



Classification of Covid-19 Using Feature Extraction GLCM and SVM Algorithm

Muhamad Saenudin¹, Fauzan Haq², Riza Ibnu Adam³

^{1,2,3} Informatics Engineering Study Program, Faculty of Computer Science, Singaperbangsa Karawang University, Karawang, 41361, Indonesia

E-mail : muhamad.saenudin17142@student.unsika.ac.id, fauzan.haq17096@student.unsika.ac.id , riza.adam@staff.unsika.ac.id

ARTICLE INFO

Article history:

Received: 01/04/2021

Revised: 20/04/2021

Accepted: 30/04/2021

Keywords:

Covid-19, GLCM, SVM

ABSTRACT

Coronavirus Disease (COVID-19) is a new variant of the corona virus that mutates and spreads rapidly between humans. The high rate of transmission and spread is not matched by the fast process of diagnosis because it has to go through a polymerase chain reaction (PCR) test in the laboratory. To identify quickly, efficiently and effectively, then a classification system is used using x-ray images of the chest using the Gray Level Co-occurrence Matrix (GLCM) method and the Support Vector Machine (SVM) algorithm. The data used are digital images of normal chest X-Ray (CXR) diagnosis and positive diagnosis of Covid-19 with 408 training data and 128 test data. The test is carried out on the GLCM parameters, namely with a distance of $d = 1, 2, 3$ and angles $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$ and feature extraction with contrast, correlation, energy, homogeneity, and dissimilarity. The test results show that the highest accuracy is at the distance $d = 1$ and the angle $\theta = 0^{\circ}$ is 90.47% and the lowest accuracy is at the distance $d = 3$ and the angle $\theta = 90^{\circ}$ is 80.35%.

Copyright © 2021 Jurnal Mantik.

All rights reserved.

1. Introduction

Coronavirus Disease (COVID-19) has become a global pandemic disaster that has spread to almost all countries in the world. Since it started spreading in December 2019 to date, it has caused more than 3 million people to die, more than 141 million cases have been confirmed positive and more than 120 million people have been declared[1]. The high level of spread has made it overwhelming for medical authorities to identify patients infected with Covid-19 due to various symptoms. It is difficult for medical personnel to perform fast medical action because the process of diagnosing Covid-19 has to go through laboratory tests and takes days. Speed in the process of diagnosing patients will be very important because the procedure for treating positive Covid-19 patients must isolate in a closed room so as not to infect people who are still healthy[2].

The use of Chest X-ray images in detecting covid-19 provides an opportunity to identify it faster and is also more cost-effective because it does not have to go through laboratory tests first[3]. In the image processing process, feature extraction using the Gray Level Co-occurrence Matrix (GLCM) method is used and then the output of the GLCM method enters the Support Vector Machine (SVM) algorithm model to be processed so that it can produce output from positive or negative diagnosis of Covid-19.

There are many studies that have used the Gray Level Co-occurrence Matrix (GLCM) method in the classification of pneumonia using the KNN method with parameters $K = 5$ and resulted in an accuracy of 66.20%.[4]. Then research on the use of the SVM method on Covid-19 detection with Chest X-Ray using Wavelet gave quite high results with an accuracy of 93.91%.[5].

[6]In research on recognizing facial expression images using feature extraction with the Gray Level Co-occurrence Matrix (GLCM) method with the parameters of contrast, correlation, energy and homogeneity feature extraction, the average accuracy rate is 33%, angry expressions 48%, happy expressions 73 %, bad mood expression 44%, fear expression 15%, sad expression 54%, and shocked expression 68%.

[7]In this study, it discusses the application of image histograms with the SVM method in the classification of breast disease by comparing the use of RBF, linear, quadratic and polynomial kernels in obtaining good accuracy. Based on testing the polynomial kernel has an accuracy of 90%.



2. Theoretical Review

2.1 Covid-19

Covid-19 is a new type of corona virus caused by infection with the SARS-CoV-2 virus which was initially thought to be transmitted from animals to humans and then can spread from human to human[8]. In the process of transmission of the corona virus, which is through tiny sparks or (droplets) that come out of the human respiratory tract when the sufferer coughs, sneezes, or talks and can be transmitted through the surface of objects touched by the sufferer, then other people touch hands, objects, eyes, nose or mouth[8]. There are some common symptoms that indicate a person is infected with the corona virus, namely chest pain, anisimia, vertigo, damage to lung tissue, dry cough, shortness of breath, chronic fatigue and easy forgetfulness[9].

