Sarifah - Deep Learning for Histopathological Image Analysis: A Convolutional Neural Network Approach to Colon Cancer Classification

by Sarifah Agustiani

Submission date: 20-Dec-2023 06:36PM (UTC+0700)
Submission ID: 2257699252
File name: ional_Neural_Network_Approach_to_Colon_Cancer_Classification.doc (1.04M)
Word count: 4939
Character count: 25855

Telematika – Vol. xx, No. xx, xxxxxx (2020) pp. xx-xx e-ISSN 2442-4528 | p-ISSN 1979-925X

Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA "2" Kemenristek/BRIN, No. 85/M/KPT/2020

UNIVERSITAS AMIKOM PURWOKERTO

Deep Learning for Histopathological Image Analysis: A Convolutional Neural Network Approach to Colon Cancer Classification

Sarifah Agustiani[1], Yan Rianto[2]

[1,2] Computer Science, Information Technology
Universitas Nusa Mandiri
E-mail:14002398@nusamandiri.ac.id[1], yanr001@gmail.com[2]

ARTICLE INFO

History of the article:
Received January 1, 2019
Revised January 21, 2019
Accepted January 12, 2019

Keywords:
Colon Cancer,
Histopathology Image,
CNN

Correspondece:
Telepon: +62 89646363255
Email:
14002398@nusamandiri.ac.id

ABSTRACT

Colon cancer is a cancer that attacks the colon or the last part of the human digestive system. The increase in cancer is influenced by the environment and lifestyle, especially diet, low-fiber food consumption and high consumption of animal proteins and fats also affect the cause of this cancer. Common symptoms can usually be seen from changes in bowel habits, abdominal pain, bleeding on the rectum and others. However not all symptoms are seen in every patient. What's more, it is difficult to determine the presence of cancer, as it must use thorough diagnostic procedures such as CT scans, MRI, PET scans, ultrasounds, or biopsies. But it also cannot be analyzed easily, it takes enough time and experts to be able to know the presence of the cancer. Therefore, this study aims to classify colon cancer based on histopathological imagery using CNN method which is one of the implementation of deep learning for digital image processing. In this study, the architecture used is pure architecture that is self-modified. The results of this study showed that the model was able to classify colon cancer with a fairly high accuracy of 99.55%

INTRODUCTION

Cancer is a non-communicable disease characterized by the growth of abnormal cells that are not controlled and are able to move and attack between cells in body tissues. The World Health Organization lists that cancer is one of the main causes of death in the world (Pangribowo, 2019).

Colon cancer is cancer that attacks the last part of the human digestive system or commonly called the large intestine. 90% of patients are elderly people over 60 years, although this disease can also attack any age (Rambe, 2019). Based on data from the *Global Burden of Cancer* (GLOBOCAN) released by the World Health Organization (WHO) stated that in 2020 colon cancer was ranked fourth with the number of cases 1,148,515 cases with a death rate reaching 576,858 people or equivalent to 5.8% (Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, 2020).

The increase in the incidence of colon cancer is influenced by the environment and lifestyle. Diet plays an important role and can lead to colon cancer. The level of consumption of low-fiber foods and high consumption of protein and animal fats are factors that influence the cause of colon cancer or colon cancer (Puspitasari & Waluyo, 2021). General symptoms can usually be seen from changes in bowel habits, abdominal pain, bleeding in the rectum, and iron deficiency anemia, although other gastrointestinal diseases also have these symptoms (Puspitasari & Waluyo, 2021).

However, not all symptoms are seen in every patient. Moreover, it is difficult to determine the presence of cancer without thorough diagnostic procedures such as CT scan, MRI, PET scan, ultrasound, or Biopsy. In many cases, patients show little or no symptoms in the early stages, because they appear only after the tumor is large enough or the cancer has spread to surrounding tissues or organs, and sometimes it is too late.

For this reason, early detection is needed, which means the earlier a person is diagnosed, the more time the doctor will have to design a treatment for him, and the patient has a better chance of beating the disease. Early diagnosis and appropriate treatment are currently the only way to reduce the number of cancer deaths (*Cancer-Symptoms and Causes-Mayo Clinic*, 2021). This diagnosis can be made through X-rays, CT scans, through histopathological images from the results of the biopsy process in the colon tissue, which from these results the doctor or medical personnel can determine the type of cancer that affects the follow-up to be carried out in the treatment of the cancer. but to do the classification manually will take a relatively long time, so we need a technology or system that can help facilitate the process of classifying the cancer, one of which is the use of *Machin Learning*.

