Global Transitions Proceedings 3 (2022) 225–229

Contents lists available at ScienceDirect

Global Transitions Proceedings

journal homepage: http://www.keaipublishing.com/en/journals/global-transitions-proceedings/

Food classification using transfer learning technique

VijayaKumari G. [a], Priyanka Vutkur [a,*], Vishwanath P. [b]

[a] Department of CSE, H. K. E. Society's S. L. N. College of Engineering, Raichur, 584135, India
[b] Department of E& CE, H. K. E. Society's S. L. N. College of Engineering, Raichur, 584135, India

ARTICLE INFO

Keywords:
Food-101
Computer vision
Transfer learning
Feature extraction
EfficientNetB0

ABSTRACT

In the subject of object detection using computer vision, image classification is becoming a prominent and promising aspect. However, studies have just scratched the surface. Till now, the superficials of food image classification in order to assess the nutritional abilities of people of different nationalities, The categorization of their traditional cuisine has a significant influence. Existing models categorize different sorts of foods. These models can only categorize a small number of meals at a given time. However, in a single model, the maximum number of foods must be recognized. This work focuses on the creation of a recognition model that uses transfer learning techniques to categorize various food products into their appropriate categories. Using Efficientnetb0, a transfer learning technique, the developed model classified 101 distinct food kinds with an accuracy of 80%. When compared to other state of art models, our model performed with best accuracy.

1. Introduction

People all across the world are becoming more health aware as the world becomes more competitive and dynamic. Overweight is becoming a worldwide concern at an alarming rate. India ranks #1 among countries with the largest vegetarian population, with 40% of Asian Indians being vegetarian [1]. It appears that vegans in India are going through a "nutrition transition," with less eating of whole plant meals and more refined carbs, fried foods, and processed foods [2]. This study investigates the link between a vegetarian diet and weight loss.

With the rise of nutrition-related diseases, there is an increasing global awareness of the importance of eating a well-balanced and healthy diet. Obesity, diabetes, and cancer can all be prevented by eating a nutritious diet [3]. Although there are a variety of programmes available to detect and classify food, they all require prior knowledge to do so [4]. However, when a new and unfamiliar cuisine is presented, a difficulty occurs. Several studies on the classification of food images have previously been completed [5]. Food image classification is a relatively new sector in the coming applications of deep learning developments. Prior to the development of Deep Learning algorithms, several food categorization works employed the standard Machine Learning technique for classification [6,7].

Food-101 data is divided into several subsets. The goal is to use photographs that have been downscaled to allow for speedy testing. HDF5 has been used to reformat the images [8].

The proposed research is carried out with the help of the Python programming language and the Tensorflow package. The results were compared to other transfer learning systems when they were completed.

2. Related work

Lei Zhou [9], gave a basic introduction to deep learning and thorough descriptions of the construction of certain popular deep neural network architectures as well as training methods. hundreds of publications were reviewed that employed deep learning as a data analysis technique to handle difficulties and challenges in the food domain, such as food recognition, calorie calculation, quality detection of fruits, vegetables, meat, and aquatic goods, food supply chain, and food contamination [10]. Each research looked into the individual challenges, datasets, preprocessing methods, networks and frameworks used, performance achieved, and comparisons with other popular solutions. Deep learning's potential for application as an advanced data mining tool in food sensory and consumption studies was also investigated. Amatul Bushra Akhi [11], used to train an image category classifier, employed a pre-trained Convolutional Neural Network (CNN) as a feature extractor. A multiclass linear Support Vector Machine (SVM) classifier trained with extracted CNN features is used to classify fast food photos into ten different kinds. A multiclass linear Support Vector Machine (SVM) classifier trained with extracted CNN features is used to classify fast food photos into 10 different kinds and reached a success rate of 99.5% after working on two distinct benchmark databases, which is higher than the accuracy achieved using bag of features (BoF) and SURF. Chairi Kiourt [12], explained the three main lines of solutions, namely de-

* Corresponding author.
E-mail address: priyankavutkur6@gmail.com (P. Vutkur).

