inception V3 and they achieved 81.45% as Top-1 accuracy in 2016.

Finally, several different studies were published for the Food101 dataset. In 2021 [20], Prakhar Tripathi suggested a method in transfer learning technologies based on DenseNet-161 and Top-1 accuracy was reported as 93.27%. hereas in 2020, the MobileNet architecture was used as a pre-trained model and three layers (average pooling, fully connected, and SoftMax) were removed from the original network [21]. Second, global average pooling, batch normalization, and SoftMax layers were added, resulting in an accuracy of 72.59%. Yet in 2022 [22], the EfficientNetB0 technology was introduced to utilize the distinct visual components and achieved 80% accuracy. Additionally, in the study [3], the authors scored 88.28% as a Top-1 accuracy. Earlier in 2014, Top-1 accuracy of 50.76% was received with an approach using a random forest-based method for mining discriminatory visual components and classification [23].

III. METHODOLOGY

A. *Data Collection*

In this study, FOOD-101 dataset was utilized as the basis for training and testing models in food image detection. The dataset contains 101,000 samples divided into 101 categories, each category with 1,000 samples (1000 images per category) characterized as real images of meals.

This dataset is important because it contains different types of food from different countries and cultures. On the other hand, some classes are similar to each other e.g., 'chocolate_cake' is very similar to 'chocolate_mousse'. Some of the other classes include sampling displays of food served in different ways. It is very difficult to find similarities between these samples, leading to a new challenge of the large disparity between models within the same class (as in apple_pie or clam_chowder).

Fig. 1: Food-101 images samples.

Fig. 2: Augmented Food-101 images samples

Another challenge is that in most cases, the real-world images contain not only the plate of the meal but other objects that can be considered as noise. Fig. 1 shows a sample of FOOD-101 images.

B. *Data Augmentation*

Although the total number of images in Food-101 dataset is high, there is a need for more image sampling per class as the number of samples per class is not sufficient for training and testing. Therefore, we applied a data augmentation approach to increase the image samples per class by adding new images. While applying synthetic data augmentation techniques such as vertical-to-horizontal flipping or rotating images, can increase the number of images in the dataset, the syntactic data augmentation tend to add bias to models. Thus, we decided to expand the dataset by adding new image samples in each class to better cover variations in each class so that it can improve the food image detection capabilities of the neural networks.

The Google search engine was used to add images. Adding images manually was slow and tedious process because adding each image was done very carefully. The reasons behind this are that the images on the Internet are varied and many of them are repetitive. Most of them are invalid because they contain noise such as writings and watermarks, which forced us to cut out parts of some images. In many cases, the same images were repeated more than once and under different labels, and sometimes the images were too small. In addition, some of the images included the cooking stage (that is, they are not the image of the finished dish). These reasons made the process of collecting images from the internet challenging and thus required a lot of time, and attention.

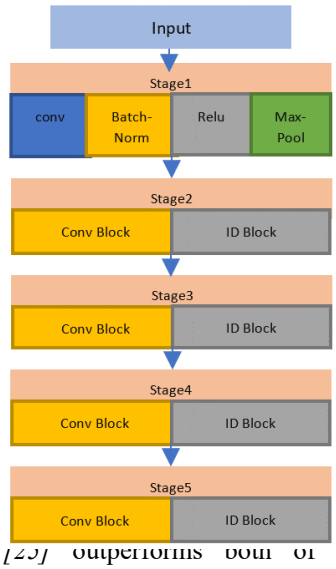
It should be noted that some classes have insufficient images, for example the “Beef-tartare” class. Despite the aforementioned challenges, 400 images have been added for each class, bringing the total number of each class to 1400 images. It means that at this stage a total of 4400 images have been added to the dataset. Fig. 2 demonstrates some sample images that have been added to augmented dataset.

