khazanah informatika

Jurnal Ilmu Komputer dan Informatika

Classification of Colon Cancer Based on Hispathological Images using Adaptive Neuro Fuzzy Inference System (ANFIS)

Nur Hidayah [1], Alvin Nuralif Ramadanti [2], Dian C R Novitasari [3*]

[1, 2, 3] Department of Mathematic
UIN Sunan Ampel Surabaya
Surabaya, Indonesia
*diancrini@uinsby.ac.id

Abstract-Cancer is a disease that is widely known and suffered by people in various countries. One type of cancer classified as the third contributor to death is colon cancer, with a mortality rate of 9.4%. Colon cancer is cancer that attacks the large intestine or rectum. Classification of colon cancer promptly is necessary to carry out appropriate treatment to reduce the death rate from colon cancer. This study uses the ANFIS method to classify colon cancer with its texture analysis using GLRLM. In addition, the evaluation model used in this study is the K-fold cross-validation method. This research uses colon cancer histopathological image data, which is 10000 image data divided into 2 classes, namely 5000 benign class and 5000 adenocarcinoma class. The best result in this study is when using k = 5 at an orientation angle of 135°, the accuracy value is 85.57%, sensitivity is 91.72%, and specificity is 80.55%.

Keywords: ANFIS, classification, colon cancer, GLRLM, K-fold cross validation

Article info: *submitted February, 24, 2022, revised July, 27, 2022, accepted July, 24, 2023*

1. Introduction

A chronic disease is carried by people worldwide is cancer [1]. Cancer can occur because abnormal and uncontrolled cell growth can affect the performance of other body organs [2]. Cancer is the leading cause of death [3]. In 2020, there were 19.3 million new cancer cases with a death rate of 10 million people [4]. Colon cancer is often diagnosed with 10% of new issues with a mortality rate of 9.4%. Colon cancer is the third leading cause of death after breast cancer, accounting for 11.7% of new cases with a mortality rate of 6.9%, and lung cancer accounting for 11.4% and a mortality rate of 18% [4]. In 2020, in the United States, 147.950 people were confirmed to have colon cancer [5].

Patients with colon cancer experience abnormal cell growth in the lining of the colon or rectum [6]. Common symptoms of colon cancer are changes in bowel movements such as diarrhea or constipation, a full stomach, less stool, often feeling tired throughout the day [7]. Colon cancer can by lifestyle, diet, and smoking [8]. Research on colon cancer has been widely carried out, such as that conducted by Caecar et al. [9], Ito et al. [10], and T. Okamoto et al. [11].

One of the efforts to determine the type of cancer in the human body is the classification process. The form of technology application that can use for classification is with the help of digital images. The classification process using digital images requires texture analysis, and one method for this problem is GLRLM. GLRLM or Gray Level Run Length Matrix has several advantages to support research results. One of the advantages of GLRLM is that it can distinguish images that have fine and coarse textures [12]. Various studies have been carried out with GLRLM and can provide optimal results. Research conducted by D.C. Rini et al. to compare the effects of breast cancer classification using the GLRLM, Gray Level Difference Matrix (GLDM), and Gray Level Co-Occurrence Matrix (GLCM) methods. The best accuracy results in this study are the GLRLM feature extraction with a value of 93.9757% [13]. In addition, a survey by N. Ding et al. in predicting postoperative abdominal aortic aneurysm using GLDM, GLRLM, and GLCM extractions. The study yielded the best accuracy of 87.23% using GLRLM extraction [14]. In conducting this research, it is necessary to use a classification method used to classify the types of cancer, and one of the methods is ANFIS.

