COVID-19 X-Ray Images Classification using Support Vector Machine and K-Nearest Neighbor

2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE) | 978-1-6654-7150-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICITACEE55701.2022.9924124

Yessi Jusman
Department of Electrical Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
yjusman@umy.ac.id

Dimas Wildan Mubarok
Department of Electrical Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
dimas.wildan.ft18@mail.umy.ac.id

Slamet Riyadi
Department of Informatics Technology
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
riyadi@umy.ac.id

Siti Nurul Aqmariah Mohd Kanafiah
School of Mechatronic Engineering
Universiti Malaysia Perlis
Perlis, Malaysia
aqmariah@unimap.edu.my

***Abstract*—COVID-19 has significantly influenced living in recent years. Almost all countries have carried out all limitations to reduce its spread. Detection is highly required for further handling of COVID-19. In this study, the detection was performed using classification on 1,184 X-ray images, specifically 404 X-ray images of COVID-19 *positive* people, 390 X-ray images of normal people and 390 X-ray images of pneumonia positive people. The image data were extracted with the Haar wavelet algorithm and classified using the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN); each had three classification models. The Quadratic SVM model obtained the best result with an accuracy of 79.8%.**

Keywords—COVID-19, X-ray Images, Haar Wavelet, Classification

I. INTRODUCTION

Coronavirus Diseases (COVID-19) spreads rapidly throughout the world. The mutation of its virus has become a challenge due to its many variants requiring the vaccine to adjust to various conditions to minimize its spread. Several previous studies conducted COVID-19 detection experiments using X-ray images of people affected by COVID-19. Mohamed El Aziz, Khalid M Khosni et al. [1] employed Fractional Multichannel Exponent Moments (FrMEMs) in the extraction resulting in an accuracy rate of 96.09% and 98.09% for each data set.

The World Health Organization (WHO) reveals that the number of people who have died due to COVID-19 has reached 6,108,976. Currently, the United States has the most fatalities, more than 1,005,056. COVID-19 detection is the first step that must be taken for the next crackdown. The detection is performed on chest X-rays containing the visual of the respiratory organs. The characteristics of the COVID-19 disease attacking the respiratory organs can serve as a reference for selecting the chest X-rays.

Classification using image processing has been carried out by researchers in health, especially for detecting COVID-19 using machine learning with image recognition methods using Bayesian networks and classification systems using SVM and KNN [2]. Then, there are COVID-19 symptom classification systems with KNN algorithms, neural networks, random forest, and Naive Bayes [3].

Many researchers have implemented machine learning to classify COVID-19 images. The images consisted of two types: X-ray and CT scan. The X-ray images were utilized by [1 - 19]. Meanwhile, the CT scan images were employed by [9], [14], [22], and [23].

Both Computerized Tomography (CT) scan and X-ray images were implemented by [3-4], [10], [12], and [21]. Several studies on COVID-19 used non images [22], [23]. Some utilized merely X-ray images, such as [2], [5], [8], [11], [13], [15], [16], and [19]. Subsequently, [1], [6], [7], [17], [18], and [20] used solely CT scan images.

For those using CT scan and X-ray images with machine learning and deep learning, the accuracy ranged from 76% to 99% [3]. A study gained 99% accuracy with the bagging tree classifier, with Light Gradient Boosting Machine (GBM) acquiring 98% accuracy [12]. Moreover, [21] obtained an accuracy of 89.41%, 99.02% and 98.11% from dataset-1 (CT), dataset-2 (X-ray) and dataset-3 (CT).

The research employed only X-ray image data and the FrMEM extraction system, with results of 96.09% and 98.09% for each data set [2]. Subsequently, for the two-class classification of the paper in [5], the accuracy, sensitivity, and specificity were, respectively, 100%, 100%, and 100% for COVID-19 vs. normal; 96.34%, 95.35% and 97.44% for COVID-19 vs. bacterial pneumonia; and 97.56%, 97.44% and 97.67% for COVID-19 vs. non-COVID-19 viral pneumonia. The combined accuracy and AUC were 79.52% and 0.87% for the multi-class classification. The study using Histogram of Gradient (HOG) and the SVM and KNN classification systems obtained 89.2% to 98.66% accuracy [11]. Moreover, the study applied SVM, with the accuracy reaching 94.12% [13]. Furthermore, the research utilized the Logistic Regression (LR) and the Convolutional Neural Networks (CNN) models and acquired the accuracy of 95.2% to 97.6% without Principle Component Analysis (PCA) and 97.6% to 100% with PCA [16]. Using only CT scan imagery as the data, other study obtained an accuracy of 99.68% with the Grey-Level Size Zone Matrix (GLSZM) extraction method [1]. In addition, [6] deployed SVM with Fused-Feature-Vector (FFV) and gained an accuracy of 89.80%. Moreover, the (Random Forest) RF classification system gained an accuracy of 96%, while the Recursive Feature Elimination (RFE) and RF combination obtained an accuracy of 97% and the ANOVA and RF combination acquired an accuracy of 94% [20].