C. *Technologies used*

Several techniques were used in this paper to classify 101 different types of food from different countries. The entire work was implemented using Keras libraries. A comparison was made between the results obtained from the original data and when using the augmented data strategy. In both cases, dataset without augmentation and the augmented case, the dataset was split into 80% training data (80,800 samples for the first case and 112787 samples for the second case) and 20% test data consisting of 20,200 images (the same test dataset was used in each case). Additionally, in both cases, automatic data augmentation strategy was also applied proportionally. More specifically, synthetic data was also generated from existing images to increase the total amount of data by adding slightly modified copies of existing images. We added four augmented filters as layers to our models, which quadrupled our data. It is worth noting that data augmentation technique was used to prevent overfitting, which acts as regularization and helps train models.

Furthermore, we used epochs 50, and 16 as batch sizes in all models because when trying with larger batch sizes, e.g., 32, unfit results were observed. In addition, all deep learning models used were based on transfer learning. However, we needed to add a Normalization layer. The last step in each model was to add the prediction layer. In compiling step for each model, the Adam optimizer was used with a learning rate of 0.0001. Below is a brief explanation of the architectures of models used:

a) ResNet50: A convolutional neural network with 50 layers, the ResNet-50 was developed by Kaiming He et al. in 2016 [24] and is based on the Residual Block. It includes a total of five stages. Each stage has a convolution and an Identity block. In addition, there are three convolution layers in each convolution block and three convolution layers in each identity block. Fig. 3 shows the architecture of ResNet50, which has roughly 23 million trainable parameters.



*b) Resn*roved version of ResNet50, 1 ResNet101, ResNet50V2 [25] outperforms both of them on the ImageNet dataset. The propagation formulation of the links between blocks was altered in ResNet50V2.

c) Resnet101: The ResNet-101 convolutional neural network has 101 layers of depth as shown in Fig. 4. This

network is pre-trained on the ImageNet database *[26]*, which contains more than one million images. The network was taught to recognize more than 1,000 types of items, including a mouse, keyboard, and a variety of animals. Thus, the network collected a variety of feature representations. The network accepts images with a resolution of 224 * 224 pixels.

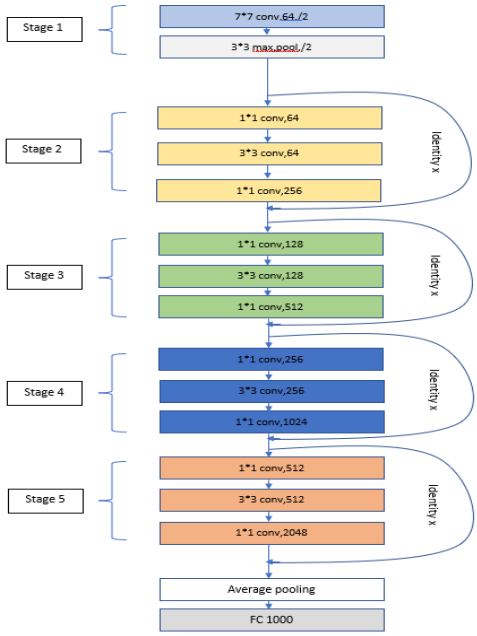


Fig. 4: Resnet101 Architecture

d) MobileNet: The mobility of this convolutional neural network makes it perfect for mobile and embedded vision applications. For mobile and embedded devices, it is based on a reduced architecture that uses deep-separable convolutions to produce lightweight, low-latency deep neural networks. Fig. 5 illustrates the architecture of MobileNet.