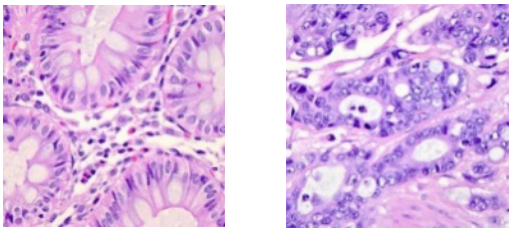
The ANFIS method has been used several times as an alternative to classifying a problem. In 2018, K. Jaya organized cervical cancer by applying ANFIS as its classification system. Khartiga produced a good result in this study because the sensitivity and accuracy were 97.42% and 99.36%, respectively [15]. In addition, another study conducted by Pravin R Kshirsagar, implemented ANFIS to detect and classify brain tumors. Pravin's research on the problem of brain tumors was carried out in 2020 using MRI images and was able to obtain high precision of 99.3% [16].

From several previous studies, this research implements the ANFIS method as an alternative classification with feature extraction using GLRLM based on histopathological image data of colon cancer. Colon cancer histopathological image data have been used in the study of Bukhari et al., which yielded the highest accuracy of 93.91% [17].

2. Theoritical Review

a. Colon Cancer

Colon cancer is one of the leading causes of health problems globally due to its high malignancy rate and the second leading cause of death from other types of cancer [18]. Colon cancer is a malignant cell that attacks the colon or rectum [19]. Cancer cells can generally extend through the intestinal wall to the lymph nodes and be carried by blood vessels to other organs around the intestine [20]. Based on the level of malignancy, colon cancer is divided into two classes, namely benign and adenocarcinoma. Benign cancer is benign cancer that occurs in the human digestive tract. At the same time, adenocarcinoma is the most common type of cancer that occurs in the mucosal epithelial tissue of the digestive tract from the large intestine to the rectum [21]. Benign cancer tissue and adenocarcinoma are shown in Figure 1.



(a) (b)

Figure 1. (a) Jaringan Kanker Benign; (b) Jaringan Kanker Adenocarcinoma

b. Gray Level Run Length Matrix (GLRLM)

GLRLM is a method that can use to distinguish between acceptable and coarse textures [22]. Galloway said that various types of statistical characteristics could be extracted based on the run-length matrix described in Table 1.

Table 1. GLRLM Statistical Features

Statistical Features	Formula
Short Run Emphasis (SRE)	$SRE = \frac{1}{n_r}\sum_{i=0}^{M}\sum_{j=0}^{N}\frac{p(i,j)}{j^2}$
Long Run Emphasis (LRE)	$LRE = \frac{1}{n_r}\sum_{i=0}^{M}\sum_{j=0}^{N} p(i,j)j^2$
Gray Level Non-uniformity (GLN)	$LN = \frac{1}{n_r}\sum_{i=0}^{M}\left(\sum_{j=0}^{N} p(i,j)j\right)^2$
Run Length Non-uniformity (RLN)	$RLN = \frac{1}{n_r}\sum_{i=0}^{M}\left(\sum_{j=0}^{N} p(i,j)j\right)$
Run Percentage (RP)	$RP = \frac{n_r}{\sum_{i=0}^{M}\left(\sum_{j=0}^{N} p(i,j)j^2\right)}$
Low Gray Level Run Emphasis (LGRE)	$LGRE = \frac{1}{n_r}\sum_{i=0}^{M}\left(\sum_{j=0}^{N}\frac{p(i,j)}{i^2}\right)$
High Gray Level Run Emphasis (HGRE)	$HGRE = \frac{1}{n_r}\sum_{i=0}^{M}\sum_{j=0}^{N} p(i,j)\, i^2$

Where:
I = histogram matrix Equalization
i = grey degree value
j = consecutive pixels (*run*)
M = the number of degrees of grey in the image
N = number of consecutive pixels in the image
r(j) = consecutive number of pixels by run length
g(i) = number of pixels by the degree of grey
s = the total number of run values generated in a given direction
p(i,j) = matrix entry
n = number of lines * number of columns

c. Principal Component Analysis (PCA)

PCA is a method that can simplify the complexity of high-dimensional data by maintaining data trends and patterns [23]. PCA can make data into smaller dimensions and allow for object visualization [24]. The first step in the PCA method is to calculate the mean and variance of the image, which is then used to calculate the covariance matrix. By obtaining the covariance matrix, it is continued with the analysis of the eigenvalue and eigenvector values. Furthermore, variables that have eigenvalues below 0.5 will be eliminated. The remaining variables will be searched for their eigenvector values and then transformed. [25].

d. K-fold Cross Validation

K-fold cross-validation is a method for evaluating the model by breaking the data into k parts/folds where each fold has the opportunity to become a testing set [26]. In the i data set, it becomes a testing set, and then the other data sets become a training set on classification [27]. Illustration of k-fold cross-validation will be shown in Figure 2.