https://doi.org/10.1016/j.gltp.2022.03.027

Available online 2 April 2022

sign from scratch, transfer learning, and platform based methodologies, are presented, tested, and compared to show the inherent strengths and weaknesses, specifically for the task at hand. Basic background material, a part devoted to important datasets that are critical in light of the empirical methodologies used, and some final notes that highlight future directions complete out the chapter. Michele De Bonis [13], The paper presents an effective way for constructing a mobile application for food recognition using Convolutional Neural Networks (CNNs). The GoogLeNet, which has the best accuracy and a model size second only to the SqueezeNet with 40Mb of model size and approximately 70% accuracy, and the SqueezeNet, which has a decent degree of accuracy but an extremely limited model size with 3Mb of model size and about 60% accuracy, have been identified. Malina Jiang [14], built convolutional neural networks from scratch and using pre-trained weights learnt on a bigger picture dataset (transfer learning) to solve the issue of food image classification, reaching an accuracy of 61.4%. In Chang Liu [15], The parameters were adjusted to match 101 food categories, and then 300,000 iterations were utilized using a base learning rate of 0.01, a momentum of 0.9, and a base learning rate of 0.01. Fine-tuned the model on the Food-101 dataset using the 1000-class pre-trained model from the ImageNet dataset; the accuracy is displayed in the table below, and we reached a top-1 accuracy of 77.4%. Muhammad Zain Amin [16], employ the transfer learning technique to retrain the final layer of the famed Inception V3 architecture developed by Google for our classification strategy, which is based on the Tensorflow platform's Inception V3 model. For 16 class classification, data augmentation approaches based on geometric transformation were used to increase the amount of training images with an overall accuracy of 91% in accurately detecting food images while avoiding the overfitting problem. Kawano et al [17] created an Android application to gather and classify food images. They also established the "UEC-256 food picture data collection," which is a database of food images. They started with this data set and ran several trials with SIFT features and SVM, which yielded significantly better results than PFID [18]. Then they used AlexNet [19] to the identical data sets and found that it performed significantly better than the SIFT- SVM-based technique [20]. In Nursuriati Jamil [21], There are 50 different types of Malaysian cuisine. A pre-trained MobileNet learning model on the Tensorflow Lite deep learning environment is used to recognize food. The average accuracy of food recognition was about 80% at the time of writing. Dipti Pawade [22], have described the Convolutional Neural Network-based model, which requires the input of a preprocessed, scaled grayscale food image. The system then estimated nutrition factors such as carbohydrate, protein, fat, and calories based on the model's prediction of the dishes in the photograph. We looked at the accuracy, loss graph, and confusion matrix when evaluating the model. Our model has a training accuracy of 99.8%. For 10 classifications, the average accuracy, recall, and F1 score are 0.72, 0.67, and 0.68, respectively. The Google Inception-V3 model serves as a foundation, and a fully linked layer is created on top of it to improve the categorization process. In Vishwanath [9], dataset consists of thousands of photos from 16 different cuisine classes. During the testing phase, a classification accuracy of 96.27% was achieved. Seon-Joo Park [23], image dataset for use in building a complicated Korean food identification algorithm. In order to expand the dataset size, augmentation techniques were used. Diksha Solanki [24], When it comes to identifying and recognizing food images, they use a convolutional neural network (CNN). Image identification of food products is frequently difficult due to the huge range of foods available. Convolutional Neural Networks (CNNs) have been shown to be more robust and expressive than hand-crafted highlights. Asif Mahbub Uddin [25], Even with its high accuracy, processing power combined with efficiency, and automated detection of crucial elements without the need for human intervention, the CNN technique has been chosen. TBFI classification has also been done using a transfer learning strategy with fine-tuned VGG16. Deepak Rana [26], CNN was utilized because the convolution layers may be tweaked and are simple to install. Second, customize the model with GUI features and nutritional analyses [27]. In addition, they included typical Indian cuisine categories to the FOOD-101 dataset [28,29]. This study provides a classification system for 101 distinct food item classification.