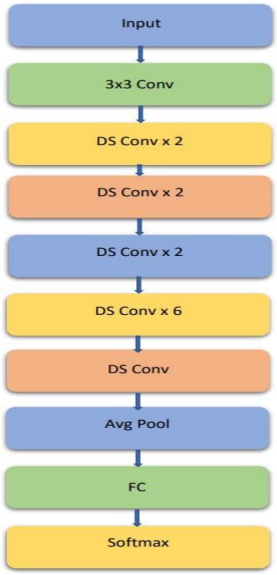
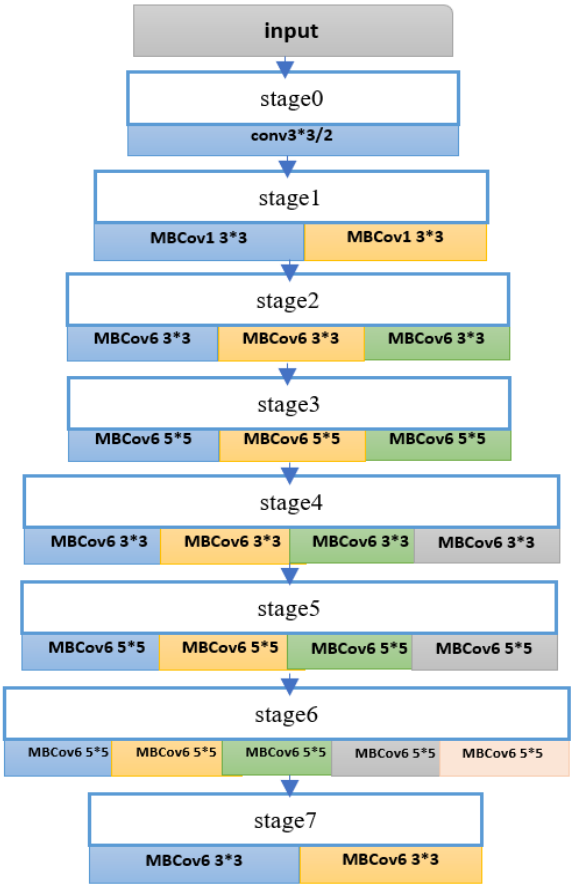


Fig.5: MobilNet Architecture

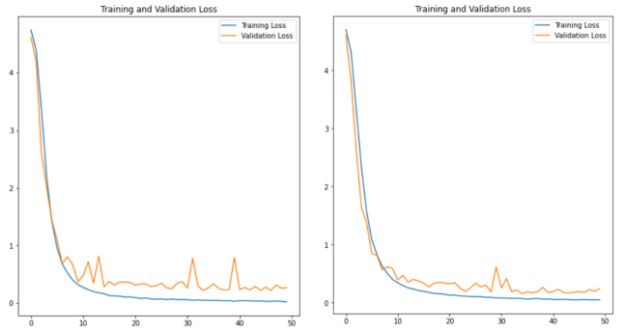
e) EfficientNetB1: is a method for designing and scaling convolutional neural networks that makes use of a compound coefficient to scale all dimensions of depth, breadth, and resolution in an equal manner. In contrast to normal practice, the EfficientNet scaling technique employs a set of predetermined scaling coefficients to uniformly scale network width, depth, and resolution. EfficientNet-B1 requires the employment of decoder blocks in the upsampling network. Each decoder block is made up of two concatenated feature maps from the encoder section, which are upsampled two times by a stride of two. Finally, the decoder passes the concatenated tensor through two convolution layers with ReLU activation and batch normalizes it. Softmax convolution is the last layer, and its channel number and output picture size are both matched to



thc the arc

IV. RESULTSTO EVALUATE THE IMPACT OF OUR DATA AUGMENTATION APPROACH, WE COMPARED THE PERFORMANCES OF FIVE PRE-TRAINED NETWORKS, NAMELY RESNET50, RESNET50V2, RESNET101,

MobilNet, and EfficientNetB1 on both the food-101 dataset and the augmented food-101 dataset. The models were trained by using Google Colaboratory and Kaggle framework. The results obtained from these networks are compared based on overall accuracy as shown in Table I .