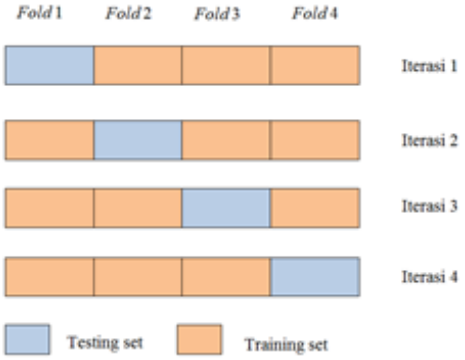


Figure 2. Illustration of k-fold cross-validation

e. Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS combines Fuzzy Logic (FL) and Neural Network (NN) methods. The fuzzy inference system uses a fuzzy inference system known as the first order Tagaki – Sugeno - Kang (TSK) model [28]. The neuro-fuzzy system contains five arrangements that have different functions. ANFIS is trained using a hybrid algorithm consisting of forwarding and backward steps.

In the next step, the premise parameters are set stable using the Recursive Least Square Estimator (RLSE) method by improving the consequent parameters based on the input and output data. Then, what will pass on the output data to the adaptive network to compare the output results and the actual data. Whereas in the backward step, set the consequent parameter to be stable, and the error value between the adaptive and actual network output will be propagated using gradient descent. In gradient descent, the premise parameters will be fixed. [29].

f. Confusion Matrix

The confusion matrix is a technique used to determine the performance results of an algorithm [30]. The confusion matrix is commonly used to calculate the results of the classification process [31]. There are 4 terms to calculate the results of the classification process, which are presented in Table 2.

Table 2. Confusion Matrix

Actual Class	Classification Result	
	Adenocarcinoma	**Benign**
Adenocarcinoma	True Positif (TP)	False Negatif (FN)
Benign	False Positif (FP)	True Negatif (TN)

- True Positive (TP): Number of patients with adenocarcinoma who were correctly classified as adenocarcinoma
- True Negative (TN): Number of benign patients classified as benign
- False Positive (FP): Number of patients with adenocarcinoma classified as benign
- False Negative (FN): Number of benign patients classified as adenocarcinoma

In this study, the success rate of the classification process is measured based on three parameters, namely sensitivity, specificity and accuracy, which can find using equations 1, 2, and 3.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + FN} \quad (3)$$

3. Methods

The method for the classification of colon cancer includes several stages. The first stage is to collect histopathological image data, consisting of 2 classes, namely benign and adenocarcinoma. This study used 10000 histopathological image data, divided into 5000 benign class data and 5000 adenocarcinoma class data. The steps of this research are presented in Figure 3.