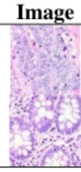
The existence of *Machine Learning* in the computer field has become a trend in information technology and has penetrated into various sectors including health (Sugara & Subekti, 2019). *Machine Learning* are used in the field of health has a variety of applications in pathology, ranging from detection of disease to an intelligent system, which is seen from the patient's symptoms, can prescribe drugs (Das et al., 2018). Thus, the detection results can be received more quickly by doctors / health experts, so that the action or handling of a disease will be faster (Sugara & Subekti, 2019).

In the application of *Machine Learning* technology, a good method or algorithm is needed so that it can predict accurately. Because the determination of the model and algorithm can affect the prediction results obtained, this study will propose a method using *Convolutional Neural Network* (CNN) which is one of the implementations of *Deep Learning* for digital image processing. CNN has become a popular method because it has been ranked first in the image classification category since 2012 (Felix et al., 2020). The research was conducted by examining the dataset obtained from the Kaggle repository with 10,000 images consisting of *Colon Adenocarcinoma* (colon aca) and *Colon Benign* (colon n) images (*Lung and Colon Cancer Histopathological Images | Kaggle*, n.d.). The results show that the model is able to classify colon cancer with high accuracy.

RESEARCH METHODS

The data used in this study is secondary data (public data) obtained from the *Kaggle Repository* known as the LC25000 dataset. This dataset, created by Andrew A. Borkowski and colleagues, and collected at the James A. Haley Veterans Hospital located in Tampa, Florida, has 10,000 images with 2 classes of colon cancer and each class has 5,000 images with 768 x 768 pixels in size and in jpeg file format. *The datasets were* generated from original samples from appropriate sources and validated by the *Health Insurance Portability and Accountability Act* (HIPAA).

Table 1. Colon Cancer Dataset

No	Image	Class	Amount
1		*Colon Adenocarcinomas*	5.000
2		*Colon Benign*	5.000
		Total Dataset	**10.000**

The research stages were carried out starting from data collection, image calling, pre-processing, *dataset* sharing, building CNN architecture, determining *hyperparameters*, conducting model training, model evaluation, and model implementation. This is done to get the right model in classifying colon cancer using histopathological images. The stages of the research can be seen in Figure 1.

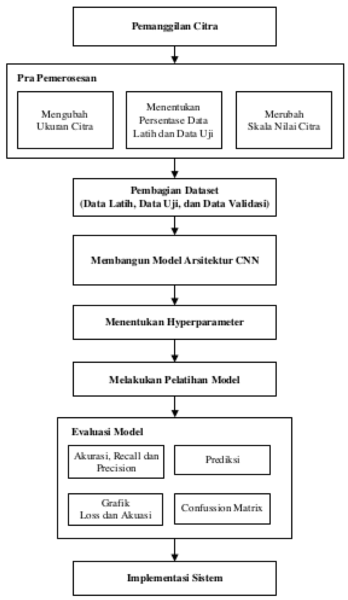


Figure 1. Research Stages

Based on the picture, the stages of the research can be explained as follows:

1. **Image Summon**

   The first step for calling the image is to determine the *dataset* storage directory to be used. where the *dataset* has 2 types of colon cancer classification, namely *Colon Adenocarcinoma* and *Colon Benign* . Then call all the *libraries* needed in building the architectural model, including: *Module OS* , *pandas* , *glob* , cv2, *numpy* , *matplotlib* , *sklearn* , *tensorflow* , *hard* , and others.

2. **Pre-Processing**

   At this stage it consists of three stages of process including changing the image size to 50x50 pixels, this is done so that the image is small so that it can speed up the image recognition process at the modeling stage, the next stage is to determine the percentage of training data and test data, in which the ratio used is 80:20, meaning that 80% is used for training data, and 20% for test data. Then change the scale of the image value, this serves to represent the image into *a* 3D *array* or also called a tensor which has a value range between 0-255.

3. Dataset Sharing

At this stage, the *dataset is* divided into training data, test data, and validation data, where the percentage of training data and test data distribution is in accordance with the percentage determined in the previous stage, which is 80:20. This is done to ensure that the classifier is generalized well, so that the data is divided into three categories, namely training data, test data, and validation data, each of which consists of 7950 training data samples, 2000 test data samples, and 50 samples of validation data taken. of the amount of training data.

