Machine Learning Performances for Covid-19 Images Classification based Histogram of Oriented Gradients Features

2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) | 978-1-6654-8684-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEMTRONICS55184.2022.9795854

1st Yessi Jusman*
Department of Electrical Engineering, Faculty of Engineering Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
*Corresponding
Email:yjusman@umy.ac.id

2nd Wikan Tyassari
Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

3rd Difa Nisrina
Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

4th Fahrul Galih Santosa
Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

5th Nugroho Abdi Prayitno
Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

***Abstract*— Coronavirus disease (Covid-19) is an infectious disease that attacks the respiratory area caused by the severe acute respiratory syndrome (SARS-CoV-2) virus. According to the World Health Organization (WHO) as of April 2022, there were more than 500 million cases of Covid-19, and 6 million of them died. One of the tools to detect Covid-19 disease is using X-ray images. Digital X-ray images implementation can be developed classification method using machine learning. By using machine learning, the diagnosis of this disease can be faster. This study applied a features extraction method using the Histogram of Oriented Gradients (HOG) algorithm and the Linear Support Vector Machine (SVM), K-Nearest Neighbor (KNN) Medium and Decision Tree (DT) Coarse Tree classification methods. The study can be used in the diagnosis of Covid-19 disease. The best method among the classification methods is features extraction from HOG algorithm and DT Coarse Tree. The highest values of accuracy, precision, recall, specificity, and F-score were 83.67%, 96.30%, 78.79%, 98.25, and 76.48%.**

Keywords—Covid-19, Histogram of Oriented Gradients, Support Vector Machine, K-Nearest Neighbor, Decision Tree.

I. INTRODUCTION

In the last three years, the world has been hit by a global pandemic that has affected the entire world community. Coronavirus disease (Covid-19) is an infectious disease that attacks the respiratory area caused by the Severe Acute Respiratory Syndrome (SARS-CoV-2) virus [1]. According to data from the World Health Organization (WHO) as of April 2022, there were more than 500 million cases due to Covid-19, and 6 million of them died [2].

This Covid-19 disease is quite dangerous apart from being deadly because the transmission of Covid-19 is very fast. With its rapid spread and very easy transmission, fast and accurate detection is needed to prevent its spread. There are several methods of detecting Covid-19 disease, one of which is detection using x-ray images [3]. Although diagnosis using Polymerase Chain Reaction (PCR) is more widely used, detection with x-ray images is cheaper because it can detect a large number of patients and is not limited by test equipment [4]. However, the diagnosis of Covid-19 disease using x-ray images can only be carried out by the relevant expert or doctor. So it takes quite a long time and is limited to the number of experts or doctors available [5] [6].

To speed up the diagnosis of Covid-19 disease, it is carried out on a computer-based basis. The development of computer-based Covid-19 disease detection is one of the studies that has been widely developed. In recent years, studies [7] [8] have used computers to help in the early diagnosis of Covid-19.

Machine Learning (ML) is well known for its high-performance image processing and its application in certain image classifications. Thus, ML is an alternative for automatic detection of Covid-19 disease using x-ray images. The use of ML in the diagnosis of Covid-19 disease has been carried out by several studies [9] [10].

Image processing is an important aspect in the ML process. For an image to be analyzed using the ML method, the image must be extracted (feature). One feature extraction method that is quite often used is the Histogram of Oriented Gradients (HOG). HOG feature extraction focuses on the shape and angle of the feature gradient. Several related studies that use HOG as a feature extraction method in detecting Covid-19 images are. [5][11][12] yielded the highest accuracy values of 98.66%, 98.11%, and 98.5%.

978-1-6654-8684-2/22/$31.00 ©2022 IEEE

To determine the image of Covid-19 and not the image of Covid-19, a classification method is used. Several image classification methods include Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Decision Tree (DT). Several studies related to the diagnosis of Covid-19 disease using the SVM, KNN, and DT classification methods include [13] [12]. Several studies use other methods to be applied in the detection of Covid-19 disease [14].

With only a few studies on Covid-19 images combined with the HOG feature extraction method with the SVM, KNN, and DT classification methods. Therefore, in this study the HOG method, with the classification of SVM, KNN, DT is used in the diagnosis of Covid-19 disease with x-ray images.