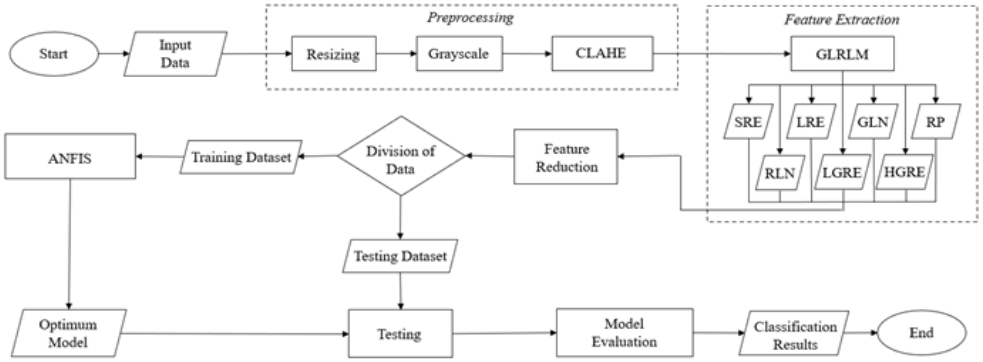
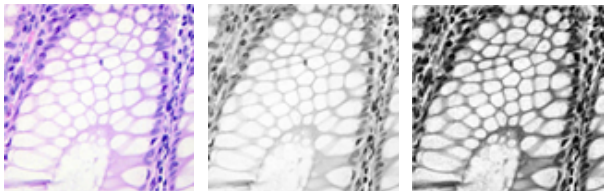


Figure 3. Research Flowchart

The second stage is preprocessing, which consists of a resizing stage that aims to change and resize the image to 300x300 pixels. Then grayscale is performed to display the grey level of the image and followed by the Contrast Limited Adaptive Histogram Equalization (CLAHE) process. After the preprocessing process has been carried out, it is continued with the feature extraction process using GLRLM at angles of 0°, 45°, 90°, and 135°. Furthermore, the feature extraction results will be reduced using the PCA method and then divide the data into data training and data testing by testing the kfold = 5 and 10 values. The next stage is the classification process carried out using the ANFIS method followed by model evaluation using the confusion matrix.

4. Results and Discussion

Classification of colon cancer is done by using histopathological image data on benign and adenocarcinoma classes. Some images are not the same size. Therefore, resizing is done on all data to produce a new length of 300 x 300. Next, what will convert the image to grayscale to reduce the scale of the image. After obtaining an image with a greyscale, the image contrast is enhanced using Adaptive Histogram Equalization. The results of the resizing process to increase the contrast are presented in Figure 4.



(a) (b) (c)

Figure 4. (a) Image Resizing Results; (b) Grayscale Process Results; (c) Contrast Enhancement Results After increasing the contrast of the image, feature extraction will be carried out using GLRLM with seven features, namely SRE, LRE, GLN, RP, RLN, LGRE, and HGRE with degrees 0°, 45°, 90°, and 135°. The feature extraction results in the benign class using GLRLM 0° are shown in Table 3.

Table 3. Feature Extraction Results using GLRLM

Data	SRE	LRE	GLN	RP	RLN	LGRE	HGRE
1	0.715091	6.05749	3859.344	0.570533	24312.22	125.3446	3859.344
2	0.764583	4.549788	4061.156	0.642389	31314.38	101.9458	4061.156
3	0.787527	3.653389	4138.749	0.678956	35191.18	97.08127	4138.749
4	0.796169	2.754085	4263.356	0.711944	37739.07	98.06859	4263.356
5	0.787337	2.99568	4321.206	0.699533	36232.51	101.003	4321.206
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
10000	0.777024	3.152989	4131.755	0.683056	34424.36	100.8773	4131.755

Table 4. Test Results on the Value of k = 5

Angle	Value	GLRLM					GLRLM - PCA				
		k = 1	k = 2	k = 3	k = 4	k = 5	k = 1	k = 2	k = 3	k = 4	k = 5
0	Accuracy	68.94	66.53	65.26	72.18	70.81	59.75	62.20	59.10	57.55	61.33
	Sensitivity	61.97	60.22	59.09	64.53	63.76	60.89	64.99	60.44	58.38	62.99
	Specificity	94.76	92.88	96.85	96.63	92.25	58.82	60.29	58.07	56.87	60.05
45	Accuracy	82.82	65.86	81.11	76.82	68.44	84.31	83.68	84.57	85.44	84.52
	Sensitivity	76.29	59.59	76.22	69.62	61.56	90.02	89.25	95.35	92.04	89.71
	Specificity	93.63	95.17	88.54	92.15	95.11	80.05	79.51	78.01	80.64	80.55
90	Accuracy	71.74	75.00	77.61	78.01	71.18	71.84	71.34	73.46	73.34	71.51
	Sensitivity	64.74	67.63	70.35	70.23	63.64	73.80	72.90	79.52	76.36	73.11
	Specificity	91.27	92.96	92.74	95.32	96.30	70.18	69.97	69.49	70.94	70.12
135	Accuracy	78.24	88.41	57.62	68.27	56.95	75.01	75.21	75.78	76.29	74.86
	Sensitivity	70.89	91.72	54.11	61.31	53.77	76.83	76.70	80.38	79.80	77.11
	Specificity	93.53	85.34	94.77	97.40	94.84	73.43	73.89	72.40	73.52	72.97