4. CNN Model Building

In this stage, it is done by determining the number of architectural *layers* to be used starting from determining the value of the input layer which is the input of image data that is converted into a matrix with 3 dimensions where the value of each dimension is red, blue and green (Felix et al., 2020) the number of convolution layers It is the main part of CNN. The operations performed are the same as convolution operations which are usually performed in image processing, where there is a kernel and sub-images. The kernel used in CNN is generally 3x3 in size. Then for each sub-image that is the same size as the kernel, a convolution operation is performed (Nugroho et al., 2020), *pooling layers is the stage after the convolutional layer. The pooling layer consists of filters of a certain size and step. Each shift will be determined by the number of steps to be shifted across the feature map or activation map area* (Santoso & Ariyanto, 2018), *fully-connected layers is a multi-layer perceptron (MLP) classification process. In the fully connected layer, each neuron is fully connected to all activations in the previous layer* (Candra & Prapanca, 2020) and *output layers*. The steps involved in making the CNN model are described in more detail in Table 2 below:

Table 2. CNN Architecture

Layer		***Size and Filters***	***Kernel Size***	***Activation***	***Pooling size***
1	Image	50 x 50 x 3	-	-	-
	Convolution	16x16	2x2	*relu*	-
	Maxpooling	-	-	-	2x2
2	*Convolution*	32x32	2x2	*relu*	-
	Maxpooling	-	-	-	2x2
3	*Convolution*	64x64	2x2	*relu*	-
	Maxpooling	-	-	-	2x2
4	*Convolution*	128x128	2x2	*relu*	-
	Maxpooling	-	-	-	2x2
Flatten		-	-	-	-
Dropout		0,2			
Dense		200		*relu*	
Dropout		0,3			
Dense		500		*relu*	
Dropout		0,4			
Dense		2		*sigmoid*	

The research was conducted using four convolution layers with an image input of 50x50 pixels according to the size of the image that had been initiated earlier. All convolution operations on each layer use a 2x2 kernel size starting from the first layer to the last layer, but with different filter sizes for each layer starting from filter 16 for the first layer, *filter* 32 for the second layer, *filter* 64 for the third layer and *filter* 128 for the fourth layer. As for the use of activation at each convolution layer using ReLu activation and *pooling layer* using *max pooling* with a size of 2x2.

In addition, *zeropadding is* also applied to each convolution process, which is the process of adding zero values on each side of the input. After going through the convolution process, the final result of *max pooling* will be converted into a two-dimensional vector. The *hidden layer* uses *dense* with a *vector* number of 200, a dropout value of 0.2 along with a dense with a *vector* number of 500 and a dropout value of 0.3 and uses the activation of the ReLU function. And in the *output layer* there are *dense* as many as 2 according to the number of classes in the dataset and add *a dropout value* of 0.4 and a *sigmoid* activation function . The CNN model development architecture can be seen in the following figure:

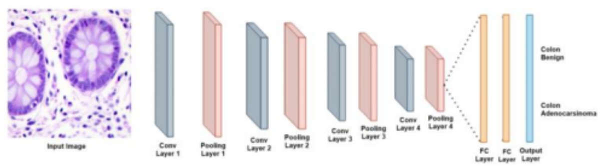


Figure 2. CNN Architecture Model

5. Determination of Hyperparameter

At this stage determine the *hyperparameters* for the architectural model that is built, but setting these parameters is done outside the construction of the CNN architectural model, such as determining the type of optimization, the type of probabilistic *loss* and the measurement of the evaluation model to be used.

6. Model Training

The model training stage of the architecture that has been built, in this stage determines the number of *epochs* to be used, where the number of *epochs* is a *hyperparameter* that determines how many times the learning algorithm will work through the entire training *dataset* . And determine the number of *batch size* which is the number of samples processed before the model is *updated* . The number of *epochs* used is 50 and the *batch size* used is 32.

7. Model Evaluation

The next stage is the model evaluation stage which functions to see the performance of the model that has been built, which can be seen from the accuracy value obtained, predictions for testing image samples based on the model, viewing graphs generated from iterations of model training, and *confusion matrix* tables .

8. System Implementation

The last stage is the implementation stage of the architectural model that has been built into a prediction system, where this system will be built using *flash web development* using the *Python* programming *language*

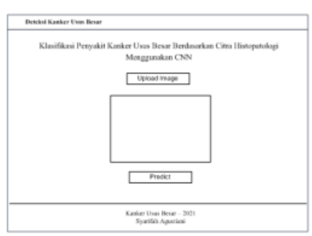


Figure 3. System Interface Design

RESULTS AND DISCUSSION

This study aims to classify colon cancer based on histopathological images into two classes, namely *Colon Adenocarcinoma* and *Colon Benign* using the CNN method with a modified architecture. In this study, several pilot scenarios were conducted to measure the success rate of the proposed method. Where this test is carried out by testing the use of the number of convolution layers, the use of optimization types, data sharing, as well as testing the *transfer learning methods,* namely *MobileNet, DensNet* , VGG16, and *ResNet50* . The following are the results of the tests that have been carried out in this study:

1. Convolution Layer Number Test Results

In this test, it is done by testing the number of convolution layers used in the test model. This is done to compare the accuracy values when using 1 layer, 2 layer, 3 layer, 4 layer, and 5 layer sequentially for the proposed CNN architecture. The results of this test can be seen in Table 4.6 as follows:

Table 3. Convolution Layer Number Test Results

No	Number of Convolution Layers	Accurasy
1	1	0.4964999854564667
2	2	0.9670000076293945
3	3	0.9890000224113464
4	**4**	**0.9955000281333923**
5	5	0.4964999854564667

Table 3 shows that the highest accuracy value resulted from the use of 4 convolution layers with an accuracy value of 99.55%. from the table it can also be seen that the use of too few or too many layers affects the accuracy value produced, especially in the tests applied to the case of this study. This can be seen from the accuracy results obtained by testing using 1 convolution layer and 5 convolution layers which produce the same accuracy value of 49.64%. the accuracy values generated from the two tests have a range of values that are quite far from the accuracy values produced by tests using 2, 3, or 4 convolution layers which are quite stable in the range of 90% and above.

2. Optimization Type Test Results

In this scenario, it is done by testing the type of optimization used in the test model that produces the best accuracy value in the previous test scenario, namely a model with 4 convolution layers according to the proposed model. This is done to compare the accuracy values when using *adam, nadam, adadelta, adagrad, adamax, RMSprop* , and SGD optimizations sequentially for the proposed CNN architecture, in this test the only difference is the use of the optimization type. It aims to see the effect of the use of optimization. The following are the results of testing on models with different types of optimization.

Table 4. Optimization Type Test Results

No	Optimization Type	Accuracy
1	**Adam**	**0.9955000281333923**
2	RMSProp	0.9940000176429749
3	Nadam	0.9940000176429749
4	Adamax	0.9919999837875366

5	SGD	0.9769999980926514
6	Adagrad	0.7070000171661377
7	Adadelta	0.5034999847412109

Table 4 shows that the use of *Adam's* optimization produces the highest accuracy value compared to other optimizations, although other optimizations also produce no less high accuracy, namely in the range of 97-99% except for *Adagrad* optimization with an accuracy of 70.70% and *Adadelta* which produces the lowest accuracy. that is 50.34%.

3. Test Results Percentage of Data Sharing

In this scenario, testing is still carried out using the proposed model, only changes are made during the distribution of training data and test data, which are carried out three times with different data comparisons ranging from 9:1, 8:2 and 7 ratios. :3, this is done to see the performance of the proposed model, which data comparison produces the highest accuracy value. The following are the results of testing the model with different data comparisons.

Table 5. Test Results of the Percentage of Data Sharing

No	Data Percentage	Accuracy
1	9:1	0.9909999966621399
2	**8:2**	**0.9955000281333923**
3	7:3	0.9950000047683716

Table 5 shows that the difference in the percentage of data sharing can produce different accuracy values, in this scenario the use of 8:2 produces the highest accuracy value of 99.55% with further accuracy obtained from a comparison of 7:3 with an accuracy value of 99.50%, and a comparison of 9:1 produces an accuracy value of 99.09%.

4. Transfer Learning Method Test Results

At this stage, a trial using the *transfer learning* method was conducted, which aims to determine the best accuracy between the proposed method and the *transfer learning* method . *Transfer learning* architecture methods used include *Resnet50,* VGG16, *Densnet121,* and *Mobilenet* . The results of the experiments carried out are as follows:

Table 6. Test Results of *Transfer Learning* Method

No	Transfer Learning Method	Accuracy
1	VGG16	0.9725000262260437
2	DensNet	0.9559999704360962
3	MobileNet	0.926999986171722
4	Resnet	0.925000011920929
5	**Proposed Method**	**0.9955000281333923**

Based on table 6 shows that the architecture of the proposed method has a higher accuracy value of 99.55% compared to the transfer learning method. In this test, the highest accuracy produced by the *transfer learning* method was obtained by the VGG16 method with an accuracy of 97.25%, and the lowest accuracy was obtained from the ResNet method with an accuracy of 92.50%.

5. Model Evaluation

The results of the research conducted with several test scenarios that have been carried out starting from comparing the accuracy values based on the number of convolution layers used, the type of optimization, the percentage of data sharing to making comparisons between the proposed method, namely CNN non- *transfer learning* and the *transfer learning* method. Based on the results obtained, it appears that the best results are obtained using the proposed method with an accuracy value of 99.55%.

To see the performance of the model that has been built, it can be seen from the accuracy value obtained, predictions for testing image samples based on the model, viewing graphs generated from iterations of model training, and *confusion matrix* tables.

a. Model Evaluation with Predicted Results

Besides being able to be evaluated based on the accuracy value, the performance of the proposed CNN method can also be seen from the results of image prediction, in this case the proposed method can predict the image well as shown in Figure 4. below.

