



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijrset@gmail.com

 www.ijrset.com

A Comprehensive Survey on Exploring Convolutional Neural Networks for Image-Based Nutrition Analysis

Poreddy Vara Sreeja¹, Deverakonda Neha², Sana Tabassum³, Nadia Anjum⁴

B.E Student, Department of Artificial Intelligence and Data Science, Stanley College of Engineering and Technology for Women, India^{1,2,3}

Assistant Professor, Department of Artificial Intelligence and Data Science and Computer Engineering (CME), Stanley College of Engineering and Technology for Women, India⁴

ABSTRACT: This survey paper delves into the realm of nutritional content analysis in the context of global fitness awareness and the imperative role of a well-maintained diet in mitigating health concerns. The primary challenge addressed is the need for efficient monitoring of nutritional and caloric content in daily food intake. This review paper provides a comprehensive exploration of image classification methodologies, with a specific focus on the application of Convolutional Neural Networks (CNNs) in the realm of nutritional analysis. The versatility of CNNs in automatically learning intricate spatial hierarchies of features, such as textures, shapes, and compositions within food images, is a central theme. While yielding promising results, the paper candidly acknowledges challenges and limitations observed during real-life applications. This research significantly contributes to the advancement of automated nutritional analysis systems, specifically within the realm of image-based food classification, demonstrating a nuanced awareness of practical considerations.

KEYWORDS: Convolutional Neural Network (CNN), Image Classification

I. INTRODUCTION

1.1 Image classification

Image classification, a quintessential task in computer vision, revolves around the assignment of labels or classes to input images. Positioned within the realm of supervised learning, this problem entails the training of a model on a labeled dataset, comprising images paired with corresponding class labels. The trained model is then deployed to predict the class labels of new, unseen images.

In the expansive landscape of image classification architectures, convolutional neural networks (CNNs) stand out as a prominent and widely embraced choice. This review paper embarks on a comprehensive exploration of image classification methodologies, with a particular emphasis on the efficacy and popularity of CNNs.[7],[9].

1.2 Convolutional Neural Networks

CNNs have emerged as a cornerstone in image recognition, object detection, and classification tasks, showcasing their versatility in various domains such as facial recognition and object detection. Our model harnesses the power of CNNs, comprising four essential layers: Convolutional layer, Rectified Linear Units (ReLU) layer, Pooling layer, and Fully Connected layer.[2-5].

1.2.1 Convolutional layer

The Convolutional layer, the first in the sequence, plays a pivotal role in reducing the number of parameters associated with image data. By convolving the input and passing the results to subsequent layers, this layer mimics the response of neurons in the visual cortex, enabling the network to recognize intricate patterns. The convolution operation addresses challenges posed by the vast amounts of data in images, providing a solution to issues like disappearing or exploding gradients encountered in traditional neural networks.[8]

In the context of image inputs, the convolutional layer engages with three-dimensional matrices, encapsulating the spatial dimensions of height, width, and depth corresponding to RGB channels. The feature detector, traverses the receptive fields of the image, executing a convolution process that detects the presence of specific features.

The feature detector, represented as a two-dimensional array of weights, undergoes a convolution process by applying a dot product to the input pixels. This operation results in an output array, referred to as a feature map, activation map, or convolved feature. A crucial aspect of this process is parameter sharing, where the weights in the feature detector remain fixed across the image, facilitating the identification of consistent features. During training, certain parameters, such as weight values, undergo adjustments through the iterative processes of backpropagation and gradient descent. However, the convolutional layer operation is nuanced by three pivotal hyperparameters: the number of filters, the stride, and zero-padding. The number of filters influences the depth of the output, introducing a multi-dimensional aspect to the feature maps. Stride determines the spatial distance covered by the kernel during convolution, impacting the size of the output. Zero-padding comes into play to ensure compatibility when filter dimensions do not align with the input image, presenting three distinct types: Valid padding, Same padding, and Full padding.[2],[6].

1.2.2 ReLu layer

Following the Convolutional layer, the ReLU layer introduces nonlinearity to the system, crucial for capturing complex features in the data. Replacing traditional nonlinear functions like tanh and sigmoid, ReLU facilitates faster training without compromising accuracy.[12]

1.2.3 Pooling layer

The Pooling layer, be it local or global, is instrumental in reducing data dimensionality and enhancing computational efficiency. By combining the outputs of neuron clusters, pooling layers contribute to the overall simplification of the computational process. Different pooling strategies, such as max pooling and average pooling, cater to diverse requirements in feature extraction.[14]

1.2.4 Fully Connected layer

The Fully Connected layer serves as the culmination of the CNN architecture, amalgamating the features extracted by previous layers for the ultimate purpose of image classification. In this layer, each neuron is directly connected to every neuron in the subsequent layer, providing a weighted sum of pixels as input. While Convolution, ReLU, and Pooling layers focus on feature extraction, the Fully Connected layer plays a pivotal role in decision-making, utilizing a softmax activation function for classification tasks.