3. Proposed system

CNN is a deep learning technique that uses a customized neural network with a number of layers to interpret pictures. Convolution layers are the main layers that execute filtering processes. A pre-trained model for classification and recognition is used in the Transfer Learning technique. Due to their greater performance, convolutional neural networks outperform classic neural networks for image, audio/video, and voice signal inputs. As shown in Fig. 1, it primarily comprises of three types of layers: convolutional, pooling, and fully connected layers. A convolutional network's first layer is the convolutional layer. More convolutional layers or pooling layers can be added after convolutional layers, but the fully-connected layer is the last layer. The CNN becomes more complicated with each layer, detecting more areas of the picture. The first layers concentrate on the most basic features, such as colors and borders. As the visual input flows through the CNN layers, it begins to recognize the item's bigger parts or characteristics, finally recognizing the target object.

3.1. Convolutional layer

The convolutional layer is the most important part of CNN because it is where the majority of the computation takes place. To create the feature map, the collection of pictures is fed into the convolutional layer. A grayscale image is represented as a 2D matrix of pixels, with the input in three dimensions: height and breadth. The feature detector, also known as a kernel or a filter, will examine the picture for the existence of features. The filter size can vary, although a 3×3 matrix is commonly employed. The final result of a set of Convolution procedures is a feature map, activation map, or convolved feature as shown in Fig 1.

After the convolution layer the nonlinear activation operation will be performed using Rectified Linear Unit (ReLU).

3.2. Feature map

Pooling layers are used to minimize the size of the feature map. As a result, the amount of processing in the network and the number of parameters to learn are both lowered. The pooling layer summarizes the characteristics discovered in a given section of the convolution layer's feature map.

3.3. FNN block

As a fundamental example of neural network design, the feedforward neural network has a constrained architecture. Signals are passed from one layer to the next. Some feed forward designs are even more straightforward. A single layer perceptron model, for example, has only one layer and a feedforward signal that moves from one layer to an individual node. Feedforward models with many layers are also known as multi-layer perceptron models.

3.4. Transfer learning

Transfer learning (Fig. 2) is a machine learning design technique that starts with a learned model and then trains it for a new problem domain. The pre-trained network is used to transport data from one domain to another. Other strategies, including as feature transfer and fine-tuning, as well as freezing and retraining certain network layers, are available. Transfer learning is used in many deep learning applications currently and in the future. This is mostly due to deep learning training systems' vast scale, which necessitates a substantial amount of resources.

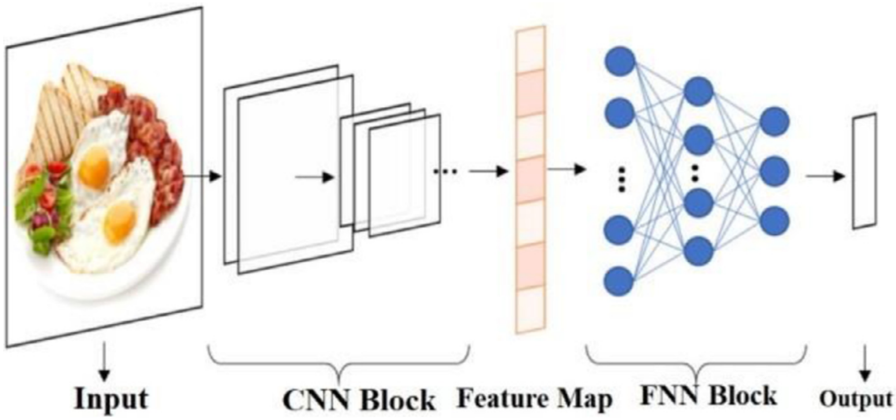

Fig. 1. CNN architecture.