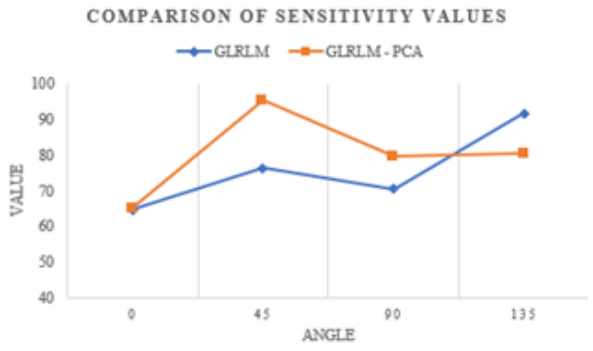
Table 5. Test Results on the Value of k = 10

Angle	Test	Value	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k=10
0	GLRLM	Accuracy	66.67	69.36	62.84	63.15	67.40	67.92	68.51	68.36	64.66	74.34
		Sensitivity	60.13	62.33	57.42	57.66	60.62	61.24	61.61	61.16	58.64	67.12
		Specificity	95.88	94.30	97.24	98.14	96.29	93.47	95.57	95.63	96.52	91.50
	GLRLM - PCA	Accuracy	60.60	59.80	57.70	58.30	59.30	59.30	63.10	60.20	56.30	63.06
		Sensitivity	61.78	61.40	59.10	59.20	60.89	60.64	66.17	61.33	57.31	64.25
		Specificity	59.64	58.60	56.67	57.56	58.12	58.26	61.01	59.27	55.54	62.06
45	GLRLM	Accuracy	83.99	88.42	70.97	84.96	65.90	55.40	64.69	84.38	64.76	63.73
		Sensitivity	81.20	88.87	63.85	80.07	59.66	52.88	58.92	81.08	58.84	58.01
		Specificity	87.85	87.95	93.36	91.79	94.92	93.55	90.22	88.76	94.05	97.90
	GLRLM - PCA	Accuracy	84.97	84.27	84.97	85.69	84.28	83.47	83.40	83.30	85.09	84.85
		Sensitivity	89.37	90.69	91.23	93.00	89.22	90.31	88.30	94.32	90.51	90.42
		Specificity	81.47	79.62	80.38	80.51	80.46	78.63	79.61	76.76	80.95	80.67
90	GLRLM	Accuracy	77.59	72.55	71.82	78.76	75.30	76.98	79.53	79.80	76.08	73.52
		Sensitivity	71.12	65.87	64.64	71.13	68.69	70.04	73.99	73.14	68.57	65.73
		Specificity	89.83	89.16	92.58	94.98	89.71	91.36	88.98	91.85	93.90	96.80
	GLRLM - PCA	Accuracy	71.80	71.60	72.04	73.40	72.04	72.17	71.14	72.92	73.57	71.27
		Sensitivity	72.06	74.11	75.52	76.71	73.55	76.37	71.88	77.89	76.17	73.25
		Specificity	71.54	69.57	69.42	70.82	70.73	69.14	70.46	69.49	71.45	69.61
135	GLRLM	Accuracy	83.55	81.81	86.90	72.00	59.16	68.70	70.34	66.07	75.68	77.38
		Sensitivity	77.74	76.82	83.58	64.51	55.03	62.42	63.02	59.93	67.93	69.97
		Specificity	92.69	89.11	90.99	95.45	98.94	87.85	95.95	91.71	95.09	93.61
	GLRLM - PCA	Accuracy	75.55	75.88	76.00	75.80	74.97	74.80	74.45	74.25	76.78	74.75
		Sensitivity	76.68	77.45	78.89	80.00	77.12	77.68	76.02	78.76	79.60	76.97
		Specificity	74.52	74.48	73.64	72.63	73.15	72.46	73.07	70.98	74.45	72.88