Haar Wavelet algorithm has been used by several research in image processing to extract the texture features from the used images. The performance of the algorithm has been proven that it can be used to extract the features. Based on the literature review, this research come out to design the algorithm by using Haar Wavelet and machine learning to build the classification system. The system can be used as another method to classify the images using another perspective in the same research area.

II. METHODOLOGY

This research went through several stages, beginning with collecting image data from three different classes. Then, the image pre-processing was performed after the extraction by Haar wavelet transformation by dividing the second order into two, followed by the classification using SVM and KNN.

A. System Design

The system design was carried out to create a cystion flow to allow the classification to run well; the second order of system design is displayed in Fig. 1.

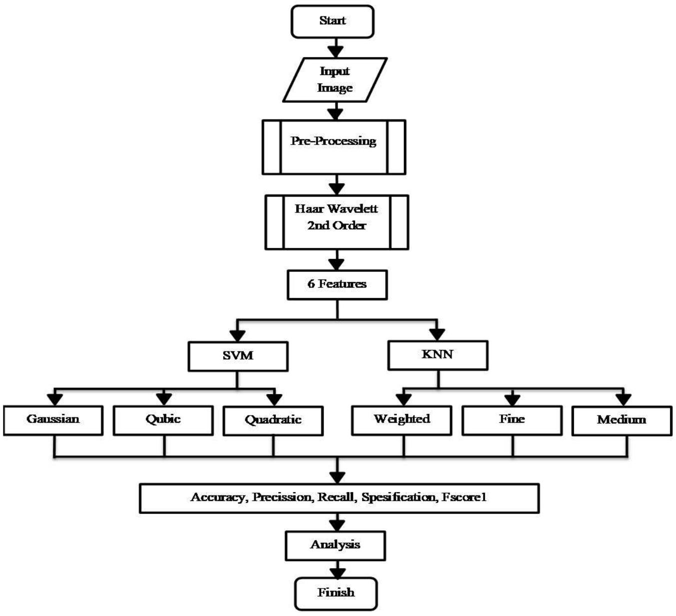


Fig. 1. Flowchart of System

The system designed in Fig. 1 applied Haar wavelet transformation extraction, where this study utilized the conditions in the second order.

B. Data and Device

The image data consisted of three classes: COVID-19, normal and pneumonia conditions. This data is obtained from the open source Github, where initially the images were obtained from various hospitals in various countries with details as described in Table I after preprocessing step.

TABLE I. TRAINING AND TESTING DATA

		Training	Testing
COVID-19	404	363	41
Normal	390	351	39
Pneumonia	390	351	39
Total	**1,184**	**1,065**	**119**

The device used in this research was Windows 10 with Intel(R) Core (TM) i5-9400 CPU @ 2.90GHz processor specifications, 64-bit type system and 16 GigaBytes of RAM.

C. Pre-Processing

The initial stage in pre-processing was changing the size of the image alignment to 300 x 300. Several images, specifically COVID-19, were rotated, flipped and converted from RGB to grayscale types as described in Fig. 2. The last stage of image enhancement was processing to improve the quality of image using histogram equalization algorithm. Due to limitation of total number for COVID-19 images, the rotated and flipped process (augmentation) are needed to multiply the number of images.

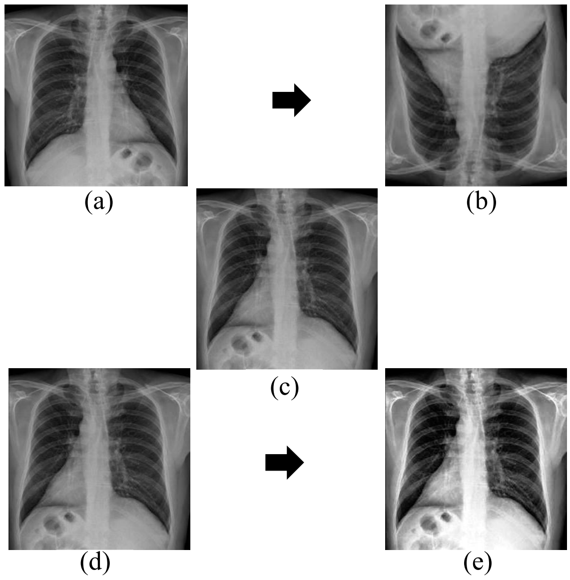


Fig. 2. Image pre-processing results: (a) resizing, (b) rotating, (c) flipping, (d) RGB to grayscaling and (e) histogram equalizing

D. Extraction System with Haar Wavelet Transformation

The extraction results generated by Haar wavelet transformation were leveled for this research. The data utilized the second order of the extraction, resulting in six features.

E. Classification Methods

The classification was divided into two, with SVM and KNN, and each was redivided into three different models. The SVM method comprised Gaussian SVM, Cubic SVM, and Quadratic SVM models, while the KNN method consisted of Weighted KNN, Fine KNN, and Medium KNN models.

