

Convolutional Neural Networks in Computer-Aided Diagnosis of Colorectal Polyps and Cancer: A Review

Kamyab Keshtkar¹

School of Electrical and Computer Engineering, University of Tehran, Iran

Abstract. As a relatively high percentage of adenoma polyps are missed, a computer-aided diagnosis (CAD) tool based on deep learning can aid the endoscopist in diagnosing colorectal polyps or colorectal cancer in order to decrease polyps missing rate and prevent colorectal cancer mortality. Convolutional Neural Network (CNN) is a deep learning method and has achieved better results in detecting and segmenting specific objects in images in the last decade than conventional models such as regression, support vector machines or artificial neural networks. In recent years, based on the studies in medical imaging criteria, CNN models have acquired promising results in detecting masses and lesions in various body organs, including colorectal polyps. In this review, the structure and architecture of CNN models and how colonoscopy images are processed as input and converted to the output are explained in detail. In most primary studies conducted in the colorectal polyp detection and classification field, the CNN model has been regarded as a black box since the calculations performed at different layers in the model training process have not been clarified precisely. Furthermore, I discuss the differences between the CNN and conventional models, inspect how to train the CNN model for diagnosing colorectal polyps or cancer, and evaluate model performance after the training process.

Keywords: convolutional neural networks (CNNs), deep learning, computer-aided diagnosis, colorectal polyps, colorectal cancer, colonoscopy

I. Introduction

Machine learning is an artificial intelligence method that a computer learns to do specific tasks from available data without anyone explicitly programming it. Machine learning approaches are divided into three core methods: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is most broadly used among other core methods, and popular algorithms, include linear regression, logistic regression, artificial neural networks, support vector machines, decision trees, and decision forests. In these algorithms, the goal is to find a relationship between an (x, y) pair and determine the function f , $f(x)=y$, using available samples of (x, y) shapes [1]. In an (x, y) pair, x and y can be regarded as an input and output of a system, respectively.

Artificial neural networks, or simply called neural networks, is a subclass of machine learning in which the learning process is inspired by the mathematical modeling of nerve cells in the human body. The overall structure of neural networks is layered. Each layer consists of several nodes connected in two consecutive layers (Figure 1). Each node models a neuron in the nerve system and connections among them represent synapses in the nerve system that transmit information between two nodes [2]. This method is called deep learning when the number of hidden layers in the neural network exceeds a certain amount (5 to 20 layers) [3].

¹ e-mail: k.keshtkar@ut.ac.ir

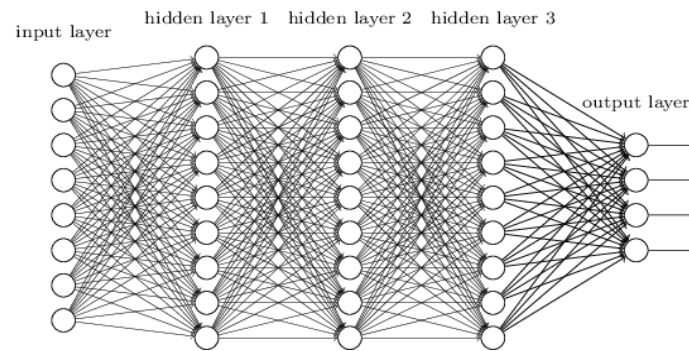


Figure 1. Artificial neural network architecture. Figure from [4].

Convolutional Neural Network (CNN) is a deep learning model used in various computer vision tasks such as object detection and face recognition, emerging in the early 1990s [3]. In 2012, CNN was exploited to classify 1.2 million images into 1000 various classes, with phenomenal results being presented in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [5][6]. After that, CNNs became the prevalent method in computer vision tasks [3].

Colorectal cancer is the third most common cancer and the second leading cause of mortality worldwide [7], with 1.1 million new cases and 576 thousand deaths among 185 countries in 2020 [8]. Polyps in the colon can precede colorectal cancer; thus, timely diagnosis and removal of these polyps by colonoscopy can prevent colorectal cancer and its mortality. Moreover, diagnosing colorectal cancer or malignant polyps early on by colonoscopy can increase the survival rate in the long term [9]. However, according to studies, up to 30% of adenomas polyps are missed during colonoscopy screening by endoscopists. In this situation, developing a computer-aided diagnosis (CAD) tool based on deep learning, or artificial intelligence in general terms, can help minimize the polyps missing rate [10].