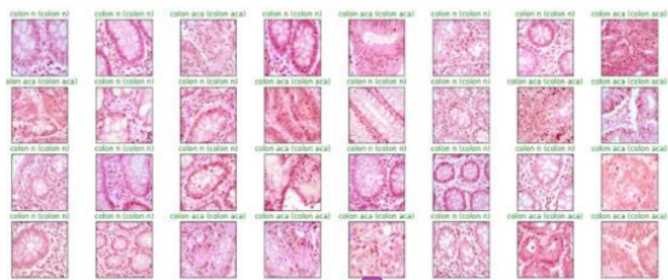


Figure 4. Prediction Results of Colon Cancer Image

Figure 4 shows some of the predictions generated by the CNN model proposed in this study. Based on the picture, information is given to the types of colon cancer, namely colon aca for the colon adenocarcinoma class, and colon n for the benign colon class, where the green color indicates that the predictions are correct and correspond to the actual values, while the incorrect predictions are marked in red. Overall the proposed model predicts the image correctly, this is evident from the prediction shown in green which means the prediction is correct.

b. Model Evaluation with Loss Graph and Accuracy

The following is a graph of accuracy, val accuracy, loss and val loss generated from the proposed method.

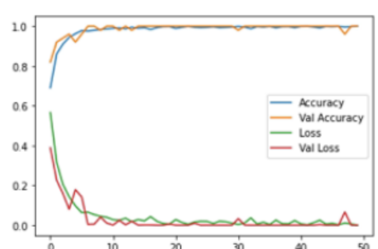


Figure 5. Graphics of the Proposed Method

Figure 5 displays a graph of the accuracy value generated for each epoch. Based on the graph, it can be concluded that the accuracy value of each epoch in the average training data has increased quite stable starting from the first epoch with an accuracy value of 0.6907 and continues to increase until the last epoch where the 50th epoch produces a high accuracy value of 0.9999, This means that the greater the accuracy value, the better the proposed model. Likewise with the validation data which has increased or decreased which is quite stable.

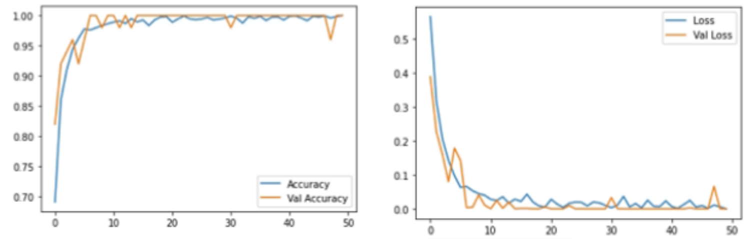


Figure 6. Accuracy Graph and Loss Graphs

The loss value in the test data produced in this study shows that it is decreasing where the first loss value is 0.5656 to 0.0058 towards the last epoch, the smaller loss value indicates a low error rate, in this case the loss value in the validation data has increased or decreased. the decline at certain epochs is quite stable.

c. Evaluate Model with Confusion Matrix

Other evaluation results are shown by the confusion matrix table, where in this table shows the number of positive predictive values that turn out to be true positive values, this is called true positive (TP), and the number of negative predictive values that turn out to be true negative values or is called true negative (TN). as well as the number of positive predictions but the actual result is negative or false positive (FP) in addition, it shows the number of negative predictions, but the actual value is positive or called false negative (FN). The following is a table of confusion matrix generated in this study.

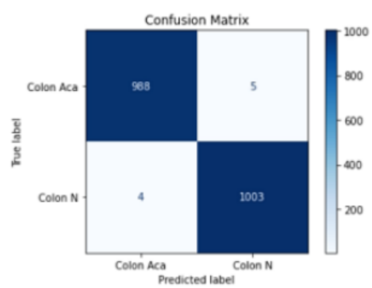


Figure 7 Confusion Matrix Value

Based on the results of the confusion matrix in Figure 7, it can be seen from the amount of test data used as many as 2000 images, from 1007 images for the colon n class, 1003 images were correctly detected and 4 other images were detected by the aca colon, while for the colon aca class, 988 images were correctly detected. and detected colon n as many as 5 images.

d. Evaluation of the Proposed Model with Previous Research

This evaluation was conducted to see the level of comparison between the results obtained from the proposed method and several colon cancer classification methods that have been discussed previously. However, most of these results cannot be directly compared because several studies were carried out using different images, but the purpose remains the same so that the research can be used as a comparison. Several previous studies were conducted using the type of colonoscopy imagery, but there were also studies that used the same dataset but with different methods, and there were also studies that used the same dataset and even the same method. Comparison of the accuracy values generated from previous studies with the proposed method can be seen in more detail in table 4.10 as follows:

Table 7 Comparison of Accuracy Values

No	Research	Image Type	Method	Accurasy
1	(Babu et al., 2018)	*Histopathological*	RF	85,30%
2	(Akbari et al., 2018)	*Colonoscopy*	CNN	90,28%
3	(Yuan et al., 2017)	*Colonoscopy*	AlexNet	91,47%
4	(Usama et al., 2020)	*Histopathological*	RESNET-50	93,91%
5	(Urban et al., 2018)	*Colonoscopy*	CNN	96,40%
6	(Mangal et al., 2020)	*Histopathological*	CNN	96,61%
7	**Proposed Method**	***Histopathological***	**CNN**	**99.55%**

Table 7 shows that the proposed method produces a higher accuracy value than the accuracy in previous studies. This is due to differences in the model development on the CNN architecture, in previous studies using 3 convolution layers, with a 150x150 image resize with one FC layer of 512 neurons and the output layer according to the number of classes. Meanwhile, in this study, the CNN model was built using 4 convolution layers, with 2 FC layers having values of 200 and 500 neurons, and the output layer according to the number of classes. So that the resulting accuracy results are different even though using the same dataset and method used.

6. System Implementation

From the results of tests conducted on research with the proposed CNN architecture, obtained a model that can be used to be implemented into a program system. Programs created using the Python programming language with the Flask Python Web Development framework. The test is carried out using one of the images in the colon adenocarcinoma class, then click predict to see the prediction results generated from the system that has been created. The results of testing the implementation of this system are carried out to compare the results of the existing image types with the predictions obtained from the implementation of the model that has been made. System implementation shows the following results:

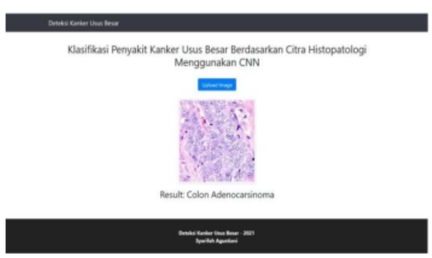


Figure 8. Test Results for Colon Adenocarcinoma Class

Figure 8 shows the results obtained from testing the system. Based on these tests, it can be seen that the predictions generated by the system are in accordance with the type of image from the class that is used as an example case. With the implementation of this system, it is hoped that it can help doctors or health analysts to diagnose colon cancer based on histopathological images with a faster and more accurate time.

CONCLUSIONS AND RECOMMENDATIONS

Based on the results of research conducted regarding the Classification of Colon Cancer Based on Histopathological Image Using the Convolutional Neural Network Method, several conclusions can be drawn in this study, namely as follows:

1. Modelling with the CNN method uses a modified architecture, namely the use of 4 convolution layers with the filter size in the first layer containing 16 filters, followed by 32, 64, to 128 for the second, third and fourth layers, kernel size 2x2 and 2 flatten layers providing performance which is good with the resulting accuracy value is quite high compared to the use of 1 layer, 2 layer, 3 layer or 5 layer convolution.
2. The application of hyperparameters in the proposed CNN modelling can give good results where in this study using Adam optimization, the type of loss binary cross entropy gives good results compared to the optimization of Nadam, Adamax, RMSprop, SGD, Adadelta and Adagrad.
3. The experimental results of the application of epoch 50 with a batch size of 32 and an image resolution size of 50x50 showed good performance as evidenced by the increase in the accuracy value displayed.
4. The proposed CNN method can be used for colon cancer classification and the implementation of the proposed CNN model into a web-based system can classify these types of cancer correctly.

ACKNOWLEDGEMENT

Thank you to all those who helped correct and publish this article, especially to the Nusa Mandiri university.

REFERENCES/BIBLIOGRAPHY

Akbari, M., Mohrekesh, M., Rafiei, S., Reza Soroushmehr, S. M., Karimi, N., Samavi, S., & Najarian, K. (2018). Classification of Informative Frames in Colonoscopy Videos Using Convolutional Neural Networks with Binarized Weights. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2018-July*, 65–68. https://doi.org/10.1109/EMBC.2018.8512226

Babu, T., Gupta, D., Singh, T., & Hameed, S. (2018). Colon Cancer Prediction on Different Magnified Colon Biopsy Images. *2018 10th International Conference on Advanced Computing, ICoAC 2018*, 277–280. https://doi.org/10.1109/ICoAC44903.2018.8939067

Cancer-Symptoms and Causes-Mayo Clinic. (2021). Mayo Foundation for Medical Education and Research (MFMER).https://www.mayoclinic.org/diseases-conditions/cancer/symptoms-causes/syc-20370588

Candra, P. N., & Prapanca, A. (2020). Klasifikasi gambar asli dan manipulasi menggunakan Error Level Analysis (ELA) sebagai proses komputasi metode Convolutional Neural Network (CNN).