As we traverse the intricacies of these layers, the survey aims to elucidate the role and significance of each component in CNNs. By understanding the interplay between these layers, we seek to contribute to the broader comprehension of CNNs and their application in image-based tasks, positioning them as indispensable tools in the realm of deep learning.[14]

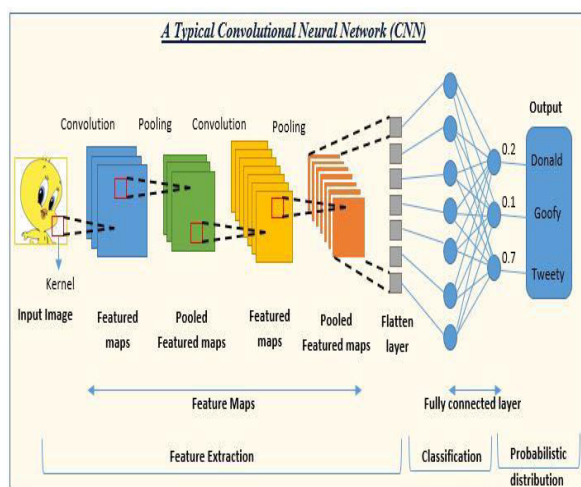


Fig 1: Convolutional Neural Network

II. LITERATURE SURVEY

Landu Jiang, et al., [2] focused on food recognition and dietary assessment using deep learning, employing the Faster R-CNN model for object detection and nutrition analysis. The approach involves generating Regions of Interest (ROIs) and utilizing a deep neural network for feature extraction in food item recognition.

Fengqing Zhu, et al., [3] developed a dietary assessment system for mobile devices, addressing the challenge of accurately measuring dietary intake in nutrition and health research. The developed tool aims to replace traditional food record methods and is being continually refined for improved accuracy and usability.

Naralasetti Hema Latha, et al., [4] focused on the dataset is unique for its considerations of camera type, shooting angle, and lighting variations, contributing to a diverse and comprehensive set for algorithm evaluation. Which employs color-texture segmentation, graph cut segmentation, and deep neural network algorithms to showcase the dataset's utility.

Paolo Napoletano, et al., [5] introduced a novel dataset for evaluating food recognition algorithms, comprising 1,027 canteen trays with 3,616 instances of 73 food classes acquired in a real canteen environment. The images are manually segmented, and an automatic tray analysis pipeline is designed, utilizing Convolutional Neural Networks (CNNs) for optimal performance.

Horea Muresan and Mihai Oltean, [6] Introduced a complex database of fruit images and conducted numerical experiments using the TensorFlow library for image classification. Future objectives focus on enhancing neural network accuracy by experimenting with network structure, including the replacement of fully connected layers with convolutional layers and exploring alternative activation functions.

Simon Mezgec and Barbara Korousi Seljak, [7] developed a NutriNet, a deep convolutional neural network architecture for food and drink image detection and recognition with a focus on high classification accuracy and efficient model training. Future research aims to enhance NutriNet's performance on real-world images with noise, explore automated optimization methods for architecture modifications, and address overfitting by incorporating additional patient-contributed images in a comprehensive food and drink image recognition framework.

Mustafa Al-Saffar and Wadhah Baiee, [8] proposed a system that analyses meal images to generate titles and ingredient lists. Pioneering the estimation of food ingredients from photographs, the model achieves a notable 85% accuracy in predicting and quantifying food attributes.

Baidaa Mutasher Rashed and Nirvana Popescu, [9] performance on evaluation of machine-learning and deep-learning algorithms for medical data diagnosis, chest X-ray, and melanoma skin cancer datasets. Machine-learning algorithms, with deep-learning CNN outperforming. Emphasizes the impact of dataset characteristics and preprocessing methods on classification outcomes, providing insights into the application of modern ML/DL methodologies in disease research and suggesting the potential for improved results with hybrid algorithms.