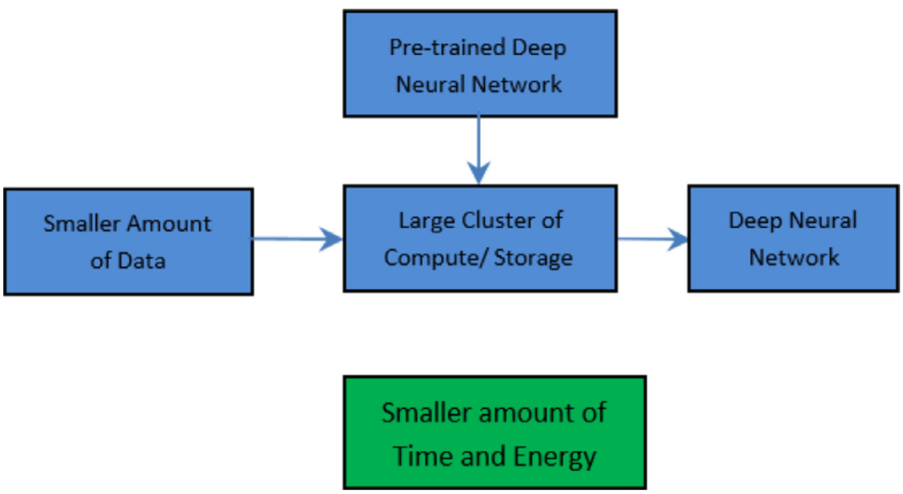

Fig. 2. Transfer learning with a pre-trained network.

For transfer learning, the generality of the learning model attributes is crucial.

The deep learning network's features are supplied into the categorization output layer. The trained network may be reused for a different challenge since these features can be reused. Another important aspect of transfer learning is training data. Transfer learning procedures include feature transfer, fine tweaking, and using a pre-trained model. Training data is another crucial part of transfer learning.

3.4.1. Feature transfer

One of the most basic techniques of transfer learning is feature transfer. Fig. 3 depicts a deep learning network with several levels (a). In this application, the network depicts a three dimensional picture. This is the input layer, which is in charge of mapping data from one layer to the next. The following layer is the feature extraction layer, which consists of a number of internal layers as well as a combination of convolutions and pooling. The feature extraction layer generates "features" that may be used to represent visual features before being translated to higher-level features in a hierarchical manner. The characteristics from this layer are combined in the final classification layer to perform the classification.

The classification layer is in charge of identifying an item in the picture based on the identified characteristics. Using the input and feature extraction layers that have been learnt with a specific data set, feature transfer is used to train a new classification layer for the relevant domain.

3.4.2. Fine –tuning

A new classification layer was introduced to the feature extraction by freezing the previous levels in deep learning (Fig. 3(b)). It's a straightforward technique to create a classification layer and then fine-tune the earlier layers using the fresh training data set through repeated training. This layer can be fine-tuned to fine- tune the layers that are more specific to the features of the classification challenge. This strategy is suitable when the issue domains are separated by a significant amount of distance and new qualities must be classified.

3.4.3. Using a pre-trained model

Feature transfer allows a model to be trained for one problem and then utilized for another. Another option is to retrain a model that has already been trained. Efficientnetb0, MobileNet, Object Detection, Sentiment Discovery, YOLO, Car Classification, Show and Tell, and Lip Reading are just a handful of the pre-trained models that may be used across platforms and activities. You can update a pre-trained model by transferring features or freezing parts of the early convolutional layers and retraining the later ones while utilising it. Because the early Convolutional layers disclose broad features that are independent of the problem, retraining the later convolutional layers, where features are more exact and dependent on the problem, may be useful. If the problem domains are the same, this paradigm is excellent. The EfficientNet-B0 architecture was created by the neural network itself, not by engineers. They used a multi objective neural architecture search to create this model, which maximizes both accuracy and floating-point operations. Using B0 as a starting point, the authors created a family of EfficientNets rang-

