Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 7 No. 1, April 2022, pp. 8 – 15

New Image Texture Feature for Chest X-Ray Classification

Prajanto Wahyu Adi[*1], Fajar Agung Nugroho[2],
Universitas Diponegoro, Jl. Prof. Sudarto No.13, Indonesia (+6224) 7465403
*E-mail : prajanto@live.undip.ac.id[*1]; fajar@live.undip.ac.id[2]*
**Corresponding Author*

Yani Parti Astuti[3]
Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, Indonesia (+6224) 3517261
E-mail : yanipartiastuti@dsn.dinus.ac.id[3]

Abstract - This study proposes a new feature extraction model to identify CXR images of covid-19 and pneumonia has a high visual resemblance. The feature extraction model starts by using histogram equalization and average filters as lowpass features and high pass features obtained through Laplacian and LoG filters. In the next step, covariance matrix of image along with the entire features are used to produce an eigen vector that will be used as a feature vector in the classification process. The final stage is the process of testing features on the classification algorithms KNN, SVM, LDA, Naïve Bayes, and Decision Tree through a 10-foldcross validation scheme with 0.9 training data and 0.1 test data. The first experiment for the Covid-19 and normal classes shows that the proposed model is able to produce an accuracy of 96% as the comparison model with GLCM texture extraction have an accuracy value of 91%. The second test is conducted for the class Covid-19 and pneumonia and obtained an accuracy value of 89% for the proposed model and 73% for the GLCM texture extraction. Experiments proved that the proposed model successfully outperformed the GLCM texture extraction model in all of classification algorithms used.

Keywords – classification, chest x-ray, feature extraction, covid-19, pneumonia

1. INTRODUCTION

The Covid-19 pandemic has been hit the globe for almost two years [1][2], infecting more than 243 million people with more than 4 million of deaths. Various efforts in multidisciplinary science have been made to overcome this problem. In the field of computer science, more research is done on the detection process of symptoms that arise due to the Covid-19 virus [3] and then further classify between healthy people and those infected by viruses based on medical images of the lungs such as Computed Tomography (CT)[4]–[6], Chest X-Ray (CXR) [7]–[9], or Magnetic Resonance Imaging (MRI) [10]

Current research tends to utilize deep learning architectures [11] with some model variations such as [12] with the highest accuracy of 99% and the lowest accuracy being 91%, as well as convolutional-based Neural Networks (CNN) [13] with 98% accuracy and 95% for 2 different datasets. These models require high computing costs as well as very large numbers of datasets in order to achieve the expected results. Such methods also tend to be blurry or vague, where a researcher will have difficulty understanding the model, making model improvements, and in decision making. On the other hand, the use of new feature extraction-based models and classification algorithms is also able to produce good performance, among others: Fractional Oder and Cuckoo Search (FO-CS) [14] with accuracy rates of 76% and 97%

on two different datasets, Joint Classification and Segmentation (JCS) [15] and radiomic classifier [16] with sensitivities of 95% and 75%. In addition, experiments with several existing methods such as Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG) with K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Bag of Tree were able to produce accuracy above 86% [17]. Another study used a combination of the popular Algorithms Gray Level Cooccurrence Matrix (GLCM) and Neural Network (NN) [18] with an accuracy rate of 88%. These models also have a lower level of complexity than those based on deep learning and make it easier for researchers to understand the resulting model so as to facilitate in implementation, improve the model, and make decisions.

In general, CXR imagery and CT imagery are able to clearly illustrate the difference between healthy lung conditions and those infected by disease clearly. However in some cases CT imagery results show abnormal findings [19] and are not recommended for use in the diagnosis of patients without symptoms [20]. This is in line with most studies focusing on the classification between images of healthy lungs and those infected by Covid-19, whereas the symptoms caused by Covid-19 are very similar to pneumonia symptoms in particular when viewed based on the visual features of imagery [21]–[24] The appearance of a healthy lung image will look very different from the lungs infected by Covid-19 so that it can be distinguished easily by the human eye. Therefore, the studies conducted should better consider the visual condition of the lungs infected by Covid-19 and caused by other similar diseases.