II. METHOD

A. Data Collection

The image data used in this study is Covid-19 images obtained from open datasets and Machine Learning Projects Kaggle website with the type of lung X-ray images. The number of Covid-19 images used is 1184 images, which are divided into three types of images, namely Covid-19 images (class 1), Pneumonia images (class 2), and Normal lung images (class 3). The number of images for each class is 404 images for class 1, 390 images for class 2 and for 390 images for class 3. The data is divided into 2 types of data sets, namely training and testing. The amount of training data is 90% of the total image data, which is 1065 images, and for data set testing is 10% of the total image data, which is 119 images.

B. System Design

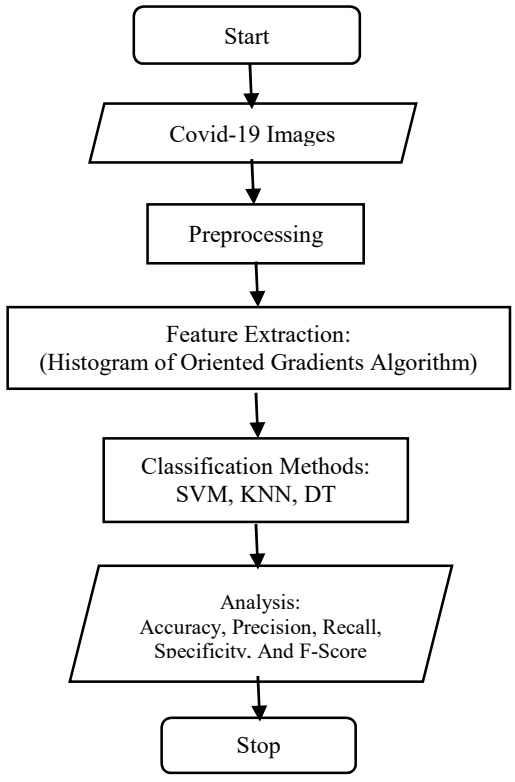

Fig. 1. System Design Flowchart

In this study, the analysis was carried out using the MATLAB software version R2020a. The computer specifications used during the Covid-19 image processing process are shown in Table 1.

TABLE I. COMPUTER SPESIFICATIONS

Hardware	Specification
Processor	Intel® Core i5 9400f
RAM Memory	16GB
GPU	Nividia RTX 2060 6GB

This COVID-19 image classification goes through several steps, including pre-processing, feature extraction, classification, and analysis. The feature extraction used for this study is the Histogram of Oriented Gradients (HOG) algorithm. The classification method used is Linear Support Vector Machine (SVM), K-Nearest Neighbor (KNN) Medium and Decision Tree (DT) Coarse Tree. The whole system design is shown in Figure 1.

C. Pre-pocessing

This Covid-19 image is divided into a training dataset and a testing data set. The next process is preprocessing for the two datasets. The original image with the 2 datasets was flipped vertically as an effort to reproduce the image. To be able to perform feature extraction using the HOG algorithm of the image must be changed from an RGB image to a Grayscale. The image that has been converted to grayscale is converted to histogram to equalize the contrast value of the image.

D. Feature Extraction

The testing and training datasets are processed to obtain feature extraction. This feature extraction stage is carried out after the image has been pre-processed. In this study, the feature extraction used is the HOG algorithm, with the extraction results totaling 36 features.

E. Classification

At this stage, classification is carried out with the extracted data using HOG algorithm. Classification is carried out using 3 different methods, the classification methods used include Linear Support Vector Machine (SVM), K-Nearest Neighbor (KNN) Medium and Decision Tree (DT) Coarse Tree. These three methods were tested ten times running with 10% data validation of the training data. The results of the classification are analyzed based on the training computation time and training accuracy. These results are displayed on the Receiver Operating Characteristic (ROC) graph.

F. Analysis

The results of the Covid-19 image classification were analyzed using a performance matrix. The image dataset used in this analysis is a testing dataset. The performance matrix analysis used includes accuracy, precision, recall, specificity, and f-score. The results of the performance matrix analysis are compared to get the best method.

III. Results And Discussions

A. Feature Extraction Result

The testing and training dataset images are feature extraction using HOG algorithm and produce a total of 36 features for each image, with different extraction values. The results of feature extraction with HOG algorithm are shown in Table 2. Table 2 shows the average extraction results of 10 images in each class.