Journal of Informatics and Computer Science, *02*, 9–18.

Das, S., Biswas, S., Paul, A., & Dey, A. (2018). AI Doctor: An intelligent approach for medical diagnosis. *Lecture Notes in Networks and Systems*, *11*, 173–183. https://doi.org/10.1007/978-981-10-3953-9_17

Felix, F., Wijaya, J., Sutra, S. P., Kosasih, P. W., & Sirait, P. (2020). Implementasi Convolutional Neural Network Untuk Identifikasi Jenis Tanaman Melalui Daun. *Jurnal SIFO Mikroskil*, *21*(1), 1–10.

Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, B. F. . (2020). International Agency for Research on Cancer 2020. *Global Cancer Observatory: Cancer Today.*, *419*, 1–2. https://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf

Lung and Colon Cancer Histopathological Images | Kaggle. (n.d.). Retrieved March 3, 2021, from https://www.kaggle.com/andrewmvd/lung-and-colon-cancer-histopathological-images

Mangal, S., Chaurasia, A., & Khajanchi, A. (2020). *Convolution Neural Networks for diagnosing colon and lung cancer histopathological images*.

Nugroho, P. A., Fenriana, I., Arijanto, R., & Kom, M. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia. *Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia*, *2*(1).

Pangribowo, S. (2019). Beban Kanker di Indonesia. *Pusat Data Dan Informasi Kemeterian Kesehatan RI*, 1–16.

Puspitasari, M. R., & Waluyo, A. (2021). Aplikasi Teori Katarine Kolbaca pada Kasus Kanker Kolon Pasca Laparatomi dengan Musik. *Jurnal Medika Husada*, *Vol 4 No.1*, Hal 1-6. http://ejournal.stikesadvaita.ac.id/index.php/MedikaUsada/article/view/86

Rambe, R. (2019). Sistem Pakar Mendiagnosa Penyakit Kanker Usus Besar pada Manusia dengan Menerapkan Metode Hybrid Case Based. *Jurnal Riset Komputer*, *6*(6), 606–611.

Santoso, A., & Ariyanto, G. (2018). Implementasi Deep Learning Berbasis Keras Untuk Pengenalan Wajah. *Emitor: Jurnal Teknik Elektro*, *18*(01), 15–21. https://doi.org/10.23917/emitor.v18i01.6235

Sugara, B., & Subekti, A. (2019). Penerapan Support Vector Machine (Svm) Pada Small Dataset Untuk Deteksi Dini Gangguan Autisme. *Jurnal Pilar Nusa Mandiri*, *15*(2), 177–182. https://doi.org/10.33480/pilar.v15i2.649

Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F., Karnes, W., & Baldi, P. (2018). Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology*, *155*(4), 1069-1078.e8. https://doi.org/10.1053/j.gastro.2018.06.037

Usama, S., Bukhari, K., Syed, A., Arsalan, S. K., Hussain, S. S., Armaghan, S. U., & Hus-, S. S. (2020). *The Histological Diagnosis of Colonic Adenocarcinoma by Applying Partial Self Supervised Learning Keywords:*

Yuan, Z., IzadyYazdanabadi, M., Mokkapati, D., Panvalkar, R., Shin, J. Y., Tajbakhsh, N., Gurudu, S., & Liang, J. (2017). Automatic polyp detection in colonoscopy videos. *Medical Imaging 2017: Image Processing*, *10133*, 101332K. https://doi.org/10.1117/12.2254671

Sarifah - Deep Learning for Histopathological Image Analysis: A Convolutional Neural Network Approach to Colon Cancer Classification

ORIGINALITY REPORT

15%	%	15%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics", Springer Science and Business Media LLC, 2021 Publication	2%
2	Mehedi Masud, Niloy Sikder, Abdullah-Al Nahid, Anupam Kumar Bairagi, Mohammed A. AlZain. "A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework", Sensors, 2021 Publication	2%
3	Marsani Asfi, Maria Ulva, Indah Triyani. "Penerapan Metode Shuffle Random pada Aplikasi Sistem Penentuan Playlist Lagu", Journal of Practical Computer Science, 2023 Publication	2%