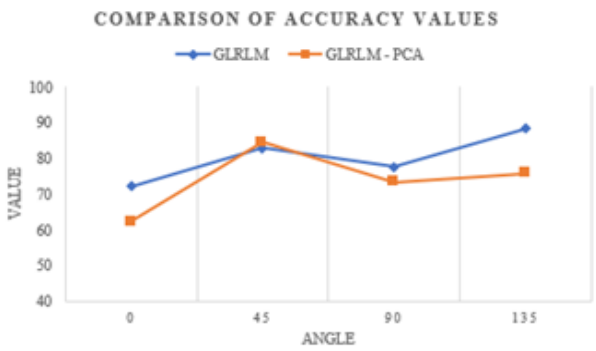
Calculations using GLRLM get different results. This shows a different pattern for each image. Furthermore, the value of the GLRLM features will be reduced using the PCA method.

The calculation results in the PCA process will be used as input for the classification process. Before classification using the ANFIS method, the PCA data were divided into training and testing data using the k-fold cross-validation method with parameters k = 5 and k = 10. The results of the classification showed that there was a difference in the accuracy of the two trials with the value of k. The best evaluation results were obtained using the trapezoidal membership function shown in Table 4 and 5.

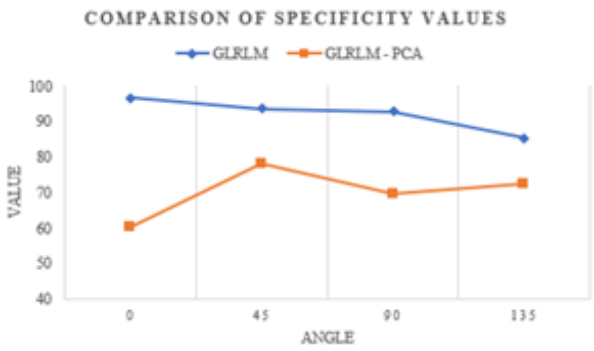
Based on the test results that have been presented in Table 4 and 5, the highest sensitivity value is obtained, namely in Table 4, which shows the importance of k = 5. This indicates that the optimal value of k to be used in data sharing is 5. The comparison of the best results in each angle at k = 5 is also shown in Figure 5.



(a)



(b)



(c)

Figure 5. (a) Sensitivity Value Comparison; (b) Comparison Value Accuracy; (c) Comparison of Specificity Value

Based on Figure 5, the best results were obtained when using GLRLM – PCA and at an angle of 45° with an accuracy value of 85.57%, sensitivity and specificity respectively 95.61% and 80.55%. What will show the best results in detail in the confusion matrix table presented in Table 5.

Table 5. Confusion Matrix Results

Actual Class	Classification Results	
	Adenocarcinoma	***Benign***
Adenocarcinoma	762	35
Benign	233	965

Based on the results in the confusion matrix presented in Table 4, the best accuracy, sensitivity and specificity values can be obtained using equations (2.1) to (2.3). Research with a similar case was also conducted by [32] using the Naive Bayes method and yielded the highest accuracy of 95.24%. These results indicate that the ANFIS method used in this study can provide better results.

5. Conclusion

From the classification system that has been made in this study, the highest sensitivity value is 95.61%. These results show numbers above 95%, which means the classification system created using the ANFIS method can recognize patterns in colon cancer well. Or in other words, what can use the ANFIS method to design a classification system for colon cancer. However, this research cannot be perfect because it is still used for classification in 2 classes of colon cancer. It will continue to be developed and refined using more complex diseases in the future.