Computer-aided diagnosis (CAD) is a tool that can assist doctors and radiologists in making effective diagnostic and treatment decisions. The CAD relies on medical image analysis, and medical imaging is no exception to the other computer vision tasks for which CNN is applied, including magnetic resonance imaging (MRI), X-ray, computed tomography (CT) [11], histological images [12]. In recent years, CNNs have been used in various fields of medical imaging such as detection of diabetic retinopathy and related eye diseases [13], breast lesion detection and classification [14], brain tumor detection in MRI images [15], skin lesion classification [16], automatic colorectal polyp detection and classification in colonoscopy videos [17], and coronavirus disease 2019 detection in chest X-ray images [18]. CNNs show remarkable medical imaging performance compared with conventional models like support vector machines (SVM), artificial neural networks, and k-nearest neighbor according to primary and secondary studies such as Shin et al. [19], Pan et al. [20], and Anwar et al. [11]. In some cases, CNNs even attained a higher accuracy than humans. For instance, Choi et al. [21] showed that CNN models were approximately 4% to 5% more accurate than expert endoscopists in colorectal cancer diagnosis and polyp classification.

This review concentrates on applying CNN models for detecting and classifying colorectal polyps in colonoscopy images or videos. This paper begins with the introduction (section I), and section II presents the differences, benefits, and drawbacks of conventional and CNN models. In section III, I describe the components of CNN models, their tasks in the learning process, and the types of outputs. Section IV discusses the preparation of colonoscopy images as inputs of the CNN model and details the training, validation and testing process. In section V, I define different evaluation metrics and how to measure them. In section VI and section VII, I explain two techniques to prevent overfitting and overcome data shortage for training the model, called data augmentation and transfer learning, respectively. Section VIII

details the limitations of CNN models and the approaches that can lead to improving the diagnostic performance of these models in future studies. Finally, conclusions are drawn in section IX.

II. What are the differences between conventional machine learning models and CNNs?

In most medical studies, conventional or traditional models like logistic regression [22], linear regression, artificial neural networks [23], support vector machines [24], and random forest [25] are used for diagnostic purposes. In these models, features are extracted manually, and this technique is called hand-crafted feature extraction [26]. These features are chosen based on methods such as histogram of oriented gradient (HOG) or hue histogram in images by persons [27]. However, CNNs extract features automatically, and features are learned by the CNN model (Figure 2), achieving superior accuracy than conventional models, especially when data is in the form of images [26].

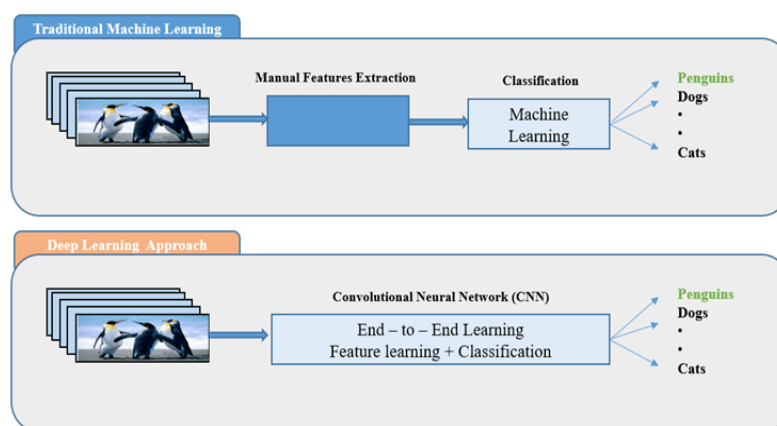


Figure 2. Comparison between workflow of conventional or traditional models and deep learning or CNN models. Figure from [28].

On the other hand, we need more data to train CNN models than conventional models since CNNs have millions of learnable features or parameters [26]. In contrast, the number of features in conventional models (less than a dozen) is much less than CNNs. Thus, the computational volume and cost go higher. To solve this problem, it is better to exploit the graphical processing unit (GPU) instead of the central processing unit (CPU) for training CNNs. GPU or graphic card processor consists of more cores than the CPU, whereas the computation power of the GPU core is less. Hence, GPUs process faster in matrix computing, and computations in the training process are of the same type.

III. How does the Convolutional Neural Network Black-Box work?

Inputs for the CNNs which are kind of artificial neural networks, should have grid patterns, e.g., images and videos. In this model, image features are automatically learned and extracted from low-level to high-level. The early layers extract simple or low-level features like edges or corners, while the late layers extract complex or high-level features like specific objects. CNN models are composed of multiple layers, which are classified into convolutional layers, pooling layers, and fully connected layers. The role of convolutional and pooling layers is feature extraction, and fully connected layers map the extracted features into the output of the CNN model [26].