Shubhangi A. Joshi, et al., [10] introduced an optimized CNN framework for binary classification of breast cancer histopathology slides, utilizing a custom model with Xception. Two other deep CNNs with EfficientNetB0 and ResNet50 backbones were compared, demonstrated their application to the IDC dataset, fine-tuning the Xception model for invasive carcinoma detection, achieving an accuracy of 88.08% and showcasing the model's generalization and transfer learning capabilities for enhanced histopathological analysis in breast cancer diagnosis.

Md. Anwar Hossain and Md. Shahriar Alam Sajib, [11] developed a model showcasing image recognition capabilities, with potential extensions to object and character recognition, as well as real-time object recognition. The experiment highlights the superiority of Convolutional Neural Networks (CNN) over other classifiers.

Poonam Dhiman, et al., [12] proposed a Faster-CNN model, applied to citrus fruit defect identification, which achieves an average accuracy of 96.92%. Through magnitude-based pruning and polynomial decay-based sparsity.



Benedetta Tondi, et al., [13] proposed a general method that effectively handles various operative conditions, including non-aligned double JPEG compression scenarios, demonstrating superior performance compared to existing state-of-the-art solutions.

Mingyuan Xin and Yong Wang, [14] proposed Deep convolutional neural networks that excel in identifying distortion-invariant images through implicit feature extraction, leveraging weight sharing to reduce network complexity.

Petrus Mursanto, et al., [15] proposed a deep learning model, particularly Densenet121, which exhibits excellent performance with AUROC values exceeding 0.98, accuracy above 0.90, precision, recall, and F1 score surpassing 0.92. Rahul Sharma, et al., [16] proposed The CNN model's superior performance compared to other machine learning classifiers like SVM, KNN, Decision Tree, and Random Forest. for plant leaf disease identification, achieving a 97.66% accuracy in detecting potato leaf disease.

Dulari Bhatt, et al., [17] focused on the significance of CNN's block-based architecture for modular learning and the potential of dimension-based CNN for recognizing three-dimensional objects, suggesting promising avenues for future research in 3D object recognition and Neural Architecture Search (NAS) techniques.

Imrus Salehin, et al., [18] explored advanced approaches like LSTM and CNN architectures for medical image classification, achieving promising outcomes with custom MedvCNN and MedvLSTM.

3 Table 1: Techniques used in various research

Sl.no	Author and Reference	Dataset	Algorithm	Method	Training-Testing ratio	Results	Summary
1	Landu Jiang et al., [2]	UEC-FOOD100, UEC-FOOD256	CNN, Faster R-CNN, SVM	Convolutional Neural Network (CNN) classifies the ROIs into various food categories, aided by a regression module for precise food location.	80% - Training 20% - Testing	93.1% Accuracy	This study uses the Faster R-CNN model and deep learning for food recognition and dietary assessment, achieving efficient results.
2	Fengqing Zhu et al., [3]	FNDDS	SVM	configurations: client-server, users sent images to the server for analysis, and standalone, data is retrieved from the FNDDS database for further	50% Training, 50% Testing	97.2% accuracy	Employing the Faster R-CNN model and deep learning, this study efficiently recognizes food and assesses diets.



				analysis.			
3	Naralasetti Hema Latha et al., [4]	Word2Vec	CNN	Picture grouping using Convolution Neural Networks. Items location, acknowledgment.		The nutritional facts present in the food item have been obtained	Provided a dataset of 3000 food images, emphasizing characteristics like camera type and lighting variations. food detection algorithms, with demonstrated results using segmentation and deep neural network approaches.
4	Paolo Napoletano et al., [5]	Food dataset	CNN, SVM,k-nn	A dataset of 1,010 canteen trays split their training and testing. classifier training is extracted from ground-truth segmented regions. The k-NN classifier used and SVM employs a radial basis function kernel determined local classification	70% Training, 30% Testing	78.9% Accuracy	The paper introduces a dataset with 1,027 canteen trays for evaluating food recognition algorithms, achieving optimal performance with CNN-based features.
5	Horea Muresan and Mihai Oltean [6]	CIFAR, Fruits 360, MNIST	CNN	A convolutional neural network (CNN) was employed, characterized by its use of convolutional layers, pooling layers, ReLU layers, fully connected layers,	75% training, 25% testing	98.73% Accuracy	The study introduces a fruit image database, utilizing TensorFlow for classification, with a focus on improving neural network accuracy through structural experimentation.