4	Xing Xu, Jie Li, Yang Yang, Fumin Shen. "Towards Effective Intrusion Detection Using Log-cosh Conditional Variational AutoEncoder", IEEE Internet of Things Journal, 2020 Publication	1%
5	Prashant Narayan Sholapur, Indiramma M. "Explainable AI and Deep Learning techniques for Colon Cancer Detection", 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2022 Publication	1%
6	Aya Hage Chehade, Nassib Abdallah, Jean-Marie Marion, Mohamad Oueidat, Pierre Chauvet. "Lung and colon cancer classification using medical imaging: a feature engineering approach", Physical and Engineering Sciences in Medicine, 2022 Publication	1%
7	Shirin Hui Tan, Catherina Anak Guan, Mohamad Adam Bujang, Wei Hong Lai, Pei Jye Voon, Edmund Ui-Hang Sim. "Identification of phenomic data in the pathogenesis of cancers of the gastrointestinal (GI) tract: A UK biobank data analysis", Research Square Platform LLC, 2023	1%

Publication

8	Rinandha Puspadhani, Tuti Purwaningsih, Ayundyah Kesumawati, Arum Handini P., R. B. Fajriya Hakim. "Onions classification automation using deep learning with convolutional neural network method", AIP Publishing, 2021 Publication	1 %
9	Amna Ali A. Mohamed, Aybaba Hançerlioğullari, Javad Rahebi, Mayukh K. Ray, Sudipta Roy. "Colon Disease Diagnosis with Convolutional Neural Network and Grasshopper Optimization Algorithm", Diagnostics, 2023 Publication	1 %
10	Dewi Putrie Lestari, Rifki Kosasih, Tri Handhika, Murni, Ilmiyati Sari, Achmad Fahrurozi. "Fire Hotspots Detection System on CCTV Videos Using You Only Look Once (YOLO) Method and Tiny YOLO Model for High Buildings Evacuation", 2019 2nd International Conference of Computer and Informatics Engineering (IC2IE), 2019 Publication	1 %
11	"Medical Image Understanding and Analysis", Springer Science and Business Media LLC, 2017 Publication	<1 %

12 Joel Adeyanju Adewuyi, Gabriel Babatunde Iwasokun, Arome Junior Gabriel. "Transformation technique for derivation of similarity scores for signatures", Iran Journal of Computer Science, 2022
Publication
<1 %

---

13 Rafi Surya, Elliana Gautama. "Cassava Leaf Disease Detection Using Convolutional Neural Networks", 2020 6th International Conference on Science in Information Technology (ICSITech), 2020
Publication
<1 %

---

14 X.-M. Zhu, H. Zabel. " Monte Carlo calculation of (111) ordering transition and surface X-ray intensities ", Acta Crystallographica Section A Foundations of Crystallography, 1990
Publication
<1 %

---

15 Kwabena Adu, Joojo Walker, Patrick Kwabena Mensah, Mighty Abra Ayidzoe, Michael Opoku, Samuel Boateng. "SqueezeCapsNet: enhancing capsule networks with squeezenet for holistic medical and complex images", Multimedia Tools and Applications, 2023
Publication
<1 %

---

16 Muhammad Fajar Azka Fadillah, Dewinda Julianensi Rumala, Mauridhi Hery Purnomo, I Ketut Eddy Purnama. "The Effect of Noisy and Blurry Data on Deep Learning: Application in
<1 %

Brain Image Classification", TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON), 2022

Publication

---

17 "Intelligent Information and Database Systems: Recent Developments", Springer Science and Business Media LLC, 2020 <1%

Publication

---

18 D F Anas, I Jaya, Nurjanah. "Design and implementation of fish freshness detection algorithm using deep learning", IOP Conference Series: Earth and Environmental Science, 2021 <1%

Publication

---

19 Dmitry Ulyanov, Dmitry Savelyev. "Development and research of learning algorithms for neural networks with reinforcement in the gaming industry", 2021 International Conference on Information Technology and Nanotechnology (ITNT), 2021 <1%

Publication

---

20 Yosua Alvin Adi Soetrisno, Sumardi Sumardi, Agung Nugroho, Jaka Windarta. "Custom Convolutional Neural Network Model Implementation for Mask Detection on Atlas Hilens Device", 2022 9th International Conference on Information Technology, <1%

Computer, and Electrical Engineering (ICITACEE), 2022
Publication

21	Kailai Sun, Xinwei Wang, Qianchuan Zhao. "A Comprehensive Review of AIoT-based Edge Devices and Lightweight Deployment", Institute of Electrical and Electronics Engineers (IEEE), 2022 Publication	<1 %
22	Rami DERBEL, Nobuo MISHIMA. "A Deep Learning Approach to Analyzing and Evaluating Historic Urban Landscapes", Research Square Platform LLC, 2023 Publication	<1 %
23	Steven Willian, Theresia Herlina Rochadiani, Thamrin Sofian. "Design of Batak Toba Script Recognition System Using Convolutional Neural Network Algorithm", Sinkron, 2023 Publication	<1 %

Exclude quotes	On	Exclude assignment template	On
Exclude bibliography	On	Exclude matches	< 10 words