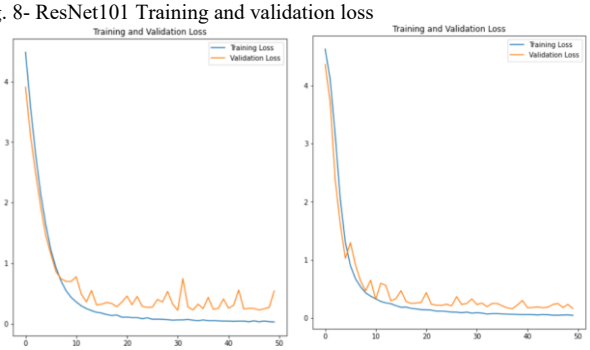


a) On Food-101 b) On augmented data

Fig. 7- MobilNet Training and validation loss

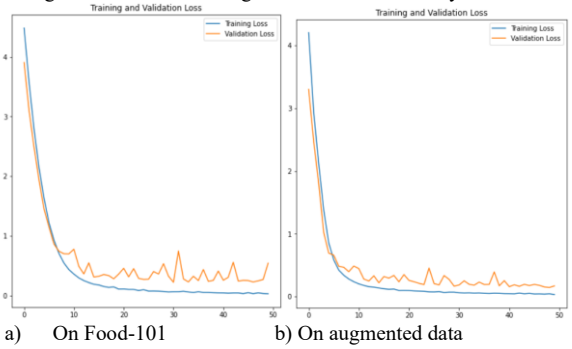
a) On Food-101 b) On augmented data

Fig. 8- ResNet101 Training and validation loss



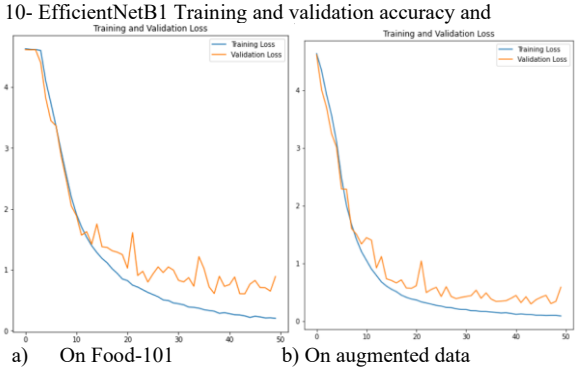
a) On Food-101 b) On augmented data

Fig. 9- ResNet50 Training and validation accuracy and loss



a) On Food-101 b) On augmented data

Fig. 10- EfficientNetB1 Training and validation accuracy and



a) On Food-101 b) On augmented data

Fig. 11- Resnet50V2 Training and validation accuracy and loss

A. *Evaluation Results on FOOD-101*

According to the original FOOD-101 dataset, the best accuracy result was achieved in Mobilenet with 95% accuracy, while the lowest result was obtained when applying ResNet101 with 76%. Moreover, we obtained 94% accuracy using Resnet50, 89% using EfficientNetB1 and 80% with ResNetV2. Finally, the part (a) in Fig.s 7, 8, 9, 10, and 11 show the train, test, and loss accuracy respectively.

B. *Evaluation Results on Augmented FOOD-101*

In the evaluations using the augmented dataset, we received better results. The best accuracy was **96.13**% when using EfficientNetB1, while the lowest was Resnst101 with **79.77**%. In addition, **94.25**% was the accuracy in Mobilnet, and 85.97% in Resnet50V2, and 95.84% was observed when using ResNet50. The part (b) in Figs. 7, 8, 9, 10, and 11, the train and test accuracy and losses are shown.

V. DISCUSSION

Although transfer learning requires a small number of data for training, our approach of applying transfer learning with the data augmentation outperformed the baseline approach. Increasing the number images in each food class by adding new image samples has paid off despite the use of transfer learning.

The syntactic data generation approaches are prone to produce bias in food image classification. The transfer learning models lack the variations observed in the food images in the real world. Thus, for the problem of classifying food images, adding new images covering variations in each class appear to be improving the model performances. Moreover, the foods are usually presented in different ways. Therefore, adding new novel samples can introduce image features about new food presentation methods and consequently help the models learn important characteristics.