The study proposes a model that utilizes the eigen vector of the CXR image's covariance value to the equalization value of the and the homogeneous regions of the image obtained with the average filter as well as heterogeneous regions of the image resulting from Laplacian and Laplacian of Gaussian (LoG) filters. The filters used aim to get the difference value of each area in the image. The next process is the calculation of covariance to find out the relationship between these values. In the final stage, the covariance values are simplified into eigen vectors that are characteristics or features of the CXR image. The features obtained will be further classified process with several machine learning algorithms such as KNN, SVM, Linear Discriminant Analysis (LDA), Naïve Bayes (NB), and Decision Tree (DT). The model is expected to improve the accuracy of classification algorithms in particular to identify CXR image of Covid-19 and pneumonia.

2. RESEARCH METHOD

The method proposed in this study is to use histogram equalization, average filters, Laplacian filters and LoG filters that will each produce feature-1, feature-2, feature-3, and feature-4. The original image along with the four features will then be calculated covariance which is the relationship between them. The next process is to calculate the eigen vector of the covariance value that will produce a feature vector. The vector features obtained will then be divided into 2 parts, 90% for training and 10% for testing with the 10-fold cross validation method. The proposed model is seen in Figure 1. The classification algorithms used in this study are widely used methods such as KNN, SVM, LDA, NB, and DT. Tests will be conducted to compare the capabilities of the proposed feature extraction method against popular feature extraction methods which is GLCM.

2.1 Dataset

The data used in the study was a `*.png` formatted Chest X-ray image obtained from https://www.kaggle.com/tawsifurrahman/covid19-radiography-database/version/3 [24] which consists of 1200 positive images of Covid-19, 1341 normal imagery, and 1345 pneumonia images. Examples of images of each class can be seen in Figure 2.

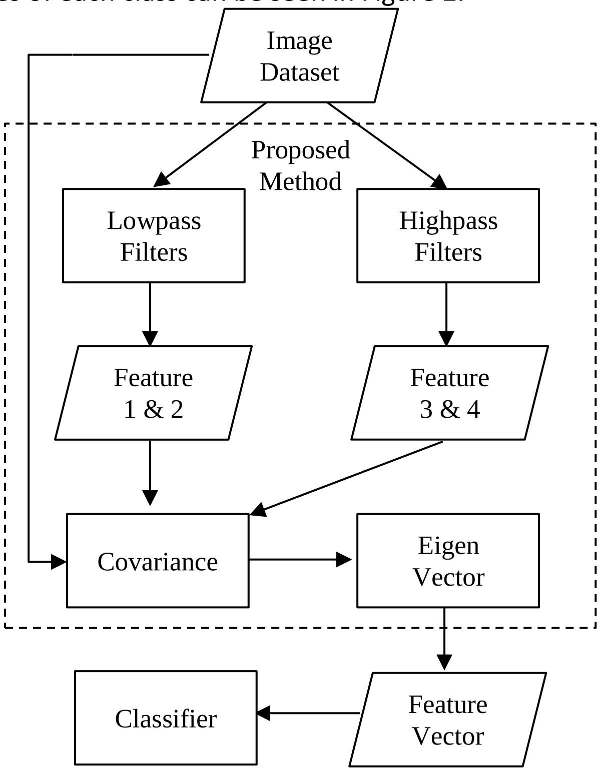

Figure 1 Proposed Model

The image bellow shows the visual representation of X-ray images of lungs from 3 classes namely normal, Covid-19, and pneumonia. From the picture it can be seen that healthy lungs look different from the lungs infected by Covid-19 and pneumonia. The main challenge in the study was to classify images of lungs infected by Covid-19 and pneumonia because both images showed very similar displays.