References

[1] S. Dutta, S. Mahalanobish, S. Saha, S. Ghosh, and P. C. Sil, "Natural products: An upcoming therapeutic approach to cancer," *Food and Chemical Toxicology*, vol. 128, pp. 240–255, 2019.

[2] N. Kumar *et al.*, "Efficiency of Nanomaterials for Electrochemical Diagnostics based Point-of-Care Detection of Non-Invasive Oral Cancer Biomarkers," *Adv. Mater. Lett*, vol. 12, p. 21081651, 2021.

[3] S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world," *Journal of hepatology*, vol. 70, no. 1, pp. 151–171, 2019.

[4] H. Sung *et al.*, "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, 2021.

[5] G. Pellino and A. Spinelli, "How coronavirus disease 2019 outbreak is impacting colorectal cancer patients in Italy: a long shadow beyond infection," *Diseases of the Colon & Rectum*, vol. 63, no. 6, pp. 720–722, 2020.

[6] R. Zhou *et al.*, "Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I–III colon cancer," *Cancer immunology, immunotherapy*, vol. 68, no. 3, pp. 433–442, 2019.

[7] M. Sayuti and N. Nouva, "Kanker Kolorektal," *AVERROUS: Jurnal Kedokteran dan Kesehatan Malikussaleh*, vol. 5, no. 2, pp. 76–88, 2019.

[8] P. Pignatelli *et al.*, "The potential of colonic tumor tissue Fusobacterium nucleatum to predict staging and its interplay with oral abundance in colon cancer patients," *Cancers*, vol. 13, no. 5, p. 1032, 2021.

[9] N. K. C. Pratiwi, R. Magdalena, Y. N. Fuadah, and S. Saidah, "K-Nearest Neighbor for colon cancer identification," in *Journal of Physics: Conference Series*, 2019, vol. 1367, no. 1, p. 012023.

[10] N. Ito, H. Kawahira, H. Nakashima, M. Uesato, H. Miyauchi, and H. Matsubara, "Endoscopic diagnostic support system for cT1b colorectal cancer using deep learning," *Oncology*, vol. 96, no. 1, pp. 44–50, 2019.

[11] T. Okamoto *et al.*, "Feature extraction of colorectal endoscopic images for computer-aided diagnosis with CNN," in *2019 2nd International Symposium on Devices, Circuits and Systems (ISDCS)*, 2019, pp. 1–4.

[12] J. P. Sari, A. Erlansari, and E. P. Purwandari, "Identifikasi Citra Digital Kura-Kura Sumatera Dengan Perbandingan Ekstraksi Fitur GLCM Dan GLRLM Berbasis Web," *Pseudocode*, vol. 8, no. 1, pp. 66–75, 2021.

[13] D. C. R. Novitasari, A. Lubab, A. Sawiji, and A. H. Asyhar, "Application of feature extraction for breast cancer using one order statistic, GLCM, GLRLM, and GLDM," *Advances in Science, Technology and Engineering Systems Journal (ASTESJ)*,

vol. 4, no. 4, pp. 115–120, 2019.

[14] N. Ding *et al.*, "CT texture analysis predicts abdominal aortic aneurysm post-endovascular aortic aneurysm repair progression," *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.

[15] B. K. Jaya and S. S. Kumar, "Image registration based cervical cancer detection and segmentation using ANFIS classifier," *Asian Pacific Journal of Cancer Prevention: APJCP*, vol. 19, no. 11, p. 3203, 2018.

[16] P. R. Kshirsagar, A. D. Yadav, K. A. Joshi, P. Chippalkatti, and R. Y. Nerkar, "Classification and Detection of Brain Tumor by using GLCM Texture Feature and ANFIS," *Journal of Research in Image and Signal Processing*, vol. 5, no. 1, pp. 15–31, 2020.