At first, images are fed into the first convolutional layer as inputs of the CNN model. Images are a kind of matrix that each pixel is equivalent to an element of the matrix. Each convolutional layer consists of filters or kernels that detect image features like edges, corners, textures, and objects. Kernels are also a type of matrix, and their elements are called weights,

the values of which are calculated in the training process. Then, the input image matrix convolves with the kernel, or in other words, the kernel matrix sweeps through the entire input matrix in the length and width directions where these two matrixes overlap (Figure 3). In the next step, the elements of the feature map matrix are summed with a constant value, called bias. Here, a non-linear activation function like ReLU (Rectified Linear Unit), sigmoid, or tanh (hyperbolic tangent) is applied to each acquired matrix element. A significant variation exists between the values of the feature map matrix elements, and the activation function diminishes variation between these values since the activation function range is limited. For instance, for all input values, the hyperbolic tangent or sigmoid outputs are between minus one and one, or zero and one, respectively (Figure 4). Between the convolutional layers, there are pooling layers that, through down-sampling, reduce the computation volume and remove redundancy from the input matrix of pooling layers by decreasing matrix dimensions. In these layers, a window with a specific size moves through the input matrix. Wherever the window is located, the maximum value of elements in the window (Max pooling) or the average of the elements' value in the window (Average pooling) creates each element of the output matrix of the pooling layers (Figure 5). Finally, the last convolutional layer's output matrix is converted to a vector (a kind of matrix with only one column) before being fed into fully connected layers as inputs or features extracted by the convolutional and pooling layers. Fully connected layers, as an artificial neural network, perform classification [29]. The overall structure of CNN, along with the types of layers described above, is shown in Figure 6.

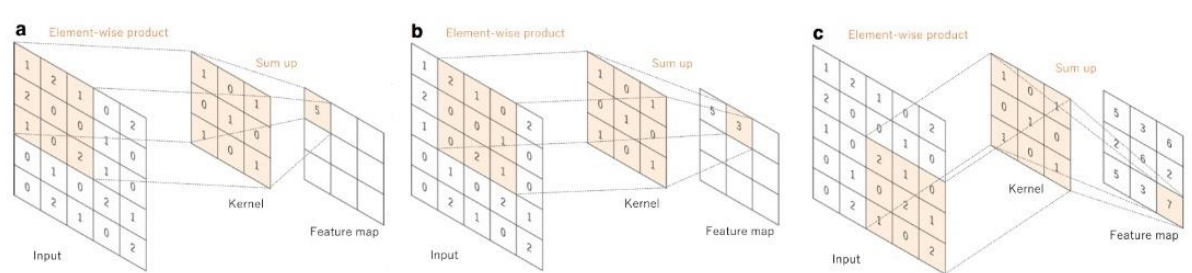


Figure 3. Three sample steps of the convolution operation. Wherever the kernel is located, the values of the corresponding elements in the two matrices are multiplied. Then, the resulting values are added together to form the corresponding element in the output matrix, called the feature map matrix. Figure from [26].

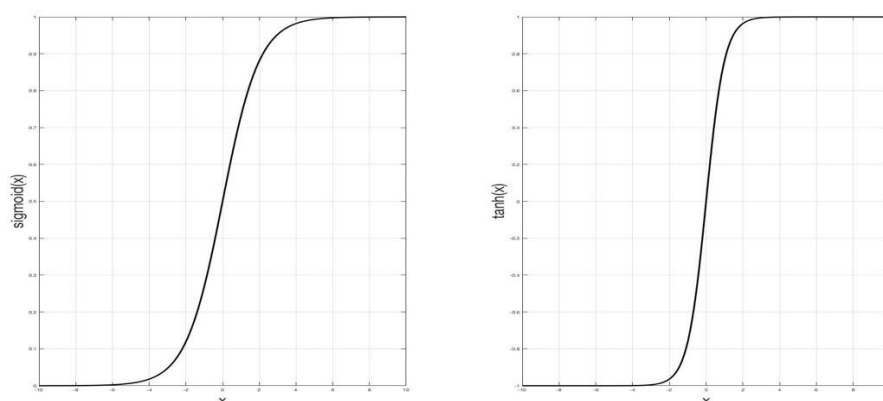


Figure 4. Sigmoid function graph (right-side) and hyperbolic tangent graph (left-side)

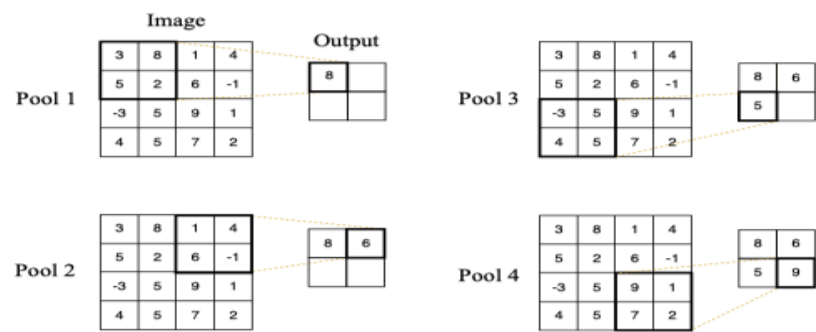


Figure 5. Steps of max pooling operation with a filter size 2*2 and a stride of 2. At each step, the maximum value is extracted in a 2*2 patch from the input matrix and creates an element of the output matrix.