Figure 2 Example of chest X-ray image left: normal; middle: Covid-19; right: pneumonia

2.2. Feature Extraction

The study proposes a new feature extraction model based on the eigen vector of feature correlation. Features extraction of CXR image are performed using histogram equalization as well as low pass and high pass filter. The detailed feature extraction process is as follows:

Step 1. Perform histogram equalization to get feature *a* as the first low pass feature

$$a_{i,j} = \left\lfloor (L-1) \sum_{n=0}^{f_{i,j}} p_n \right\rfloor \quad (1)$$

where $a_{i,j}$ and $f_{i,j}$ are feature-1 and image in rows *i* and column *j, L* is maximum value of pixels, and p_n is the probability of occurrence of pixels with values *n* from 0 to *L*-1.

Step 2. Use the average filter to get feature *b* as a second low pass feature

$$b_{i,j} = \frac{1}{MN} \sum_{i,j=0}^{M,N} f_{i,j} \quad (2)$$

where $b_{i,j}$ is feature-2 on rows *i* and columns *j, M* and *N* are rows columns.

Step 3. Calculating the *c* feature through the Gaussian filter which is the first high pass feature

$$c_{i,j} = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} \quad (3)$$

where $c_{i,j}$ is feature-3 in row *i* and column *j,* and σ^2 is standard deviation

Step 4. Get feature *d* through a LoG filter that will be used as a high pass feature to two

$$d_{i,j,t} = -\frac{1}{\pi\sigma^4} \left[1 - \frac{i^2+j^2}{2\sigma^2}\right] e^{-\frac{i^2+j^2}{2\sigma^2}} \quad (4)$$

where $d_{i,j}$ is feature-4 on row *i* and column *j,*

Step 5. Calculate the covariance of the original image along with the four features that have been obtained in steps 1 to 4 to find the strength or magnitude of the relationship value between these variables

$$cov_{x,y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1} \quad (5)$$

where *x* and *y* are the first and second data that will be used to form a matrix covariance as follows

$$R = \begin{bmatrix} cov_{f,f} & cov_{a,f} & cov_{b,f} & cov_{c,f} & cov_{d,f} \\ cov_{f,a} & cov_{a,a} & cov_{b,a} & cov_{c,a} & cov_{d,a} \\ cov_{f,b} & cov_{a,b} & cov_{b,b} & cov_{c,b} & cov_{d,b} \\ cov_{f,c} & cov_{a,c} & cov_{b,c} & cov_{c,c} & cov_{d,c} \\ cov_{f,d} & cov_{a,d} & cov_{b,d} & cov_{c,d} & cov_{d,d} \end{bmatrix} \quad (6)$$

Step 6. The final process is to obtain the Eigen vector from the covariance matrix which serves as the main characteristic or feature of the CXR image data.

$$Rv = \lambda v \quad (7)$$

where λ is the eigenvalue used to find the eigenvector *v*

$$V = \begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & v_{1,4} & v_{1,5} \\ v_{2,1} & v_{2,2} & v_{2,3} & v_{2,4} & v_{2,5} \\ v_{3,1} & v_{3,2} & v_{3,3} & v_{3,4} & v_{3,5} \\ v_{4,1} & v_{4,2} & v_{4,3} & v_{4,4} & v_{4,5} \\ v_{5,1} & v_{5,2} & v_{5,3} & v_{5,4} & v_{5,5} \end{bmatrix} \quad (8)$$

The eigen vector will be used as a feature vector at the stages of training, classification and testing of the model.

3. RESULTS AND DISCUSSION

The experiment was conducted under 2 conditions, the first test is conducted to classify the Covid-19 class and normal while the second experiment conducts on covid-19 and pneumonia.

3.1. Covid-19 vs Normal

The first test was conducted to determine the basic capabilities of the proposed feature extraction model to distinguish the covid-19 class from the normal class through a classification algorithm. Visually, normal CXR imagery will be easily distinguished from CXR covid-19 imagery, so a good algorithm must be able to achieve accuracy above 90% in order to be further developed into the initial diagnosis system of the disease.