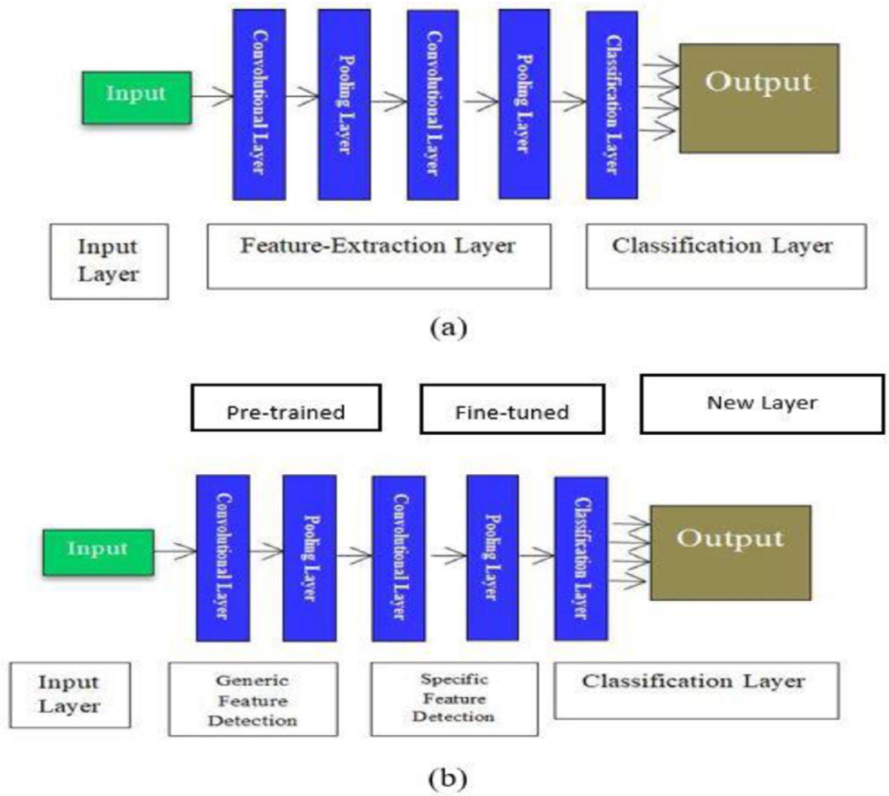

Fig. 3. Simple deep learning network illustrating the basic layers, Fine-tuning the specific feature extraction layers.

Table 1
Proposed methods.

layers	Model:1	Model:2	Model:3	Model:4
CNN	EfficientNetB0	EfficientNetB0	EfficientNetB0	EfficientNetB0
Input Size	224 × 224	224 × 224	224 × 224	224 × 224
Data augmentation	Flip ("horizontal"), (rotation, height, width, Zomm)=0.2	Flip ("horizontal"), (rotation, height, width, Zomm)=0.2	Flip ("horizontal"), (rotation, height, width, Zomm)=0.2	Flip ("horizontal"), (rotation, height, width, Zomm)=0.2
Mixed precision	-	-	Mixed_float16	Mixed_float16
Pooling	Global Average Pooling 2D()	Global Average Pooling 2D()	Global Average Pooling 2D()	Global Average Pooling 2D()
Output	Dense(activation="Softmax")	Dense(activation="Softmax")	Dense(activation="Softmax")	Dense(activation="Softmax")
Learning rate	$1e^{-2}$	$1e^{-4}$	$1e^{-2}$	$1e^{-4}$
Loss function	Categorical_crossentropy	Categorical_crossentropy	Categorical_crossentropy	Categorical_crossentropy
Accuracy	57.76%	60.78%	70.53%	80.16%

ing from B1 to B7 that achieved state of the art accuracy on ImageNet while being extremely efficient in comparison to its competitors. B0 is a mobile-sized architecture with 11M trainable parameters that employs 7 inverted residual blocks, each with its own set of settings. Squeeze and excitation blocks, as activation, are used in these blocks. In this essay, we'll go through all three in depth.

Table 2
Precision, recall, and F1 score values.

Models	precision	recall	f1-score
Model:1	59	58	58
Model:2	63	61	61
Model:3	74	71	71
Model:4	83	81	81

4. Experimental results

The data we are working with come from the Food-101 dataset, which contains 101,000 (1000 images per category) real-world images of food meals divided into 101 categories. The training data contains 75,750 images and test data contains 25,250 images. Using this data based on the following Table 1 model parameters experiment was performed for 100 epochs and achieved better results, Fig. 4 shows the comparison of accuracy of proposed methods.