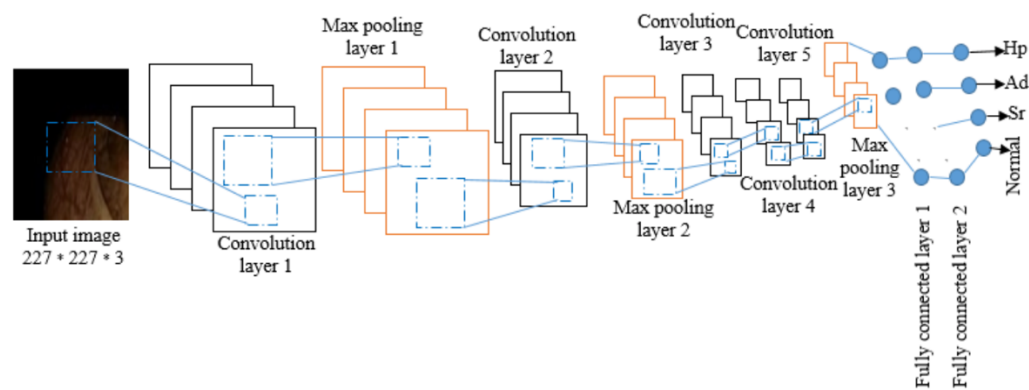


Figure 6. The overall structure of a CNN model, which classifies colonoscopy images into hyperplastic (Hp), adenomatous (Ad), serrated (Sr), and normal groups. Figure from [30].

In the related primary studies, the output of the CNN model can be divided into four overall categories, polyp classification, polyp detection, polyp localization, and polyp segmentation. In polyp classification, the CNN model recognizes the type of polyp (e.g., adenomatous, hyperplastic, serrated, or normal) in a colonoscopy image (Figure 6). In polyp detection, the CNN model only recognizes whether a colonoscopy image contains at least one polyp or not. In polyp localization, the CNN model marks the position of each polyp in a colonoscopy image with a rectangle but not with its exact shape (Figure 7). In polyp segmentation, the CNN model draws a margin around each polyp that it detects (Figure 8) [31]. CNN models have different architectures; each model has a specific name, such as U-Net [32], VGG [33], ResNet [34], and Faster R-CNN [35].

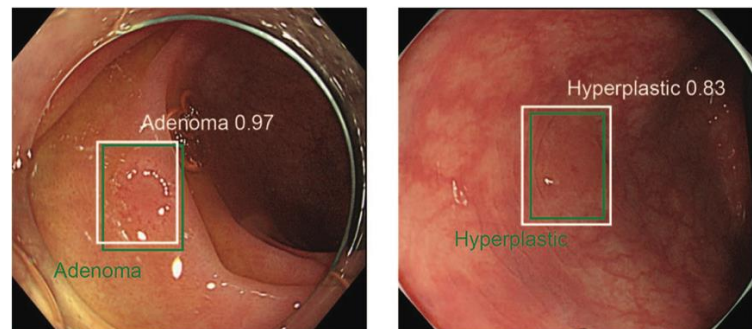


Figure 7. Samples of polyp localization. Green boxes show ground truth (gold standard), and white boxes show polyp localization by the CNN model. Figure from [17].

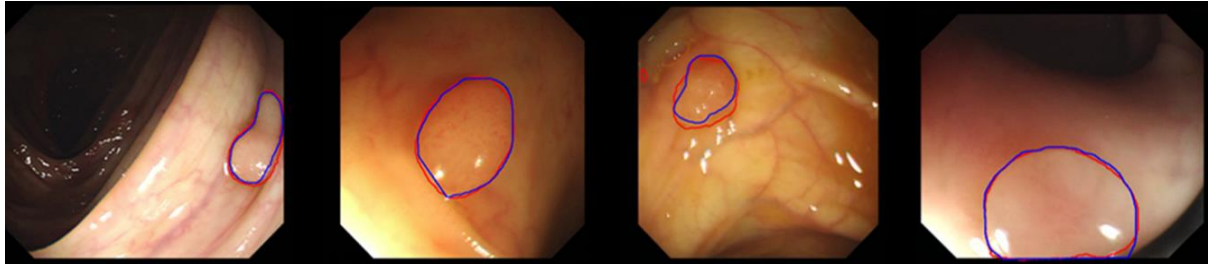


Figure 8. Samples of polyp segmentation. Blue contours represent ground truth (gold standard), and red contours represent polyp segmentation by the CNN model. Figure from [36].