2.2 Gray-Level Co-occurrence Matrix (GLCM)

Gray Level Co-occurrence Matrix (GLCM) is a feature extraction method in an image processing that works by calculating the proximity value between pixels that have a gray intensity and orientation from an angle. The direction or orientation of the angle, represented in degrees formed from the angles 0^0 , 45^0 , 90^0 , and 135^0 [10]. There are several feature extraction from the GLCM method, including: contrast, correlation, energy, homogeneity, and entropy, dissimilarity, and ASM. [11].The following are texture feature extraction parameters that are often used in the GLCM method, namely as follows[12]:

a. Contrast

This feature measures the contrast level in an image, which when the contrast is higher, it will increase the difference in the intensity of neighboring pixels.The value of contrast can be calculated using equation number 1.

$$\sum_{x,y} (x - y)^2 p(x, y) \quad (1)$$

b. Correlation

This feature provides clues to the presence of a linear structure in the image by showing a measure of the linear dependence of the gray scale. The higher the correlation value obtained, it means that the intensity of adjacent pixels varies together.The value of correlation can be calculated using equation number 2.

$$\sum_{x,y} \frac{(x-\mu_x)(y-\mu_y)p(x,y)}{\sigma_x\sigma_y} \quad (2)$$

c. Homogeneity

This feature measures the level of softness in an image by measuring the number of similar gray levels in the image. If the pixels are more uniform, the homogeneity will be high.The value of homogeneity can be calculated using equation number 3.

$$\sum_{x,y} \frac{p(x,y)}{1+|x-y|} \quad (3)$$

d. Energy

The texture feature represents a measure of the concentration of the intensity pairs in the matrix.The higher the energy, the higher the similarity value.The value of energy can be calculated using equation number 4.

$$\sum_{x,y} p(x, y)^2 \quad (4)$$

2.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that works by processing data based on predetermined targets. In contrast to unsupervised learning algorithms, which do not have targets that are directed in processing data.The concept of SVM work is to find the best hyperplane that separates classes in an input space[13].

3. Research Methodology

3.1 Dataset

The chest X-Ray Covid-19 image dataset was taken through the website www.kaggle.com with PNG image format.There is a total of 408 training data with 204 normal images and 204 Covid-19 images. Then the test / testing data amounted to 84 each for normal and Covid-19 images.





Fig 1. Normal Chest X-Ray Image
(Source: Kaggle.com)



Fig 2. Covid-19 Chest X-Ray Image
(Source: Kaggle.com)

3.2 Research Steps

The research steps to be built are represented in a flowchart and illustrated in Fig 3.

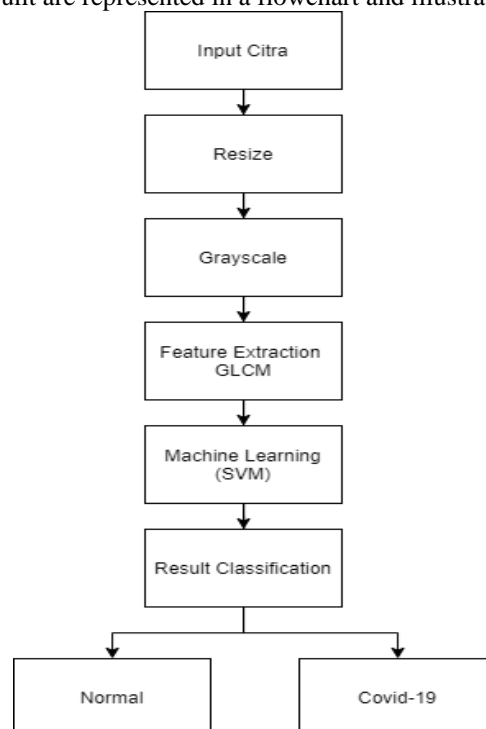


Fig 3. General Research Steps [14]

Fig 3 shows the flow of the research design starting from the image data input process to the classification results. When the image data is entered, preprocessing will be carried out in the form of a change in the size scale of the original image and then converted into a gray image. After that, the image will enter the feature extraction stage with GLCM and produce extraction features such as contrast, correlation, energy, homogeneity, and entropy, dissimilarity, and ASM. This study will use several features of contrast, correlation, energy, homogeneity, and dissimilarity. Then the results from the GLCM will enter the modeling stage using the Support Vector Machine (SVM) algorithm and then produce a classification.