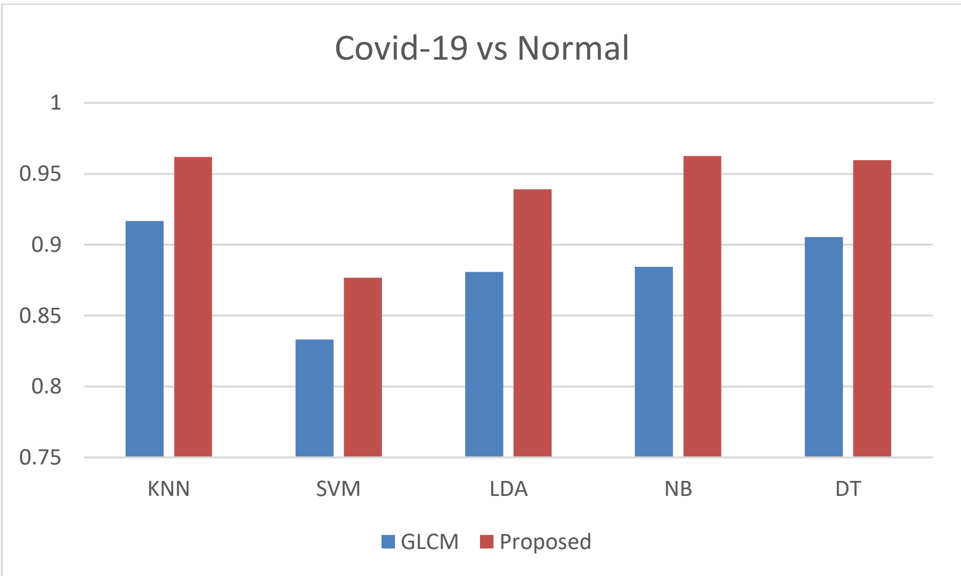

Figure 3 Comparison of feature extraction models in covid-19 vs normal class

In the proposed model, the highest accuracy is met with naïve bayes and KNN algorithms with an accuracy of 96%, while in GLCM the highest accuracy is on the KNN algorithm with 91%. The first test proved that the proposed model was able to produce better accuracy than the extraction of GLCM features across classification algorithms as seen in Figure 3. This is because the proposed model is able to produce features that have a higher standard deviation value compared to GLCM. The highest standard deviation obtained was 0.879 at *v*2.1 with an average deviation of 0.214, a higher value than the highest GLCM feature extraction yield of 0.057 with an average of 0.032. High standard deviation values in image features will result in relatively higher accuracy.

3.2. Covid-19 vs Pneumonia

The second trial was conducted to classify the class Covid-19 and pneumonia. Tests were conducted to test the capabilities of feature extraction models with very similar image conditions in the Covid-19 and pneumonia classes. In the second experiment the highest results were produced with the KNN algorithm.