The precision is calculated as tp / (tp + fp), with tp signifying true positives and fp denoting false positives. Recall is calculated using the ratio tp / (tp + fn), where tp represents true positives and fn represents false negatives. The support is the number of instances of each class. The unweighted mean of each label's metrics is calculated using macro average. To establish the average weight, a weighted average is applied to the metrics of each label, Table 2 shows the results of precision, recall, and f1-score for proposed methods.

When compared to the other three models, Model 4 fared the best, with an accuracy of 80% for various metrics. Table 3 shows the State of art of different authors.

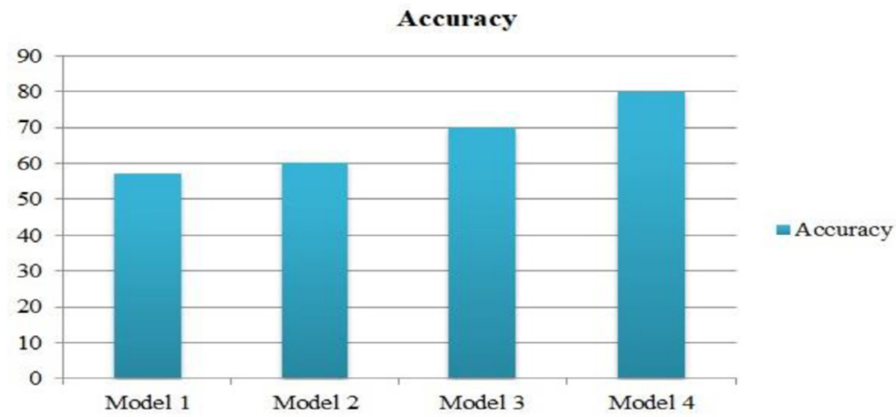

Fig: 4. Comparison of all four models.

Table 3
comparison of our method with State of art.

Sl. No	Author	Classification	Model	Accuracy(%)
1	Chang Liu [6]	101	Convolutional Neural Network (CNN)	77.00
2	Prakhar Tripathi [20]	101	Random Forest Method	50.78
3	Ignazio Gallo [21]	101	Inception V3	71.67
4	Thunchanok Tang pong [22]	101	GoogLeNet	80.00
5	Lukas Bossard [19]	101	Random Forests	53.00
6	Our Model	101	EfficientNetB0	80.00

4. Conclusion

Present a fresh large-scale benchmark dataset for food recognition in this research. Also introduced a unique approach for mining discriminative visual components and efficient classification based efficientNetB0. On the hard "Food-101" dataset, we demonstrated that it outperforms state of the art methods on food recognition, with the exception of CNN, and that it achieves 80% accurate results when compared to alternative recent part based classification approaches.

In the future, test this method with more multimodal datasets that include text and image, as well as different modalities such as audio.