An interesting finding of the study that we noticed from the results that the food image classification task using the current food classification dataset does not need a very deep model. It appears that the models with larger number of the filters demonstrated better results i.e. the results obtained by using Resent101 performed worse than the Resnet50.

VI. CONCLUSION

This work introduced several techniques based on transfer learning by fine-tuning the pre-training deep learning network to classify food according to the Food-101 dataset. In addition, the Food-101 dataset was augmented by adding more carefully selected samples to expand the samples per class in the dataset and improving the existing samples, as many of them include distorted images such as containing the remains of a dish that is not clearly defined or the image

is crowded with objects. The results obtained indicated that this dataset does not need very complex networks. Overall results with Resnet50 and Mobilenet were better than when using Resnet-101, which is considered more in-depth than Resnet50. It was also noted that the augmented dataset produced better results as the increase in the number of samples led to improvement in network training, and predictions in some groups, such as "ravioli" in the Mobiles network. It was marked by the lowest rate of 75%, but after improving the data, the percentage went up to 90%. In other words, data augmentation led to better coverage of some classes that were poorly represented in the dataset.

For future work, we intend to embed this work into a dedicated phone app that can distinguish food with the phone's camera, helping the blind and those traveling in buffet restaurants.

TABLE I. THE MODEL COMPARISON RESULTS.

RESULTS OF THE ORIGINAL FOOD-101 DATASET		RESULTS OF THE AUGMENTED DATASET	
THE NETWORKS	THE ACCURACY %	THE NETWORKS	THE ACCURACY %
MOBILENET *[1]	72.59	-	-
EFFICIENTNETB0*[2]	80	-	-
RESNET50	**94**	**RESNET50**	**95.84**
RESNET50V2	**80**	**RESNET50V2**	**85.97**
RESNET101	**76**	**RESNET101**	**79.77**
MOBILENET	**95**	**MOBILENET**	**94.25**
EFFICIENTNETB1	**89**	**EFFICIENTNETB1**	**96.13**

*[1]:The model accuracy in [21]
*[2]: The model accuracy in [22]

As shown in the Table I, we found that our approach of developing transfer learning models using data augmentation achieved better evaluation results in predicting the food images when compared to the results of the state-of-art in the field.

REFERENCES

[1] "World Health Organization." https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight#:~:text=In%202016%2C%20more%20than%201.9,kills%20more%20people%20than%20underweight.

[2] "Obesity and overweight(World hunger: Key facts and statistics 2022)," *Action against hunger*, Apr. 14, 2022. https://www.actionagainsthunger.org/world-hunger-facts-statistics

[3] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food Image Recognition Using Very Deep Convolutional Networks," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, Amsterdam The Netherlands, Oct. 2016, pp. 41–49. doi: 10.1145/2986035.2986042.

[4] M. Bojarski *et al.*, "VisualBackProp: efficient visualization of CNNs." arXiv, May 19, 2017. Accessed: Jul. 04, 2022. [Online]. Available: http://arxiv.org/abs/1611.05418

[5] R. A. Mouha, "Deep Learning for Robotics," *J. Data Anal. Inf. Process.*, vol. 09, no. 02, pp. 63–76, 2021, doi: 10.4236/jdaip.2021.92005.

[6] A. M. Hafiz and G. M. Bhat, "A Survey of Deep Learning Techniques for Medical Diagnosis," in *Information and Communication Technology for Sustainable Development*, vol. 933, M. Tuba, S. Akashe, and A. Joshi, Eds. Singapore: Springer Singapore, 2020, pp. 161–170. doi: 10.1007/978-981-13-7166-0_16.

[7] J. Thomas, L. Comoretto, J. Jin, J. Dauwels, S. S. Cash, and M. B. Westover, "EEG CLassification Via Convolutional Neural Network-Based Interictal Epileptiform Event Detection," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, Jul. 2018, pp. 3148–3151. doi: 10.1109/EMBC.2018.8512930.