IV. Model training, validating, and testing

The CNN model maps from the input data (i.e., images) to the corresponding output (ground truth or gold standard) through supervised learning method. The data obtained directly from colonoscopy, or raw data, is in video format, so videos should be converted into consecutive images to be processable to the CNN models as inputs. Then, the obtained images should be resized to the specific size required for input into the CNN model, though some CNN architectures like U-Net are compatible with various image sizes [37]. Many related primary studies have utilized pre-prepared and publicly available datasets, including ETIS-LARIB [38], CVC-CLINIC [39], ASU-Mayo Clinic Colonoscopy Video [40], and KVASIR [41]. In contrast, in some studies such as Ozawa et al. [17], Haj-Manouchehri et al. [42], Choi et al. [21] and Shafi et al. [30], colonoscopy videos were provided in collaboration with hospitals or institutes, and these videos needed to be converted to images. Finally, the dataset should be split into three parts: 90%, 5%, and 5% of the whole dataset (or in other ratios), used for the CNN model training (training dataset), validation, and testing (testing dataset), respectively.

The loss function reaches a minimum, or optimum point in the training process by optimization algorithms, referred to as back-propagation and gradient descent. The loss function (or cost function) measures the similarity among ground truth (gold standard) labels and the output of the CNN model, and the loss function's value is updated in each epoch (iteration) through the back-propagation algorithm. Sometimes, the loss function should be calculated through a subset of the training dataset instead of the whole training dataset due to lack of memory, increased efficiency and decreased computation cost. This subset is named mini-batch, and the mini-batch size usually is a power of 2 (e.g., 32). Gradient descent (Figure 9) is a kind of optimization algorithm that is defined as follows:

$$W(New) := W(Old) - \alpha \frac{\partial L}{\partial W}$$

In the above equation, L represents the loss function, and W denotes learnable parameters like kernels' weights in convolutional layers that are updated until the loss function converges

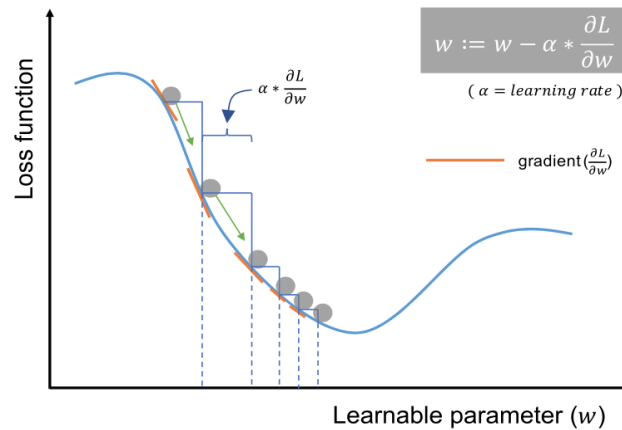


Figure 9. A schematic representation of the gradient descent algorithm. The gradient of the loss function means the direction of the steepest slope as the loss function increases. Therefore, moving in the opposite direction of the gradient with a learning rate step size leads to obtaining the loss function's minimum or local minimum point. The learnable parameter (w) value is updated iteratively until the gradient of the loss function reaches zero or the loss function reaches a minimum value. At this point, the optimal value of the learnable parameter is obtained, and our model has been trained on the dataset. Figure from [26].

a value and remains stable after at least 20 consecutive epochs. Also, α denotes the learning rate, a minute positive constant number usually between 0 and 1 [26].

In the validation phase, hyperparameters such as learning rate, mini-batch size, number of epochs, and type of loss function and optimizer are tuned, and the values are determined [26].

The CNNs performance is evaluated in the testing phase by measuring the evaluation metrics defined in section V [26]. Testing is divided into internal and external categories, based on whether the testing dataset and the training dataset are from the same place or not, respectively. An important note to consider is that in some papers, especially in medical science criteria, the testing phase is called the validation phase, which differs from the validation phase described above.

Sometimes, the CNN model performs astounding on the training dataset but does not perform well on the testing dataset or other datasets. In other words, there is a considerable interval between the test accuracy value and the training accuracy value. In this case, overfitting has occurred, and to overcome this issue, we can exploit various methods like data augmentation, transfer learning, dropout, regularization, and batch normalization [11].

V. Evaluation metrics

After the CNN model is trained on the training dataset, we have to evaluate our CNN model performance on another dataset (testing dataset). This evaluation is accomplished by measuring accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), which are defined below.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$PPV(Precision) = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

There are four indicators in these metrics, including True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). These indicators can be defined in the form of a two-by-two table, also called a diagnostic table.