				loss layers. Tensor Flow classification			
6	Simon Mezgec and Barbara Korousi Seljak [7]	UEC-FOOD100, UEC-FOOD256	CNN, Faster R-CNN, SVM	Classifying images A weekly Python script checks for new images, processes image fine-tune the model, utilizing Caffe as the deep learning framework. automatic class corrections, Google image searches, and detection model usage.	70% - Training, 10% - Validation, 20% - Testing	87.96% Accuracy	The paper introduces NutriNet, a deep CNN for accurate food and drink image recognition in a 520-class dataset, integrated into the PD Nutrition app with continuous online training.
7	Mustafa Al-Saffar and Wadhah Baiee, [8]	Yelp dataset, Nutrition 5k dataset	CNN	Prediction and transformer-based ingredient list generation with nutrition facts estimation. Preprocessing extracting food photos predicting ingredients using a pre-trained model, and estimating nutritional information.		85% Accuracy	This research presents a groundbreaking nutritional assessment system, using food photographs to generate meal titles and ingredient lists with an 85% accuracy.
8	Baidaa Mutasher Rashed and Nirvana Popescu, [9]	Medical dataset, Chest X-ray dataset, Dermoscopy skin	ANN, SVM, KNN, DT, RF, NB, LR, RS, CNN	Classification models' performance. Identifying lung diseases and melanoma skin cancer. accuracy, sensitivity,	70% - training and 30% - testing	93% Accuracy	This study analyzes medical datasets using preprocessing and segmentation. Employing diverse machine learning methods,

		cancer dataset		specificity, precision, recall, F-measure, and AUC (area under curve).			including CNN, proves effective in accurately diagnosing lung diseases and melanoma skin cancer.
9	Shubhangi A. Joshi, et al., [10]	BreakHis Histopathological Image Dataset, IDC Breast Histopathological Image Dataset	CNN	The CNN models' hyperparameters are manually tuned. Three custom convolutional layers. Relu activation. 50 epochs, Adam Optimizer, classification.	70% - Training, 15% - Validation, 15% - Testing	93.33% accuracy	The study demonstrated the model's transfer learning ability on a new dataset, emphasizing the potential for assisting histopathologists in breast cancer diagnosis.
10	Md. Anwar Hossain and Md. Shahriar Alam Sajib, [11]	CIFAR-10 dataset	CNN	(CNN) in deep learning for image recognition. CNNs extract feature maps from 2D image filters, mapping, image pixels handwriting recognition, natural object classification, and segmentation.	83% training, 17% testing	93.47% accuracy	The model achieved a recognition rate of 93.47% on 10,118 test cases after three hundred epochs. Despite misclassifying 661 images, these results are considered satisfactory, especially given the simplicity of the model, CPU training, and limited training time.
11	Poonam Dhiman, et al., [12]	Citrus fruit disease dataset	CNN	80–20 cross-validation cross-entropy loss function classifying diseases in a citrus fruit image dataset. Canker precision, recall,	80% - training, 20% - testing	96.92% accuracy	The paper addresses resource constraints in the MIC environment by implementing efficient deep learning on low-resource devices and introducing pruning. Applying Magnitude-based



				specificity, F-measure Pruning.			pruning on Faster-CNN achieves 96.92% accuracy, enhancing resource efficiency.
12	Benedetta Tondi, et al., [13]	RAISE dataset	CNN	Dataset construction procedure, and training settings. statistical analysis. alignment achieve superior accuracy.	85% training, 15% testing		The paper proposes a CNN-based method for primary quantization matrix estimation, excelling in accuracy and versatility, especially for non-aligned compression in image tampering.
13	Mingyuan Xin and Yong Wang, [14]	MNIST	CNN, KNN, SVM, BP	Image classification based on deep convolutional neural networks. training set. Evaluation, classifier's predictions.		69.47% accuracy	Deep convolutional neural networks employ weight sharing and feature detection for image tasks.
14	Petrus Mursanto, et al., [15]	Food datasets	CNN	Depth-based and multipath-based feature extraction in traditional food recognition. The depth-based CNN gradient vanishing problem. AUROC in the evaluation of spatial, depth, and multipath improvements for better recognition capabilities.	70% - training, 20% - testing, 10% - validation	98% accuracy	The research employs a deep learning model to identify and preserve 34 types of traditional foods in Indonesia with professional data acquisition. The Densenet121 model shows exceptional performance.



