A Hybrid Machine Learning Techniques with Deep Neural Network Model for Colon Cancer Diagnosis

[1,2*]Amna Ali A. MOHAMED, [3]Aybaba HANÇERLİOĞULLARI

[1]*Department of Computer Engineering, Kastamonu University, Kastamonu, Türkiye*
[2]*Tripoli University, Tripoli, Libya*
[3]*Department of Physics, Faculty of Science, Kastamonu University, Kastamonu, Türkiye*

* Corresponding author: amnaalkmati@gmail.com

ORCID Numbers: 0000-0001-8344-6937 ([1]Amna Ali A. Mohamed) 0000-0000-1700-8480 ([2]Aybaba Hançerlioğulları)

Abstract **-** Colon cancer detection is a great significant task in medical diagnosis. the detection of colorectal cancer in an early stage can significantly facilitate clinicians' decision-making and reduce mortality. The accurate detection results help to explore symptomatic treatment promptly; this can be achieved by using automatic systems with histopathological images. The combination of convolutional neural networks and supervised machine learning methods are used to achieve better classification results than using individual pre-trained deep networks. Therefore, this study is aimed to get a high performance and accuracy of CNN combined them with supervised machine learning methods. Support Vector Machine (SVM), decision tree (DT) and k-nearest neighbour (KNN) as the classification of colon cancer to get the best accuracy.

Keywords: *Colon Disease Diagnose, CNN, Machine Learning, SVM, Decision Tree, K-Nearest Neighbour.*

INTRODUCTION

Colon disease diagnosis is usually done by image processing methods, and careful learning is needed to analyze these images. The image denoising in the image pre-processing steps can reduce the features effectiveness and the classifier methods can make a mistake and the results will not be good. Also in image processing some filters cannot give us the optimum results because of losing the features. In this work, an attempt has been made to combine machine learning skills such as learning in the neural network with learning and training in humans to reveal the areas of colon disease. In the first, the convolutional neural network with SqueezeNet is used to extract the features from the colon disease images. Finally, the machine learning, support vector machine (SVM), decision tree (DT), and k-nearest neighbor (KNN) methods are used to the classification step [1-10].

ARTIFICIAL NEURAL NETWORK

A subset of artificial intelligence known as an artificial neural network (ANN) is capable of learning intricate nonlinear patterns from a set of data. Parallel computing units called ANNs have recently emerged as effective classification tools. They were first inspired by the concept of modeling mathematics and engineering to simulate the decision-making and parallel processing capabilities of the human brain. Even though ANNs still share many of the same fundamental traits as the human brain, they function much differently than biological neural networks today in terms of how they make decisions. Two primary criteria can be used to classify various ANN types [11]. The network's encoding—or, more specifically, how it stores knowledge gleaned from the data—is the first requirement. ANNs are divided into supervised and unsupervised categories using this metric. The second criterion is how the networks are decoded, or how the network handles new data after learning something from the previous input. Using this criterion, feedforward and feedback ANNs are distinguished.