		Ground Truth	
		Disease	Normal
Model Output	Positive	TP	FP
	Negative	FN	TN

The TP index indicates the number of images containing at least one polyp that the CNN model has correctly classified or detected, or the number of images in which the type of polyp is malignant and the model has correctly diagnosed the polyp type. The FN index represents the number of images containing at least one polyp while the CNN model cannot classify or detect it (or them), or the number of images containing malignant polyps falsely diagnosed as benign. The TN measure indicates the number of images without any polyps correctly classified by the CNN model as non-polyp images, or the number of images containing benign polyps correctly diagnosed as the benign type. Finally, FP represents the number of images without any polyps falsely classified by the CNN model as having at least one polyp, or the number of images in which the polyp is benign while the model has falsely diagnosed it as malignant.

VI. Data augmentation

In medical imaging, due to privacy concerns, the amount of dataset is often inadequate [26], and the dataset is unbalanced. This means the number of negative group images is as many times as positive group images. An unbalanced dataset leads the CNN model to perform weakly in positive group recognition, i.e., diagnosis of malignant polyps, and causes the values of the evaluation metrics to be lower than when the dataset is balanced [43]. Moreover, in the training phase of CNN, if the amount of the training dataset is insufficient, overfitting will ensue, which is explained in section IV [44]. Furthermore, the more layers the CNN model has, the larger the training dataset needs to be to train the CNN model [42].

Geometric transformations, as effective ways for augmentation, are applied to enlarge the training dataset. Transformations that can be used include rotation (0 degrees to 360 degrees), flip from top to bottom or right to left, zoom in or out, random brightness changes, and cropping of training images. In these ways, we acquire several new images from each original image [45]. Therefore, these transformations enhance the CNN model's diagnostic performance and the evaluation metrics values [44].

VII. Transfer learning

Transfer learning is another way to tackle the data shortage. In this method, firstly, the CNN is trained on another huge dataset (mostly ImageNet). Although ImageNet is a non-medical dataset, it can be useful when CNN is trained on medical images. Then, the CNN is initialized with the obtained pre-trained weights, and the layer weights of the CNN are updated persistently until the evaluation metric (accuracy) reaches its optimum value. In fact, updating the last layers' weights is usually adequate since, as mentioned in section III, the early layers learn the low-level features such as edges or corners, which are common in all images. However, if the category of images with which pre-trained and target models are trained differs considerably, the early layers' weights may also need to be updated [46].

VIII. Limitations of CNNs and future prospect

Although CNN models have achieved astonishing results in computer vision tasks, they may not perform optimally in some cases, and it is better to use hybrid models (combination of hand-crafted and automatic feature extraction methods) [29]. For instance, Shin et al. [27] demonstrated that the accuracy of a hybrid model involving the combination of HOG and hue histogram methods (hand-crafted feature extraction) with the dictionary learning method (automatic feature extraction) in polyp detection was 4% higher than the CNN model.

CNN models are far more data-hungry than conventional models because they have millions of learnable features or parameters. Also, due to patient privacy, medical images are less available than other types of images. To resolve this issue and to prevent overfitting, data augmentation (section VI) and transfer learning (section VII) have been utilized in most primary studies related to the subject of this paper. In addition, there is another method for overcoming data scarcity, called Generative Adversarial Networks (GANs), which was first designed by Goodfellow et al. [47] in 2014.

GAN is a machine learning framework that generates new sample images from original sample images among various domains, including medical images. This framework can be applied to images of a training dataset to enhance the variety and number of images, and boost the robustness of CNN models [48]. In Thomaz et al. [49], the Faster R-CNN model was trained on a training dataset augmented by the data augmentation method, obtained sensitivity (recall) of 61.0% in polyp segmentation on a testing dataset. In that study, the same model was also trained on a training dataset augmented by the GAN, improving sensitivity (recall) to 69.2% on the same testing dataset. As a result, it is recommended to exploit the GAN framework instead of the data augmentation method in future polyp detection and segmentation studies. Because using GANs not only improves the diagnostic performance of CNN models but also is not time-consuming in creating new images compared with other data augmentation methods.

As mentioned in Section IV, colonoscopy images should be rescaled to train CNN models. This resizing reduces images quality and misses some detailed information, e.g., small polypus lesions. These issues have negative impacts on CNN models training and thus these models' diagnostic performance. To resolve these issues in future studies, we can crop each colonoscopy image into several smaller patches and then feed these patches to the CNN model as input instead of rescaled images [50].

IX. Conclusions

In recent decades, CNN models have accomplished many computer vision tasks like object detection, image reconstruction, and medical imaging with notable results. In this paper, I have discussed the applications of CNN in diagnosing colorectal polyps or cancer while explaining the differences and advantages of CNN over conventional models. knowing the applications and benefits of the CNN models, as well as their limitations and drawbacks, will help develop computer-aided diagnosis tools. Such tools can enhance the endoscopists' performance during colonoscopy, ultimately increasing the colorectal cancer survival rate by reducing the polyps missing rate.