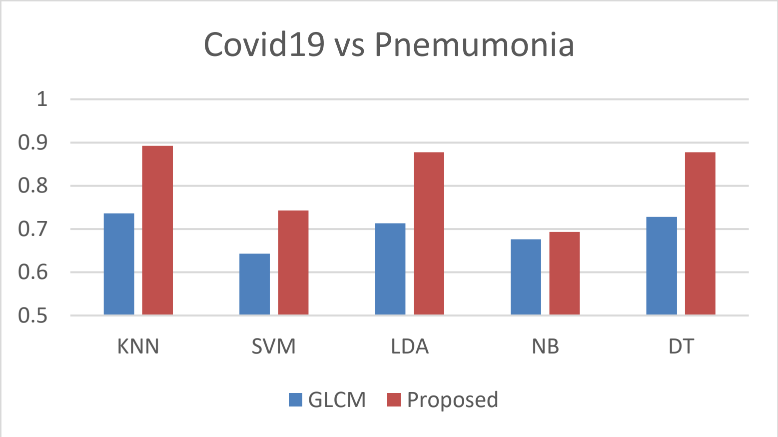

Figure 4 Comparison of feature extraction models in class covid-19 vs pneumonia

The proposed extraction model is capable of outperforming GLCM across classification algorithms as seen in Figure 4. The model has an accuracy rate of 89% while at GLCM extraction the highest accuracy only reaches 73%. When compared to the previous experiment, there was a slight decrease in accuracy of 7% in the proposed model while in GLCM extraction there was a very significant decrease of 18%. The results of the measurement of the standard deviation value of the feature also showed the same thing, where the proposed feature model has a maximum deviation value of 0.937 with an average of 0.227 while the GLCM feature has a maximum feature deviation standard of 0.055 with an average of 0.013. The results showed that the proposed feature extraction model was able to work well, even for classification on images that have a high visual resemblance this method was able to produce a higher feature deviation value so as to improve the accuracy of the classification algorithm.

4. CONCLUSION

The study proposes a feature extraction model to identify CXR Covid-19 imagery and pneumonia with very high similarities that complicate the image-based classification process. The experiment conducted 3886 data with 3 classes: 1200 images of Covid-19, 1341 images of normal, and 1345 images of pneumonia. In the first test with the Covid-19 vs. normal class, the proposed feature extraction model was able to achieve the highest accuracy of 96% on KNN classification algorithms and naïve bayes while GLCM extraction gets a maximum accuracy of 91% on KNN algorithms. In the second trial with the class Covid-19 vs pneumonia, the highest accuracy was again achieved by the KNN where in the proposed model produced the highest accuracy of 89% while in GLCM obtained accuration of 73%. The proposed model is able to produce feature values with high standard deviation so as to increase accuracy in all classification algorithms used. In further research, the process of selecting features based on the highest deviation value to improve accuracy while speeding up the classification process. In addition, comparisons can be made to feature extraction models and classification algorithms to identify CXR images of the normal Covid-19 and pneumonia classes.

REFERENCES

[1] K. Roosa *et al.*, "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020," *Infect. Dis. Model.*, vol. 5, pp. 256–263, 2020, doi: 10.1016/j.idm.2020.02.002.

[2] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. A. Abo-Elsoud, "A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier," *Knowledge-Based Syst.*, vol. 205, p. 106270, 2020, doi: 10.1016/j.knosys.2020.106270.

[3] V. K. Gupta, A. Gupta, D. Kumar, and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," *Big Data Min. Anal.*, vol. 4, no. 2, pp. 116–123, 2021, doi: 10.26599/BDMA.2020.9020016.

[4] T. Yan, P. K. Wong, H. Ren, H. Wang, J. Wang, and Y. Li, "Automatic distinction between COVID-19 and common pneumonia using multi-scale convolutional neural network on chest CT scans," *Chaos, Solitons and Fractals*, vol. 140, p. 110153, 2020, doi: 10.1016/j.chaos.2020.110153.

[5] L. Sun *et al.*, "Adaptive Feature Selection Guided Deep Forest for COVID-19 Classification with Chest CT," *IEEE J. Biomed. Heal. INFORMATICS*, vol. 24, no. 10, pp. 2798–2805, 2020.

[6] J. Zhang, Y. Chu, and N. Zhao, "Supervised framework for COVID-19 classification and lesion localization from chest CT," *Ethiop. J. Heal. Dev.*, vol. 34, no. 4, pp. 235–242, 2020.

[7] R. C. Joshi *et al.*, "A deep learning-based COVID-19 automatic diagnostic framework using chest X-ray images," *Biocybern. Biomed. Eng.*, vol. 41, no. 1, pp. 239–254, 2021, doi: 10.1016/j.bbe.2021.01.002.

[8] Z. Wang *et al.*, "Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays," *Pattern Recognit.*, vol. 110, 2021, doi: 10.1016/j.patcog.2020.107613.

[9] J. D. Arias-Londono, J. A. Gomez-Garcia, L. Moro-Velazquez, and J. I. Godino-Llorente, "Artificial Intelligence applied to chest X-Ray images for the automatic detection of COVID-19. A thoughtful evaluation approach," *IEEE Access*, vol. 8, pp. 226811–226827, 2020, doi: 10.1109/ACCESS.2020.3044858.