Reference

[1] Amatul Bushra Akhi, Farzana Akter, Tania Khatun, Mohammad Shorif Uddin, Recognition and classification of fast food images, Global J. (2022) ISSN: 0975-4172.
[2] B.D. Parameshachari, R. Karappa, K.S. Soyjaudah, S.K. Devi, Partial image encryption algorithm using pixel position manipulation technique: The smart copyback system, in: 2014 4th international conference on artificial intelligence with applications in engineering and technology, IEEE, 2014, pp. 177–181.
[3] C. Kiourt, G. Pavlidis, S. Markantonatou, Deep learning approaches in food recognition, Machine Learning Paradigms - Advances in Theory and Applications of Deep Learning, Springer, 2020.
[4] T.N. Nguyen, N.P. Nguyen, C. Savaglio, Y. Zhang, B. Dumba, The role of artificial intelligence (AI) in healthcare data analytics, Int. J. Artif. Intell. Tools 30 (06 N 08) (2021).
[5] Michele De Bonis1, Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, and Paolo Manghi," Deep Learning Techniques for Visual Food Recognition on a Mobile App". 2022.
[6] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, T. Sato, Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health, IEEE Wirel. Commun. 28 (3) (2021) 54–61.
[7] Malina Jiang, Food Image Classification with Convolutional Neural Networks, CS230: Deep Learning, Fall (2019) Stanford University, CA.
[8] P. Subramani, P. BD, Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients, Pers. Ubiquitous Comput. (2021) 1–14.
[9] Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu, Yong He, Application of deep learning in food: a review, Comprehensive Rev. Food Sci. Food Safety 18 (2019).
[10] N.T. Le, J.W. Wang, D.H. Le, C.C. Wang, T.N. Nguyen, Fingerprint enhancement based on tensor of wavelet subbands for classification, IEEE Access 8 (2020) 6602–6615.
[11] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, Yunsheng Ma," DeepFood: deep learning-based food image recognition for computer-aided dietary assessment". 2022.
[12] Y. Sun, J. Liu, K. Yu, M. Alazab, K. Lin, PMRSS: privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare, IEEE Trans. Ind. Inf. 18 (3) (2021) 1981–1990.
[13] Nursuriati Jamil, Shahrul Raimie Roslan, Raseeda Hamzah, Izzad Ramli." Food recognition of Malaysian meals for the management of calorie intake, ISSN 2278-3091 (2022).
[14] Dipti Pawadea, Ashwini Dalvia, Irfan Siddavatamb Dr., Myron Carvalhoc, Prajwal Kotianc, Hima Georgec, Cuisine detection using the convolutional neural network, I.J. Educ. Manag. Eng. (2022).
[15] Vishwanath C. Burkapalli1, Priyadarshini C. Patil2, Transfer learning: inception-v3 based custom classification approach for food images, ISSN: 0976-9102 (2022).
[16] Muhammad Zain Amin, Amir Ali." InceptFood: A CNN Based Classification Approach for Recognizing Food Images". 2022.
[17] Seon-Joo Park, Akmaljon Palvanov, Chang-Ho Lee, Nanoom Jeong, Young-Im Choand Hae-Jeung Lee. "The development of food image detection and recognition model of Korean food for mobile dietary management". 2019;13(6):521-528.
[18] Diksha Solanki, Detection and classification of food consumption using convolutional neural networks, ISSN: 2319-7064 (2022).
[19] Asif Mahbub Uddin, Abdullah Al Miraj, Moumita Sen Sarma, Avishek Das, Md. Manjurul Gani. "Traditional bengali food classification using convolutional neural network". doi:10.1109/TENSYMP52854.2021.9550874. 2022.
[20] Deepak Rana, Sonia Kukreja, Food recognition using convolution neural network, ISSN: 2395-0056 (2022).
[21] J.P. Nayak, B.D. Parameshachari, K.S. Soyjaudah, R. Banu, A.C. Nuthan, Identification of PCB faults using image processing, in: 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), IEEE, 2017, pp. 1–4.
[22] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, NIPS (2012) 4.
[23] Y. Kawano, K. Yanai, Foodcam: A real-time food recognition system on a smartphone, Multimedia Tools Appl. 74 (2015) 5263–5287.
[24] T. Nguyen, T. Nguyen, B.M. Nguyen, G. Nguyen, Efficient time-series forecasting using neural network and opposition-based coral reefs optimization, Int. J. Comput. Intell. Syst. 12 (2) (2019) 1144.
[25] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, PFID: Pittsburgh fastfood image dataset, in: Image Processing (ICIP), 2009 16th IEEE International Conference on, 2009, pp. 289–292.
[26] H.A. Li, J. Fan, K. Yu, X. Qi, Z. Wen, Q. Hua, … Q. Zheng, Medical image coloring based on gabor filtering for internet of medical things, IEEE Access 8 (2020) 104016–104025.
[27] Y. Kawano, K. Yanai, Food image recognition with deep convolutional features, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Adjunct Publication, 2014, pp. 589–593.
[28] Lukas Bossard, Matthieu Guillaumin, Luc Van Gool1. "Food- 101 – Mining Discriminative Components with Random Forests". 2022.
[29] Prakhar Tripathi. Ttransfer learning on deep neural network: a case study on food-101 food classifier. 2022.