References

- [1] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015.
- [2] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, p. e00938, Nov. 2018.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* 2015 521:7553, vol. 521, no. 7553, pp. 436–444, May 2015.
- [4] "AI Is Not Magic. How Neural Networks Learn | CodeAhoy." [Online]. Available: <https://codeahoy.com/2017/07/28/ai-is-not-magic-how-neural-networks-learn/>. [Accessed: 30-Apr-2021].
- [5] A. Krizhevsky, ... I. S.-... information processing, and undefined 2012, "Imagenet classification with deep convolutional neural networks," *proceedings.neurips.cc*.
- [6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision* 2015 115:3, vol. 115, no. 3, pp. 211–252, Apr. 2015.
- [7] N. Keum and E. Giovannucci, "Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies," *Nature Reviews Gastroenterology & Hepatology* 2019 16:12, vol. 16, no. 12, pp. 713–732, Aug. 2019.
- [8] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.
- [9] A. G. Zauber, S. J. Winawer, M. J. O'Brien, I. Lansdorp-Vogelaar, M. van Ballegooijen, B. F. Hankey, W. Shi, J. H. Bond, M. Schapiro, J. F. Panish, E. T. Stewart, and J. D. Waye, "Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths," <http://dx.doi.org/10.1056/NEJMoa1100370>, vol. 366, no. 8, pp. 687–696, Feb. 2012.
- [10] P. Wang, P. Liu, J. R. Glissen Brown, T. M. Berzin, G. Zhou, S. Lei, X. Liu, L. Li, and X. Xiao, "Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study," *Gastroenterology*, vol. 159, no. 4, pp. 1252–1261.e5, Oct. 2020.
- [11] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review," *Journal of Medical Systems* 2018 42:11, vol. 42, no. 11, pp. 1–13, Oct. 2018.
- [12] E. Ribeiro, A. Uhl, and M. Hafner, "Colonic polyp classification with convolutional neural networks," in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2016, vol. 2016-Augus, pp. 253–258.
- [13] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. S. Yeo, S. Y. Lee, E. Y. M. Wong, C. Sabanayagam, M. Baskaran, F. Ibrahim, N. C. Tan, E. A. Finkelstein, E. L. Lamoureux, I. Y. Wong, N. M. Bressler, S. Sivaprasad, R. Varma, J. B. Jonas, M. G. He, C.-Y. Cheng, G. C. M. Cheung, T. Aung, W. Hsu, M. L. Lee, and T. Y. Wong, "Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, Dec. 2017.
- [14] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with Deep Learning," *Scientific Reports* 2018 8:1, vol. 8, no. 1, pp. 1–7, Mar. 2018.
- [15] M. Toğaçar, B. Ergen, and Z. Cömert, "BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model," *Medical Hypotheses*, vol. 134, p. 109531, Jan. 2020.
- [16] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature* 2017 542:7639, vol. 542, no. 7639, pp. 115–118, Jan. 2017.
- [17] T. Ozawa, S. Ishihara, M. Fujishiro, Y. Kumagai, S. Shichijo, and T. Tada, "Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks," *Therapeutic Advances in Gastroenterology*, vol. 13, 2020.
- [18] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Scientific Reports* 2020 10:1, vol. 10, no. 1, pp. 1–12, Nov. 2020.
- [19] Y. Shin and I. Balasingham, "Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2017, pp. 3277–3280.