TABLE II. Feature Extraction Results

Images	Histogram of Oriented Gradients		
	Class 1	*Class 2*	*Class 3*
1	0.142	0.144	0.149
2	0.149	0.145	0.149
3	0.144	0.146	0.139
4	0.149	0.136	0.154
5	0.143	0.125	0.143
6	0.150	0.135	0.137
7	0.144	0.127	0.141
8	0.147	0.133	0.145
9	0.145	0.133	0.142
10	0.151	0.139	0.146
Mean	0.147 ± 0.003	0.136 ± 0.007	0.144 ± 0.005

B. Classification Result

The extracted data obtained were used for classification using the Linear SVM, KNN Medium, and DT Coarse Tree methods. Classification was carried out 10 times running for each method. From the classification with these methods, the value of training accuracy and training time is obtained. The results of the classification, namely training accuracy and computational time are shown in Table 3 and the results of the Receiver Operating Characteristic (ROC) graph. The best ROC graph is shown in Figure 2.

Based on the results of image classification in Table 3, the best training accuracy was obtained by KNN Medium with the highest accuracy value of 70.50% and the fastest time of 0.72 second. The average value of training accuracy obtained by KNN Medium is 69.91% ± 0.33% and the average computation time is 0.78 ± 0.03. In contrast to KNN Medium, the DT Coarse Tree method gets the lowest training accuracy value, which is 62.40% with a time of 0.84 seconds. The average training accuracy of the DT Coarse Tree method is 63.05% ± 0.47% and time 0.88 ± 0.13.

C. Analysis Result

At this stage, a testing dataset that has been extracted has been used as input. Testing data is tested using the best training classification results from each method. The analysis was carried out using a performance matrix comparison for each method to find out the best method for classifying Covid-19 images. The results of the performance matrix for testing are shown in Table 4. The results of the best confusion matrix of the three methods are shown in Figure 3, the best confusion matrix for SVM Linear is 1st running, KNN Medium is 3rd running, and DT Coarse Tree is 5th running.

Based on the data shown in Table 4, the best performance matrix in the classification of Covid-19 images is the DT Coarse Tree method. The accuracy values obtained by the DT Coarse Tree method are 78.10% for class 1, 83.67% for class 2, and 70.09% for class 3. The precision values are 76.48% for class 1, 96.30% for class 2, and 48.15% for class 3. The recall obtained were 63.41% for class 1, 63.41% for class 2, and 78.79% for class 3. The specificity results obtained were 87.54% for class 1, 98.25% for class 2, 66.67% for class 4. Results F- the score obtained is 69.33% for class 1, 76.48% for class 2, and class 3 is 59.78%. The results of the comparison of these three methods can be seen in the graph shown in Figure 4. From the graph, the performance of DT Coarse Tree tends to be higher than the other methods.

TABLE III. Classification Results

Run	SVM Linear		KNN Medium		DT Coarse Tree	
	Accuracy (%)	*Time (s)*	*Accuracy (%)*	*Time (s)*	*Accuracy (%)*	*Time(s)*
1	**69.40%**	**1.21**	69.80%	0.79	**63.70%**	**1.28**
2	69.30%	1.23	69.50%	0.81	62.50%	0.80
3	69.30%	1.22	**70.50%**	**0.76**	63.20%	0.80
4	68.80%	1.23	69.90%	0.82	63.20%	0.83
5	68.50%	1.21	70.00%	0.77	62.70%	0.82
6	68.20%	1.22	70.00%	0.75	62.60%	0.82
7	68.70%	1.24	69.30%	0.76	63.00%	0.82
8	68.20%	1.28	70.00%	0.72	62.90%	0.86
9	68.80%	1.33	70.30%	0.81	63.60%	0.85
10	69.10%	1.30	69.80%	0.79	62.40%	0.84
Mean	68.83% ± 0.42%	1.25 ± 0.04	69.91% ± 0.33%	0.78 ± 0.03	63.05% ± 0.47%	0.88 ± 0.13