4. Results and Discussion

The results of the size and grayscale conversion will then carry out the feature extraction stage using the GLCM method which will use five parameters, namely contrast, correlation, energy, homogeneity, and dissimilarity. Then it will be tested with pixel distances $d = 1, 2, 3$ and four different angles, namely angles $0^{\circ}, 45^{\circ}, 90^{\circ},$ and 135° . For implementing the SVM algorithm, the default kernel is the RBF kernel.

The classification results will be displayed in Table 1, Table 2 and Table 3 with some information about the test results which are divided into TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). For the explanation, namely TP (True Positive) is data that matches the label giving positive results, TN (True Negative) is data that matches the label giving negative results, FP (False Positive) is data that matches the label giving positive results but is classified as negative, and FN (False Negative) is data that fits the label giving negative results but is classified as positive [2].

To calculate the value of accuracy, you can use the formula in equation number 5.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\% \quad (5)$$

Table 1
Testing distance $d = 1$ and angle $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$

d	\emptyset	TP	FP	FN	TN	Akurasi (%)
1	0	69	15	1	83	90.47
	45	62	22	3	81	85.11
	90	59	25	4	80	82.73
	135	64	20	2	82	86.90
	All (0, 45, 90, 135)	66	18	1	83	88.69

As can be seen in Table 1, it illustrates the results of the Covid-19 classification using GLCM and SVM that at a distance $d = 1$ and angle $\emptyset = 0^{\circ}$ produces the highest accuracy that is 90.47% and the angle $\emptyset = 90^{\circ}$ has the lowest accuracy that is 82.73.

Table 2
Testing distance $d = 2$ and angle $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$

d	\emptyset	TP	FP	FN	TN	Akurasi (%)
2	0	68	16	1	83	89.88
	45	62	22	3	81	85.11
	90	60	24	4	80	83.33
	135	64	20	2	82	86.90
	All (0, 45, 90, 135)	66	18	1	83	88.69

Then in Table 2, it illustrates the results of the Covid-19 classification using GLCM and SVM that on testing the distance $d = 2$ and angle $\emptyset = 0^{\circ}$ produces the highest accuracy, namely 89.88% and angle $\emptyset = 90^{\circ}$ has the lowest accuracy, namely 83.33%.

Table 3
Testing distance $d = 3$ and angle $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$

d	\emptyset	TP	FP	FN	TN	Akurasi (%)
3	0	65	19	2	82	87.50
	45	58	26	4	80	82.14
	90	56	28	5	79	80.35
	135	60	24	3	81	83.92
	All (0, 45, 90, 135)	65	19	3	81	86.90



In the last test as can be seen in Table 3, it illustrates the results of the Covid-19 classification using GLCM and SVM that at a distance $d = 3$ and angle $\phi = 0^0$ produces the highest accuracy, namely 87.50% and angle $\phi = 90^0$ has the lowest accuracy that is 80.35%.

Based on the results of the tests that have been carried out as can be seen in Table 1, Table 2, and Table 3, it provides a conclusion that from the three tests at a distance $d = 1,2,3$. The test with the angular direction $\phi = 0^0$ gives the greatest accuracy results compared to the others with the respective accuracy of 90.47%, 89.88% and 87.50%. The same results also occurred in the test with an angle of $\phi = 90^0$ giving the lowest accuracy results compared to the others with the respective accuracy of 82.73%, 83.33% and 80.35%.

5. Conclusion

The application of the Support Vector Machine (SVM) algorithm in the classification of the Covid-19 Chest X-Ray image and feature extraction using the GLCM method provides fairly high accuracy results with an accuracy of 90.47% and the lowest accuracy result is 80.35%. Then the test with the angle $\phi = 0^0$ gives the highest accuracy results and the test at the angle $\phi = 90^0$ gives the lowest accuracy results for all tests at a distance of $d = 1,2,3$.

Most misclassification is in normal image prediction that occurs in all tests. And the least misclassification errors in the prediction of the Covid-19 Chest X-Ray image that occurred in all tests.