15	Rahul Sharma, et al., [16]	Rice dataset	CNN, SVM, KNN, Random Forest, Decision Tree	Datasets, using metrics like accuracy, precision, recall, F1 score. A confusion matrix model's performance, The study underscores the significance of these evaluations for effective plant disease classification.	80% - Training, 20% - Testing	98.28 accuracy	The paper presents a CNN model achieving 97.66% accuracy in detecting potato leaf disease using 5932 rice and 1500 mixed leaf images.
16	Dulari Bhatt, et al., [17]	PathM NIST	CNN, LSTM, RNN	Custom CNN Hybrid LSTM model. analysis, pre-trained CNNs. accuracy by capturing temporal dependencies. Thorough evaluation across seven classes measures accuracy.	83% training, 10% validation, 7% testing	85% accuracy	The study uses LSTM and CNN architectures for medical image classification, achieving superior results on the MedM-NIST dataset.
17	Imrus Salehin, et al., [18]	DICO M dataset	CNN	U-Net and LinkNet architectures. ResNet34 Inception ResNet V2 encoders segmentation.		Successful segmentation of the aorta using U-Net and LinkNet architectures with ResNet34 and Inception ResNet V2 encoders.	The paper explores aorta segmentation using U-Net and LinkNet architectures with different encoders on a small, semi-automatically labeled dataset of 154 CT scans.

III. CONCLUSION

In conclusion, this review has comprehensively explored the landscape of image classification methodologies with a specific emphasis on Convolutional Neural Networks (CNNs). Through an in-depth analysis of the literature, we have witnessed the evolution and impact of CNNs in addressing the intricate task of assigning class labels to input images.

The versatility and effectiveness of CNNs in automatically learning spatial hierarchies of features, including edges, textures, and shapes, have been a central theme throughout this review. The surveyed papers collectively demonstrate how CNNs have become a cornerstone in computer vision, particularly in image classification tasks. Their success is attributed to their ability to discern complex patterns within images, leading to improved accuracy and generalization on diverse datasets.

As we conclude, it is evident that CNNs have not only redefined the landscape of image classification but have also paved the way for advancements in related fields such as object detection, image segmentation, and feature extraction. The reviewed literature highlights ongoing research efforts aimed at enhancing CNN architectures, addressing challenges, and extending their applicability to diverse domains.

REFERENCES

1. David J. Attokaren, Ian G. Fernandes, A. Sriram, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi, "Food Classification from Images Using Convolutional Neural Networks", 2017.
2. Bappaditya Mandal, N. B. Puhan and Avijit Verma, "Deep Convolutional Generative Adversarial Network Based Food Recognition Using Partially Labeled Data", 2018.
3. Heng Zhao, Kim-Hui Yap, Alex C. Kot, Lingyu Duan, Ngai-Man Cheung, "Few-shot and Manyshot Fusion Learning in Mobile Visual Food Recognition", 2018.
4. M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.
5. R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.
6. S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
7. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
8. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
9. A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3. IEEE, 2006, pp. 850–855.
10. H. Kagaya, K. Aizawa, and M. Ogawa, "Food detection and recognition using convolutional neural network," in *Proceedings of the ACM International Conference on Multimedia*, ser. MM '14, 2014, pp. 1085–1088.
11. Y. Kawano and K. Yanai, "Food image recognition with deep convolutional features," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14 Adjunct, 2014, pp. 589–593.
12. K. Kitamura, T. Yamasaki, and K. Aizawa, "Foodlog: Capture, analysis and retrieval of personal food images via web," in *Proceedings of the ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities*, 2009, pp. 23–30.
13. F. Kong and J. Tan, "Dietcam: Automatic dietary assessment with mobile camera phones," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 147–163, 2012.
14. O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar, "MenuMatch: Restaurant-specific food logging from images," in *Applications of Computer Vision (WACV)*, 2015 IEEE Winter Conference on. IEEE, 2015, pp. 844–851.
15. Y. Kawano and K. Yanai, "Foodcam: A real-time food recognition system on a smartphone," *Multimedia Tools and Applications*, pp. 1–25, 2014.
16. F. Zhu, M. Bosch, I. Woo, S. Kim, C. Boushey, D. Ebert, and E. Delp, "The use of mobile devices in aiding dietary assessment and evaluation," *Selected Topics in Signal Processing*, IEEE Journal of, vol. 4, no. 4, pp. 756–766, 2010.
17. Z. Ahmad, N. Khanna, D. A. Kerr, C. J. Boushey, and E. J. Delp, "A mobile phone user interface for image-based dietary assessment," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 903 007–903 007.
18. M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation of food intake using a mobile device," in *Applications of Computer Vision (WACV)*, 2009 Workshop on, 2009, pp. 1–8.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details