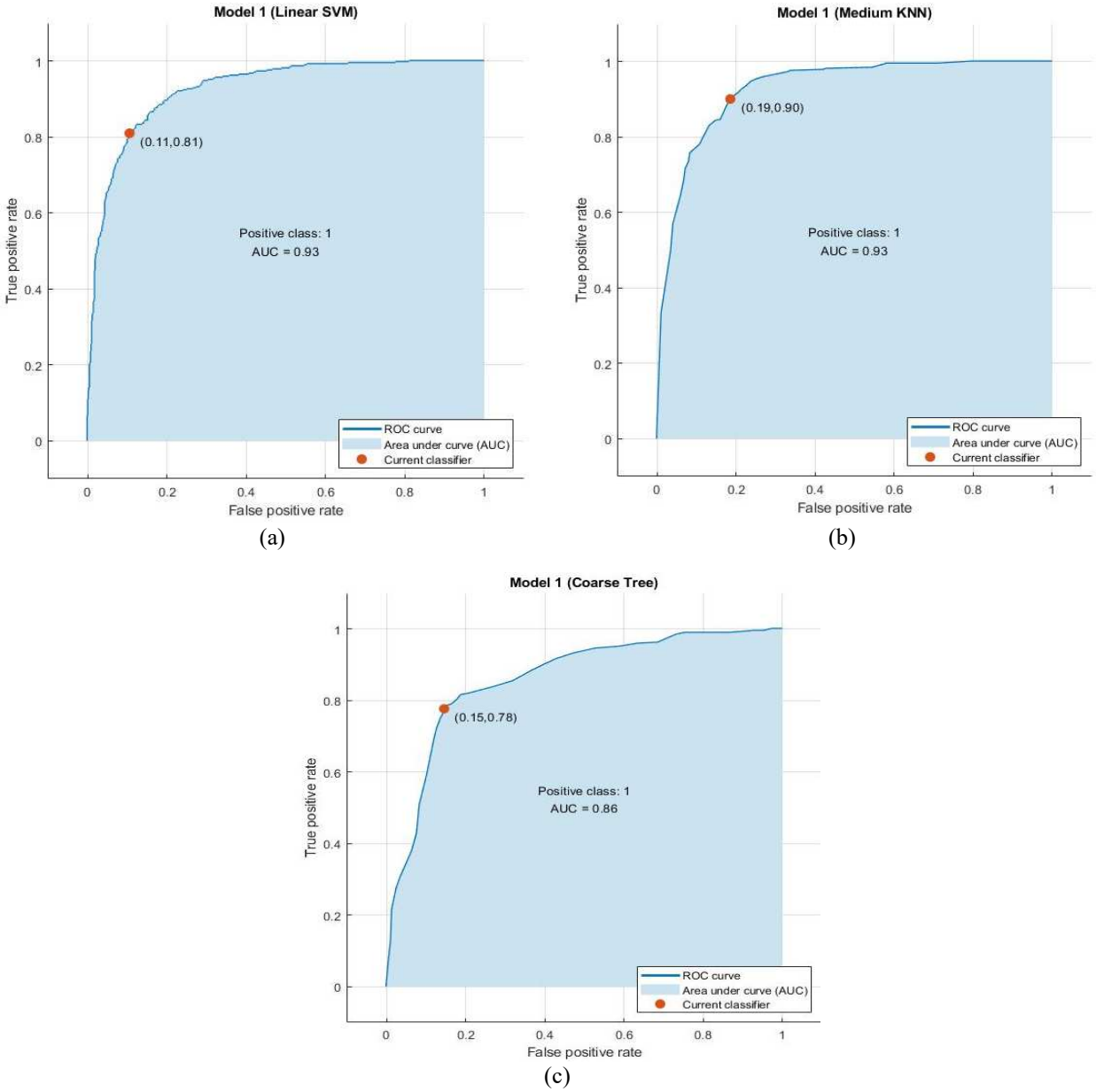

Fig. 2. Best of Receiver Operating Characteristic Graph (a) SVM, (b) KNN, (c) DT

TABLE IV. ANALYSIS PERFORMANCE MATRIX RESULT

Results	**SVM Linear**			**KNN Medium**			**DT Coarse Tree**		
	Class *1*	Class *2*	Class *3*	Class *1*	Class *2*	Class *3*	Class *1*	Class *2*	Class *3*
Accuracy	70.37%	83.52%	66.09%	70.37%	81.72%	67.26%	78.10%	**83.67%**	70.09%
Precision	59.57%	**96.55%**	54.90%	58.18%	76.19%	76.19%	76.48%	96.30%	48.15%
Recall	68.30%	66.67%	63.64%	78.05%	**82.05%**	54.24%	63.41%	63.41%	78.79%
Specificity	71.64%	**97.96%**	67.61%	65.67%	81.48%	81.48%	87.54%	98.25%	66.67%
F-score	63.64%	78.87%	58.95%	66.67%	**79.01%**	63.37%	69.33%	76.48%	59.78%