6. References

- [1] J. Gustav, "Update Corona Dunia 19 April: 141 Juta Kasus Covid-19 | 3 Juta Kematian," *Kompas.com*, 2021. <https://www.kompas.com/tren/read/2021/04/19/085000965/update-corona-dunia-19-april--141-juta-kasus-covid-19-3-juta-kematian?page=all> (accessed Apr. 19, 2021).
- [2] C. A. Nugroho, "Klasifikasi K-Nearest Neighbor Chest X-Ray Pasien COVID-19 Dengan Haralick Features dan Histogram Of Oriented Gradient," *J. Ilm. Mat.*, vol. 9, no. 1, pp. 188–195, 2021, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/mathunesa/article/view/38234/34222>.
- [3] K. S. Soin, "Detection and Diagnosis of COVID-19 via SVM-based Analyses of X-Ray Images and Their Embeddings," *Int. J. Innov. Sci. Res. Technol.*, vol. 5, no. 10, pp. 644–648, 2020, [Online]. Available: <https://ijisrt.com/detection-and-diagnosis-of-covid19-via-svmbased-analyses-of-xray-images-and-their-embeddings>.
- [4] C. Wijaya, H. Irsyad, and W. Widhiarso, "Klasifikasi Pneumonia Menggunakan Metode K- Nearest Neighbor Dengan Ekstraksi GLCM," *J. Algoritma.*, vol. 1, no. 1, pp. 33–44, 2020, [Online]. Available: <http://jurnal.mdp.ac.id/index.php/algoritma/article/view/431>.
- [5] N. W. S. Saraswati, N. W. Wardani, and I. G. A. A. D. Indradewi, "Detection of Covid Chest X-Ray using Wavelet and Support Vector Machines," *Int. J. Eng. Emerg. Technol.*, vol. 5, no. 2, pp. 116–121, 2020, doi: <https://doi.org/10.24843/IJEET.2020.v05.i02.p019>.
- [6] R. A. Rizal, S. Gulo, O. D. C. Sihombing, A. B. M. Napitupulu, A. Y. Gultom, and T. J. Siagian, "Analisis Gray Level Co-Occurrence Matrix (GlcM) Dalam Mengenali Citra Ekspresi Wajah," *J. Mantik*, vol. 3, no. January, pp. 31–38, 2019, [Online]. Available: <http://iocscience.org/ejournal/index.php/mantik/article/view/497/302>.
- [7] L. Marlina, K. Munadi, and F. Arnia, "Klasifikasi Kanker Payudara Menggunakan Teknik Support Vector Machine (Svm) Pada Citra Termografi," *J. Karya Ilm. Tek. Elektro*, vol. 5, no. 1, pp. 11–13, 2020, doi: <https://doi.org/10.24815/kitektro.v5i2.15571>.
- [8] dr. L. Felicia, "Virus Corona (COVID-19)," *sehatq.com*, 2021. <https://www.sehatq.com/penyakit/virus-corona> (accessed Apr. 19, 2021).
- [9] Z. Prihatini, "8 Gejala Tak Biasa yang Menandakan Seseorang Terinfeksi Corona," *health.detik.com*, 2020. <https://health.detik.com/berita-detikhealth/d-5306129/8-gejala-tak-biasa-yang-menandakan-seseorang-terinfeksi-corona/2> (accessed Apr. 19, 2021).
- [10] R. A. Surya, A. Fadlil, and A. Yudhana, "Ekstraksi ciri metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor untuk klasifikasi citra batik pekalongan," *J. Inform. Pengemb. IT (JPIT)*, Vol. 02, No. 02, Juli 2017, vol. 02, no. 02, pp. 23–26, 2017, doi: <http://dx.doi.org/10.30591/jpit.v2i2.520>.
- [11] Y. Agussationo, I. Soesanti, and W. Najib, "Klasifikasi citra x-ray diagnosis tuberkulosis berbasis Fitur Statistis," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 3, pp. 736–745, 2018, doi: [10.29207/resti.v2i3.523](https://doi.org/10.29207/resti.v2i3.523).
- [12] F. Agustina and Z. A. Ardiansyah, "Identifikasi Citra Daging Ayam Kampung dan Broiler Menggunakan Metode GLCM dan Klasifikasi-NN," *J. Infokam*, vol. XVI, no. 1, pp. 25–36, 2020, [Online]. Available: <http://amikjtc.com/jurnal/index.php/jurnal/article/view/196>.
- [13] A. S. Febrianti, T. A. Sardjono, and A. F. Babgei, "Klasifikasi Tumor Otak pada Citra Magnetic Resonance Image dengan Menggunakan Metode Support Vector Machine," *J. Tek. ITS*, vol. 9, no. 1, pp. A118–A123, 2020, doi: [10.12962/j23373539.v9i1.51587](https://doi.org/10.12962/j23373539.v9i1.51587).
- [14] Muhathir, "Klasifikasi Ekspresi Wajah Menggunakan Bag of Visual Words," *J. Informatics Telecommun. Eng.*, vol. 1, no. 2, pp. 73–81, 2018, doi: <https://doi.org/10.31289/jite.v1i2.1426>.