[17] S. U. K. Bukhari, A. Syed, S. K. A. Bokhari, S. S. Hussain, S. U. Armaghan, and S. S. H. Shah, "The Histological Diagnosis of Colonic Adenocarcinoma by Applying Partial Self Supervised Learning," *med*R*xiv*, 2020.

[18] A. Yaghoubi, M. Khazaei, A. Avan, S. M. Hasanian, and S. Soleimanpour, "The bacterial instrument as a promising therapy for colon cancer," *International Journal of Colorectal Disease*, vol. 35, no. 4, pp. 595–606, 2020.

[19] N. Salmi and Z. Rustam, "Naïve Bayes classifier models for predicting the colon cancer," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 546, no. 5, p. 052068.

[20] M. A. A. AlAdlee, "Lipoxygenase and Colon Cancer," 2021.

[21] Ismail, "Klasifikasi Citra Histopatologi Kanker Usus Besar Menggunakan Arsitektur Alexnet," Bandung, 2020.

[22] M. Astiningrum, A. R. Syulistyo, and M. A. Zakariya, "Identifikasi Kualitas Biji Jagung Manis Layak Jual dari Warna dan Tekstur Menggunakan HSV dan Gray Level Run Length Matrix (GLRLM)," *JIP (Jurnal Informatika Polinema)*, vol. 7, no. 1, pp. 37–44, 2020.

[23] J. Lever, M. Krzywinski, and N. Altman, "Points of Significance: Principal Component Analysis," *Nature Methods*, vol. 14, no. 7, pp. 641–642, 2017, doi: 10.1038/nmeth.4346.

[24] P. Geladi and J. Linderholm, "Principal Component Analysis," 2020.

[25] D. L. Z. Astuti and Samsuryadi, "Kajian Pengenalan Ekspresi Wajah menggunakan Metode PCA dan CNN," *Annual Research Seminar (ARS) Fakultas Ilmu Komputer*, vol. 4, no. 1, pp. 293–297, 2018.

[26] I. D. Wijaya, A. G. Putrada, and D. Oktaria, "Penggunaan Metode K-Fold untuk Data Imbalance pada Klasfikasi HWE dan QPQ dalam Kejahatan Tweet Pelecehan Seksual," vol. 8, no. 5, pp. 10987–10996, 2021.

[27] H. Rhomadhona and J. Permadi, "Klasifikasi Berita Kriminal Menggunakan NaÃ¯ve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation," *Jurnal Sains dan Informatika*, vol. 5, no. 2, pp. 108–117, 2019, doi: 10.34128/jsi.v5i2.177.

[28] A. Kumar and V. K. Tripathi, "Adaptive Neuro Fuzzy Inference System for Runoff Modelling—A Case Study," *Int. J. Curr. Microbiol. App. Sci*, vol. 8, no. 4, pp. 2054–2061, 2019.

[29] F. M. Siregar, G. W. Nurcahyo, and S. Defit, "Prediksi Hasil Ujian Kompetensi Mahasiswa Program Profesi Dokter (UKMPPD) dengan Pendekatan ANFIS," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 2, no. 2, 2018.

[30] J. H. J. C. Ortega, A. C. Lagman, L. R. Q. Natividad, E. T. Bantug, M. R. Resureccion, and L. O. Manalo, "Analysis of Performance of Classification Algorithms in Mushroom Poisonous Detection using Confusion Matrix Analysis," *International Journal*, vol. 9, no. 1.3, 2020.

[31] N. Yang, Y. Qian, H. S. EL□ Mesery, R. Zhang, A. Wang, and J. Tang, "Rapid detection of rice disease using microscopy image identification based on the synergistic judgment of texture and shape features and decision tree–confusion matrix method," *Journal of the Science of Food and Agriculture*, vol. 99, no. 14, pp. 6589–6600, 2019.

[32] N. Salmi and Z. Rustam, "Naïve Bayes classifier models for predicting colon cancer," in *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, p. 052068, 2019.