- [20] Y. Pan, W. Huang, Z. Lin, W. Zhu, J. Zhou, J. Wong, and Z. Ding, "Brain tumor grading based on Neural Networks and Convolutional Neural Networks," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-November, pp. 699–702, Nov. 2015.
- [21] K. Choi, S. Choi, ... E. K. the I. E. in M., and U. 2020, "Computer-Aided Diagnosis for Colorectal Cancer using Deep Learning with Visual Explanations," *ieeexplore.ieee.org*, 2020.
- [22] V. F. van Ravesteijn, C. van Wijk, F. M. Vos, R. Truyen, J. F. Peters, J. Stoker, and L. J. van Vliet, "Computer-aided detection of polyps in CT colonography using logistic regression," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 120–131, Jan. 2010.
- [23] W. Wei and X. Yang, "Comparison of Diagnosis Accuracy between a Backpropagation Artificial Neural Network Model and Linear Regression in Digestive Disease Patients: An Empirical Research," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.
- [24] M. J. Gangeh, L. Sørensen, S. B. Shaker, M. S. Kamel, M. de Bruijne, and M. Loog, "A Texton-Based Approach for the Classification of Lung Parenchyma in CT Images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6363 LNCS, no. PART 3, pp. 595–602, 2010.
- [25] M. Anthimopoulos, S. Christodoulidis, A. Christe, and S. Mougiakakou, "Classification of interstitial lung disease patterns using local DCT features and random forest," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, pp. 6040–6043, Nov. 2014.
- [26] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4. Springer Verlag, pp. 611–629, 01-Aug-2018.
- [27] Y. Shin and I. Balasingham, "Automatic polyp frame screening using patch based combined feature and dictionary learning," *Computerized Medical Imaging and Graphics*, vol. 69, pp. 33–42, Nov. 2018.
- [28] I. T. Ahmed, C. S. Der, N. Jamil, and M. A. Mohamed, "Improve of contrast-distorted image quality assessment based on convolutional neural networks," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 5604–5614, Dec. 2019.
- [29] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep Learning vs. Traditional Computer Vision," *Advances in Intelligent Systems and Computing*, vol. 943, pp. 128–144, Apr. 2019.
- [30] A. S. M. Shafi and M. M. Rahman, "Decomposition of color wavelet with higher order statistical texture and convolutional neural network features set based classification of colorectal polyps from video endoscopy," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 2986–2996, 2020.
- [31] L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," *Artificial Intelligence in Medicine*, vol. 108, p. 101923, Aug. 2020.
- [32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, pp. 234–241.
- [33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016.
- [35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [36] Y. Guo, J. Bernal, B. M.-J. of Imaging, and undefined 2020, "Polyp segmentation with fully convolutional deep neural networks—extended evaluation study," *mdpi.com*, pp. 377–388, 2019.
- [37] A. Tashk, J. Herp, and E. Nadimi, "Fully Automatic Polyp Detection Based on a Novel U-Net Architecture and Morphological Post-Process," in *Proceedings - 2019 3rd International Conference on Control, Artificial Intelligence, Robotics and Optimization, ICCAIRO 2019*, 2019, pp. 37–41.
- [38] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer," *International Journal of Computer Assisted Radiology and Surgery* 2013 9:2, vol. 9, no. 2, pp. 283–293, Sep. 2013.
- [39] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, Jul. 2015.

- [40] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, Feb. 2016.
- [41] K. Pogorelov, K. Ranheim Randel, C. Griwodz, B. Hospital, N. Thomas de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. Thelin Schmidt Karolinska Institutet, S. Karolinska Hospital, S. Michael Riegler, P. Halvorsen, S. Losada Eskeland, T. de Lange, P. Thelin Schmidt, and M. Riegler, "Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection," *Proceedings of the 8th ACM on Multimedia Systems Conference*, vol. 6, 2017.
- [42] A. Haj-Manouchehri and H. M. Mohammadi, "Polyp detection using CNNs in colonoscopy video," *IET Computer Vision*, vol. 14, no. 5, pp. 241–247, 2020.
- [43] R. Fonolla, F. van der Sommen, R. M. Schreuder, E. J. Schoon, and P. H. N. de With, "Multi-modal classification of polyp malignancy using CNN features with balanced class augmentation," in *Proceedings - International Symposium on Biomedical Imaging*, 2019, vol. 2019-April, pp. 74–78.
- [44] L. Duran-Lopez, F. Luna-Perejon, I. Amaya-Rodriguez, J. Civit-Masot, A. Civit-Balcells, S. Vicente-Diaz, and A. Linares-Barranco, "Polyp detection in gastrointestinal images using faster regional convolutional neural network," in *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2019, vol. 4, pp. 626–631.
- [45] A. Bour, C. Castillo-Olea, B. Garcia-Zapirain, and S. Zahia, "Automatic colon polyp classification using Convolutional Neural Network: A Case Study at Basque Country," in *2019 IEEE 19th International Symposium on Signal Processing and Information Technology, ISSPIT 2019*, 2019.
- [46] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," 2016.
- [47] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in neural information processing systems*, vol. 27, pp. 2672–2680, 2014.
- [48] F. He, S. Chen, S. Li, L. Zhou, H. Zhang, H. Peng, and X. Huang, "Colonoscopic image synthesis for polyp detector enhancement via gan and adversarial training," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2021-April, pp. 1887–1891, Apr. 2021.
- [49] V. de Almeida Thomaz, C. A. Sierra-Franco, and A. B. Raposo, "Training data enhancements for improving colonic polyp detection using deep convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 111, p. 101988, Jan. 2021.
- [50] Y. Fang, D. Zhu, J. Yao, Y. Yuan, and K. Y. Tong, "ABC-Net: Area-Boundary Constraint Network with Dynamical Feature Selection for Colorectal Polyp Segmentation," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11799–11809, May 2021.