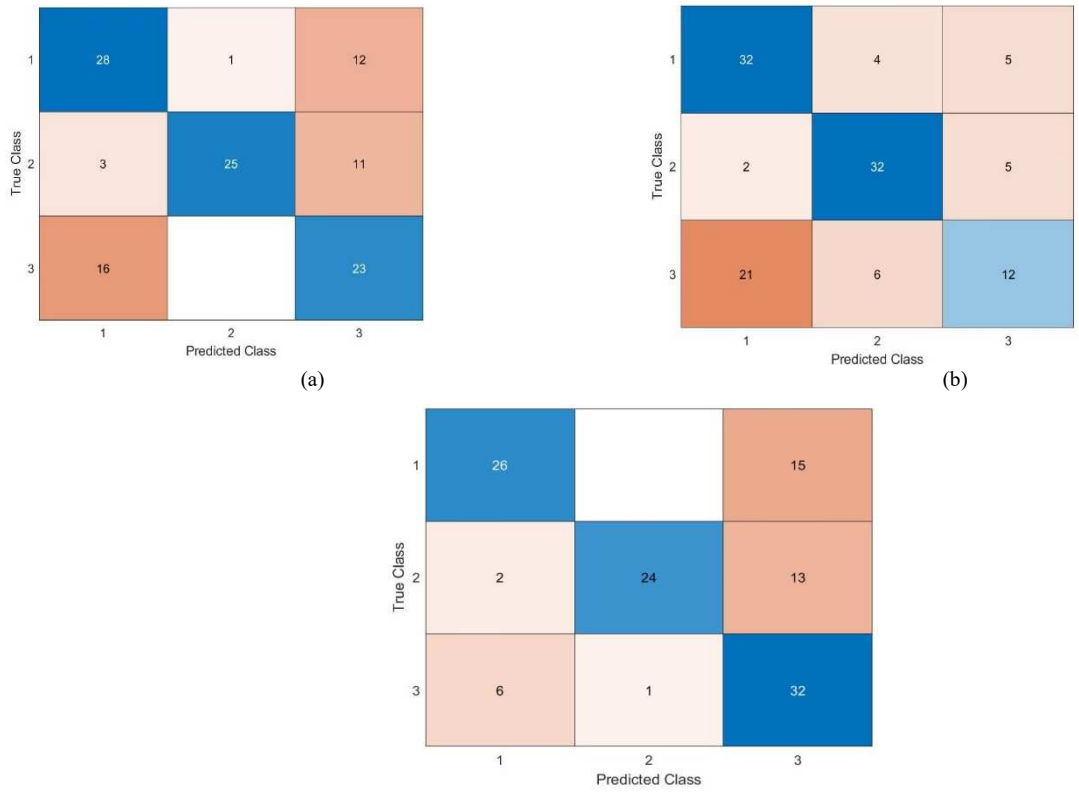

Fig. 3. Best Example of Confussion Matrix Testing (a) SVM, (b) KNN, (c) DT

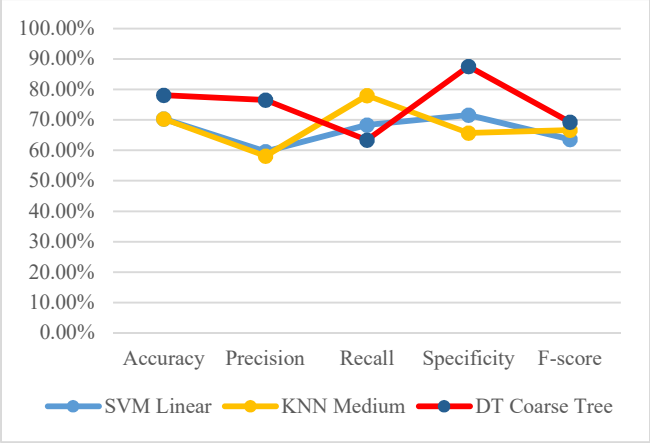

Fig. 4. Comparison Graph of Classification Models

IV. Conclusion

Covid-19 image classification using Linear Support Vector Machine (SVM), K-Nearest Neighbor (KNN) Medium and Decision Tree (DT) Coarse Tree with feature extraction by using Histogram of Oriented Gradients (HOG) algorithm can be developed to help to classify the Covid-19 disease based on X-ray images. It can be concluded that the DT Coarse Tree method is the method that produces the best performance in the classification of Covid-19 images. The results of the DT Coarse Tree produce accuracy values for class 1, class 2, and class 3 of 78.10%, 83.67%, 70.09%. The precision values for class 1, class 2, and class 3 are 76.48%, 96.30%, and 48.15%. The recall scores for class 1, class 2, and class 3 were 63.41%, 63.41%, and 78.79%. The specificity results for class 1, class 2, class3 were 87.54%, 98.25, and 66.67%. The F-score results for class 1, class 2, and class 3 were 69.33%, 76.48%, and 59.78%. Based on these results, it is necessary to test using feature extraction and other classification methods as a comparison reference for the automatic detection method for Covid-19 disease.

Acknowledgment

This research is supported by Universitas Muhammadiyah Yogyakarta and a research project grant from the Ministry of Research and Technology of the Republic of Indonesia.

References

[1] R. Rehouma, M. Buchert, and Y. P. Chen, “Machine learning for medical imaging-based COVID-19 detection and diagnosis,” *Int. J. Intell. Syst.*, p. 10.1002/int.22504, May 2021, doi: 10.1002/int.22504.

[2] S. Kadry, V. Rajinikanth, S. Rho, N. Sri, and M. Raja, “Development of a Machine-Learning System to Classify Scan Images into Normal / COVID-19 Class,” vol. 2, no. December, 2019.

[3] L. Brunese, F. Martinelli, F. Mercaldo, and A. Santone, “Machine

learning for coronavirus covid-19 detection from chest x-rays," *Procedia Comput. Sci.*, vol. 176, pp. 2212–2221, 2020, doi: https://doi.org/10.1016/j.procs.2020.09.258.

[4] M. Chieregato *et al.*, "A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data," *Sci. Rep.*, vol. 12, no. 1, p. 4329, 2022, doi: 10.1038/s41598-022-07890-1.