[8] H. Deléglise, R. Interdonato, A. Bégué, E. Maître d'Hôtel, M. Teisseire, and M. Roche, "Food security prediction from heterogeneous data combining machine and deep learning methods," *Expert Syst. Appl.*, vol. 190, p. 116189, Mar. 2022, doi: 10.1016/j.eswa.2021.116189.

[11] A. Singla, L. Yuan, and T. Ebrahimi, "Food/Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, Amsterdam The Netherlands, Oct. 2016, pp. 3–11. doi: 10.1145/2986035.2986039.

[12] A. Sengur, Y. Akbulut, and U. Budak, "Food Image Classification with Deep Features," in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Turkey, Sep. 2019, pp. 1–6. doi: 10.1109/IDAP.2019.8875946.

[13] W. Jia *et al.*, "Automatic food detection in egocentric images using artificial intelligence technology," *Public Health Nutr.*, pp. 1–12, Mar. 2018, doi: 10.1017/S1368980018000538.

[14] P. McAllister, H. Zheng, R. Bond, and A. Moorhead, "Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets," *Comput. Biol. Med.*, vol. 95, pp. 217–233, Apr. 2018, doi: 10.1016/j.compbiomed.2018.02.008.

[15] Y. Kawano and K. Yanai, "Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation," in *Computer Vision - ECCV 2014 Workshops*, vol. 8927, L. Agapito, M. M. Bronstein, and C. Rother, Eds. Cham: Springer International Publishing, 2015, pp. 3–17. doi: 10.1007/978-3-319-16199-0_1.

[16] N. Martinel, G. L. Foresti, and C. Micheloni, "Wide-Slice Residual Networks for Food Recognition," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, Mar. 2018, pp. 567–576. doi: 10.1109/WACV.2018.00068.

[17] P. Kaur, K. Sikka, W. Wang, S. Belongie, and A. Divakaran, "FoodX-251: A Dataset for Fine-grained Food Classification," 2019, doi: 10.48550/ARXIV.1907.06167.

[18] Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of Multiple-Food Images by Detecting Candidate Regions," in *2012 IEEE International Conference on Multimedia and Expo*, Melbourne, Australia, Jul. 2012, pp. 25–30. doi: 10.1109/ICME.2012.157.

[19] Z. Zahisham, C. P. Lee, and K. M. Lim, "Food Recognition with ResNet-50," in *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia, Sep. 2020, pp. 1–5. doi: 10.1109/IICAIET49801.2020.9257825.

[20] Prakhar Tripathi, "TRANSFER LEARNING ON DEEP NEURAL NETWORK: A CASE STUDY ON FOOD-101 FOOD CLASSIFIER," *IJEAST*, vol. 5, pp. 229–232, 2021.

[21] S. Phiphiphatphaisit and O. Surinta, "Food Image Classification with Improved MobileNet Architecture and Data Augmentation," in *Proceedings of the 2020 The 3rd International Conference on Information Science and System*, Cambridge United Kingdom, Mar. 2020, pp. 51–56. doi: 10.1145/3388176.3388179.

[22] V. G., P. Vutkur, and V. P., "Food classification using transfer learning technique," *Glob. Transit. Proc.*, vol. 3, no. 1, pp. 225–229, Jun. 2022, doi: 10.1016/j.gltp.2022.03.027.

[23] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101 – Mining Discriminative Components with Random Forests," in *Computer Vision – ECCV 2014*, vol. 8694, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 446–461. doi: 10.1007/978-3-319-10599-4_29.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," 2016, doi: 10.48550/ARXIV.1603.05027.

[26] Rebecca Reynoso, "A Complete History of Artificial Intelligence," *G2*, May 25, 2021. https://www.g2.com/articles/history-of-artificial-intelligence