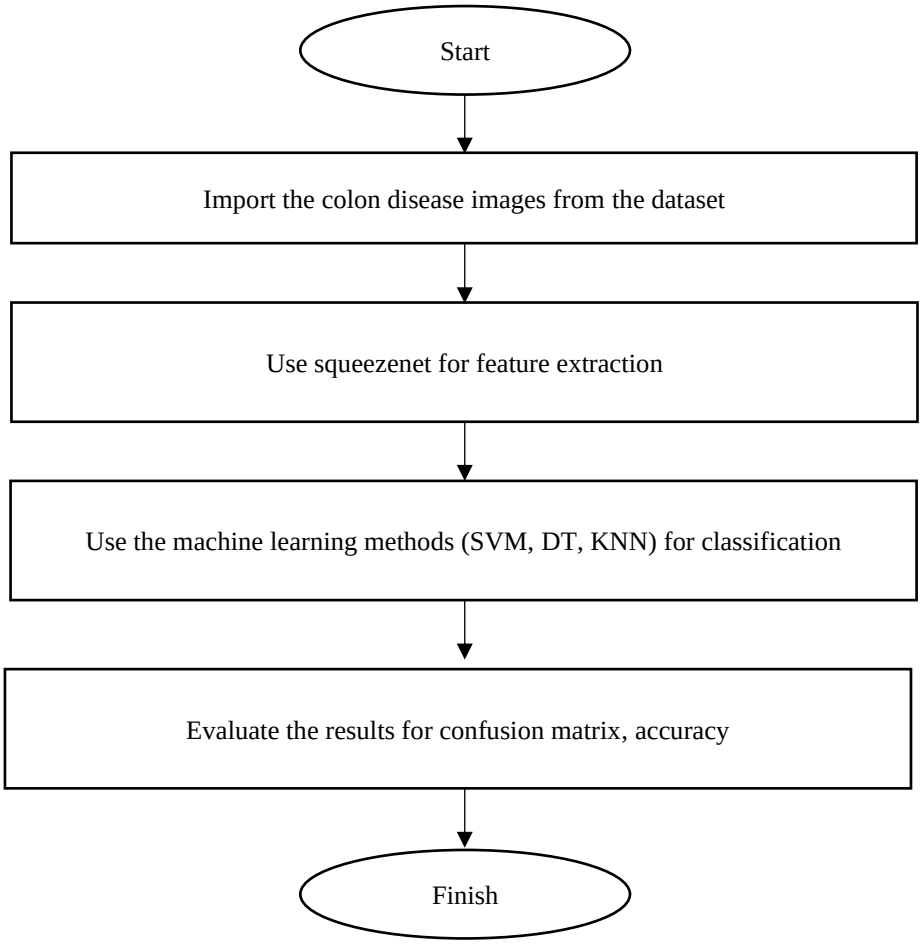

Figure 4. Flowchart of the proposed method for colon disease classification

RESULTS AND DISCUSSION

a) Dataset

In this paper, the famous dataset entitled "Lung and Colon Cancer Histopathological Images", obtained from the open-access dataset library will be used. The dataset was obtained from: (https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images)
This dataset contains 25,000 histopathological images with 5 classes. All images are 768 x 768 pixels in size and are in JPEG file format. In this paper, the last two datasets will use, Colon adenocarcinoma and the Colon benign tissue. These datasets depend on the colon histopathological images. In this paper, a total of 10,000 images will be used. 5000 images for cancer and 5000 images for non-cancer images. Visual examples of six histopathological images from the used dataset (where colon_n_refers to normal image and colon_ca_refers to an image with colon cancer).

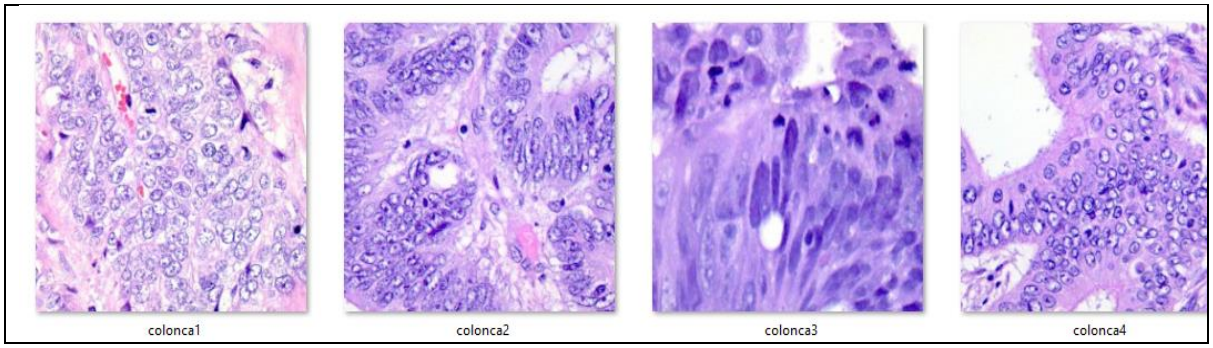

Figure 5. Some samples images from the dataset

To evaluate our results, we will use the most used measurements like Accuracy which based on confusion matrix is shown in Table 1.

Table 2: Confusion matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Support vector machines, decision trees and k-nearest neighbour have been utilized in this study to categorize colon illnesses. The suggested approach has been applied to squeezenet combined with these supervised machine learning methods as the classification of colon cancer to get the best accuracy.