[10] P. Torkian, H. Rajebi, T. Zamani, N. Ramezani, P. Kiani, and S. Akhlaghpoor, "Magnetic resonance imaging features of coronavirus disease 2019 (COVID-19) pneumonia: The first preliminary case series," *Clin. Imaging*, vol. 69, no. 5, pp. 261–265, 2021, doi: 10.1016/j.clinimag.2020.09.002.

[11] A. Gupta, Anjum, S. Gupta, and R. Katarya, "InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray," *Appl.*

Soft Comput., vol. 99, p. 106859, 2021, doi: 10.1016/j.asoc.2020.106859.
[12] E. Hussain, M. Hasan, M. A. Rahman, I. Lee, T. Tamanna, and M. Z. Parvez, "CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images," *Chaos, Solitons and Fractals*, vol. 142, p. 110495, 2021, doi: 10.1016/j.chaos.2020.110495.
[13] E. F. Ohata *et al.*, "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning," *IEEE/CAA J. Autom. Sin.*, vol. 8, no. 1, pp. 239–248, 2021, doi: 10.1109/JAS.2020.1003393.
[14] D. Yousri, M. Abd Elaziz, L. Abualigah, D. Oliva, M. A. A. Al-qaness, and A. A. Ewees, "COVID-19 X-ray images classification based on enhanced fractional-order cuckoo search optimizer using heavy-tailed distributions," *Appl. Soft Comput.*, vol. 101, p. 107052, 2021, doi: 10.1016/j.asoc.2020.107052.
[15] Y. H. Wu *et al.*, "JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 3113–3126, 2021, doi: 10.1109/TIP.2021.3058783.
[16] L. Wang *et al.*, "Multi-classifier-based identification of COVID-19 from chest computed tomography using generalizable and interpretable radiomics features," *Eur. J. Radiol.*, vol. 136, p. 109552, 2021, doi: 10.1016/j.ejrad.2021.109552.
[17] A. Saygılı, "A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods," *Appl. Soft Comput.*, vol. 105, p. 107323, 2021, doi: 10.1016/j.asoc.2021.107323.
[18] S. Varela-Santos and P. Melin, "A new approach for classifying coronavirus COVID-19 based on its manifestation on chest X-rays using texture features and neural networks," *Inf. Sci. (Ny).*, vol. 545, pp. 403–414, 2021, doi: 10.1016/j.ins.2020.09.041.
[19] Y. Peng, Y. Tang, S. Lee, Y. Zhu, R. M. Summers, and Z. Lu, "COVID-19-CT-CXR: A freely accessible and weakly labeled chest X-Ray and CT image collection on COVID-19 from biomedical literature," *IEEE Trans. Big Data*, vol. 7, no. 1, pp. 3–12, 2021, doi: 10.1109/TBDATA.2020.3035935.
[20] B. Hochhegger *et al.*, "Coronavirus Disease 2019 (COVID-19) Pneumonia Presentations in Chest Computed Tomography: A Pictorial Review," *Curr. Probl. Diagn. Radiol.*, vol. 50, no. 3, pp. 436–442, 2021, doi: 10.1067/j.cpradiol.2020.06.010.
[21] H. X. Bai *et al.*, "Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT," *Radiology*, vol. 296, no. 2, pp. E46–E54, 2020, doi: 10.1148/radiol.2020200823.
[22] F. Xiong *et al.*, "The clinical classification of patients with COVID-19 pneumonia was predicted by Radiomics using chest CT," *Medicine (Baltimore).*, vol. 100, no. 12, pp. 1–8, 2021, doi: 10.1097/MD.0000000000025307.
[23] T. Tuncer, F. Ozyurt, S. Dogan, and A. Subasi, "A novel Covid-19 and pneumonia classification method based on F-transform," *Chemom. Intell. Lab. Syst.*, vol. 210, no. January, p. 104256, 2021, doi: 10.1016/j.chemolab.2021.104256.
[24] M. E. H. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020.