[5] 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.

[6] M. R. Islam and M. Nahiduzzaman, "Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach," *Expert Syst. Appl.*, vol. 195, no. February, p. 116554, 2022, doi: 10.1016/j.eswa.2022.116554.

[7] D. Al-Karawi, S. Al-Zaidi, N. Polus, and S. Jassim, "Machine Learning Analysis of Chest CT Scan Images as a Complementary Digital Test of Coronavirus (COVID-19) Patients," *medRxiv*, p. 2020.04.13.20063479, 2020, [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.04.13.20063479v1%0Ahttps://www.medrxiv.org/content/10.1101/2020.04.13.20063479v1.abstract.

[8] A. Zargari Khuzani, M. Heidari, and S. A. Shariati, "COVID-Classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images," *Sci. Rep.*, vol. 11, no. 1, p. 9887, 2021, doi: 10.1038/s41598-021-88807-2.

[9] P. Saha, M. S. Sadi, and M. M. Islam, "EMCNet: Automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers," *Informatics Med. Unlocked*, vol. 22, p. 100505, 2021, doi: 10.1016/j.imu.2020.100505.

[10] S. Samsir, J. H. P. Sitorus, Zulkifli, Z. Ritonga, F. A. Nasution, and R. Watrianthos, "Comparison of machine learning algorithms for chest X-ray image COVID-19 classification," *J. Phys. Conf. Ser.*, vol. 1933, no. 1, 2021, doi: 10.1088/1742-6596/1933/1/012040.

[11] A. M. Ayalew, A. O. Salau, B. T. Abeje, and B. Enyew, "Detection and classification of COVID-19 disease from X-ray images using convolutional neural networks and histogram of oriented gradients," *Biomed. Signal Process. Control*, vol. 74, no. October 2021, p. 103530, 2022, doi: 10.1016/j.bspc.2022.103530.

[12] J. N. Hasoon *et al.*, "COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images," *Results Phys.*, vol. 31, no. December 2020, pp. 0–7, 2021, doi: 10.1016/j.rinp.2021.105045.

[13] A. Mohammadi *et al.*, "Diagnosis / Prognosis of COVID-19 Chest Images via Machine Learning and Hypersignal Processing," *IEEE Signal Process. Mag.*, vol. 38, no. 5, pp. 37–66, 2021.

[14] H. Yasar and M. Ceylan, "A novel comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning, and deep learning methods," *Multimed. Tools Appl.*, vol. 80, no. 4, pp. 5423–5447, 2021, doi: 10.1007/s11042-020-09894-3.