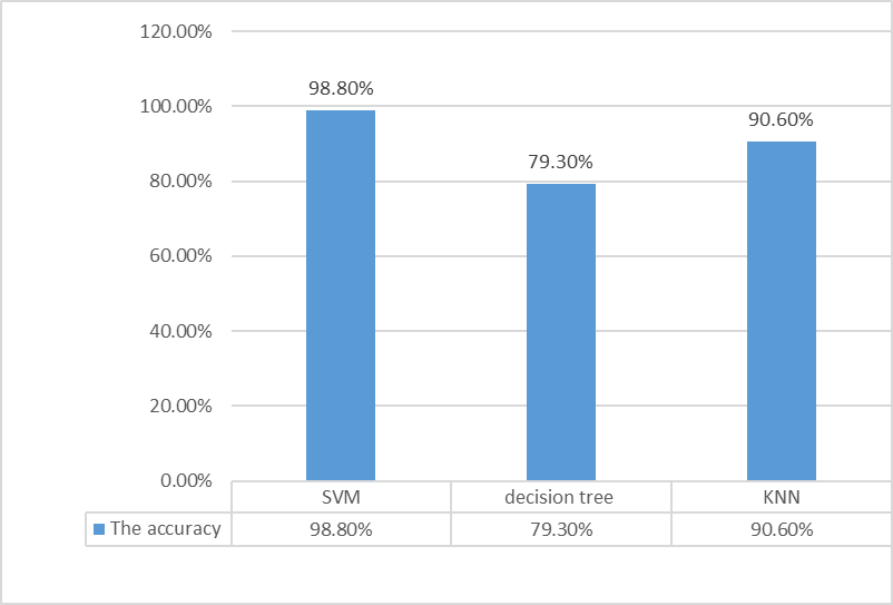

Figure 6. Mean of the accuracy index of SVM, decision tree and KNN methods based on squeezenet

The Analysis and evaluation reveal that the SqueezeNet combined by SVM approach performs quite well in terms of classification of image accuracy as it has been shown in the figure 3. In this experiment, the SVM based on SqueezeNet's accuracy is 98.80%. When the Decision-tree' experiment resulted average is 79.30% while the KNN method achieve 90.60% accuracy based on our CNN method.

CONCLUSION

In this study, a free-access database set was used for our necessary data. Using Matlab-2022a program for simulation, using different methods to reach a good accuracy in this work, we have combined machine learning based on the neural network. The machine learning algorithms SVM, decision tree and KNN have been applied to detect a colon cancer with a high accuracy. The best performance is obtained from SVM machine learning method based on the SqueezeNet pertaining neural network for colon cancer classification.

REFERENCES

[1] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, 2016, pp. 1–4.

[2] J. Alam, S. Alam, and A. Hossan, "Multi-stage lung cancer detection and prediction using multi-class svm classifie," in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, 2018, pp. 1–4.

[3] Y. J. Tan, K. S. Sim, and F. F. Ting, "Breast cancer detection using convolutional neural networks for mammogram imaging system," in *2017 International Conference on Robotics, Automation and Sciences (ICORAS)*, 2017, pp. 1–5.

[4] A. M. Godkhindi and R. M. Gowda, "Automated detection of polyps in CT colonography images using deep learning algorithms in colon cancer diagnosis," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 2017, pp. 1722–1728.

[5] T.-C. Chiang, Y.-S. Huang, R.-T. Chen, C.-S. Huang, and R.-F. Chang, "Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation," *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 240–249, 2018.

[6] S. J. Leavesley *et al.*, "Hyperspectral imaging fluorescence excitation scanning for colon cancer detection," *J. Biomed. Opt.*, vol. 21, no. 10, p. 104003, 2016.

[7] S. Rathore, M. Hussain, A. Ali, and A. Khan, "A recent survey on colon cancer detection techniques," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 10, no. 3, pp. 545–563, 2013.

[8] G. Lu, L. V Halig, D. Wang, X. Qin, Z. G. Chen, and B. Fei, "Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging," *J. Biomed. Opt.*, vol. 19, no. 10, p. 106004, 2014.

[9] R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, "A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 653–663, 2015.

[10] A. Gopi and C. S. Reshmi, "A noninvasive cancer detection using hyperspectral images," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017, pp. 2051–2055.

[11] B. Kosko, "Unsupervised learning in noise," *IEEE Trans. Neural Networks*, vol. 1, no. 1, pp. 44–57, 1990.