TABLE II. AVERAGE AND STANDARD DEVIATION

Features	**Second-Order Class**		
	COVID-19	**Normal**	**Pneumonia**
Horizontal Coefficient	0.163 ± 0.09	0.137 ± 0.04	0.104 ± 0.04
Vertical Coefficient	0.158 ± 0.09	0.289 ± 0.08	0.237 ± 0.08
Diagonal Coefficient	0.029 ± 0.02	0.032 ± 0.01	0.030 ± 0.01
Horizontal Standard Deviation	16.59 ± 9.26	6.95 ± 7.30	13.68 ± 9.30
Vertical Standard Deviation	18.18 ± 11.10	12.54 ± 16.54	13.04 ± 11.67
Diagonal Standard Deviation	21.07 ± 14.73	12.65 ± 12.47	13.938 ± 11.49

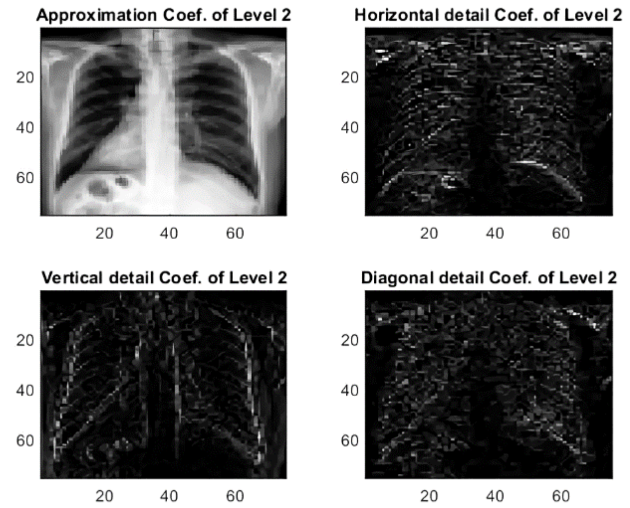


Fig. 3. Haar Wavelet features extraction results

III. Results and Discussions

A. Feature Extraction Results

The feature extraction results with the Haar wavelet transformation algorithm produced six features: horizontal coefficient, vertical coefficient, diagonal coefficient, horizontal standard deviation, vertical standard deviation, and diagonal standard deviation.

Table II describes the breakdown of the feature extraction results along with the average value and standard deviation of the COVID-19, normal, and pneumonia classes.

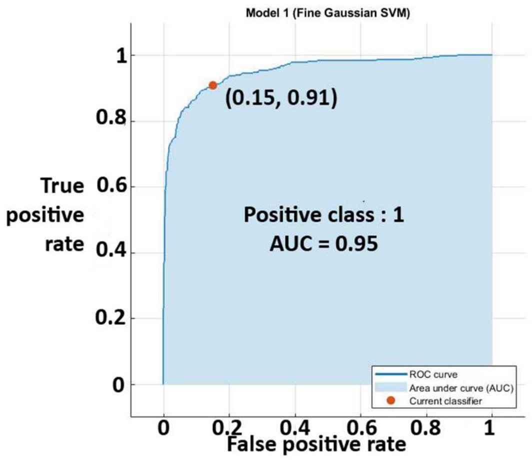
The use of Haar wavelet transformation generated visuals from the extraction levels with output in four main parts: approximation coefficient second order, vertical coefficient second order, horizontal coefficient second order, and diagonal coefficient second order. In the second order, the images had more complicated or brighter visual characteristics, as displayed in Fig. 3.

TABLE III. Training Classification Results

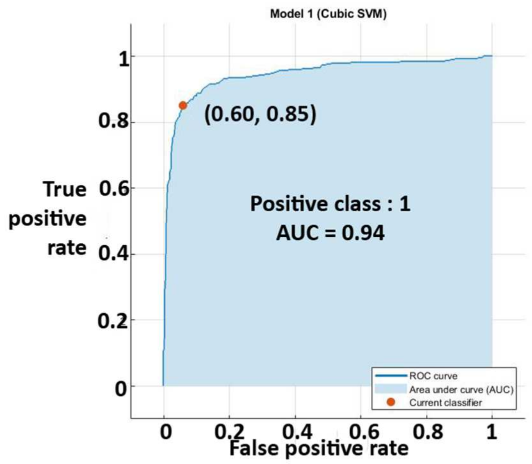
	Gaussian SVM		**Cubic SVM**		**Quadratic SVM**		**Weighted KNN**		**Fine KNN**		**Medium KNN**	
	Acc	**Time**	**Acc**	**Time**	**Acc**	**Time**	**Acc**	**Time**	**Acc**	**Time**	**Acc**	**Time**
Run 1	75.2%	1.10	77.0%	42.26	80.0%	10.62	78.3%	0.90	76.4%	0.62	76.9%	0.70
Run 2	76.6%	1.11	76.8%	48.43	78.7%	14.81	79.2%	0.56	**78.2%**	0.67	**77.5%**	0.75
Run 3	77.0%	1.14	77.5%	34.07	79.6%	10.48	79.1%	0.56	77.4%	0.66	76.7%	0.71
Run 4	75.5%	1.10	77.7%	41.69	79.9%	12.20	78.9%	0.58	76.4%	0.66	77.1%	0.72
Run 5	76.1%	1.12	77.3%	43.14	79.5%	9.72	79.2%	0.60	76.0%	0.65	76.3%	0.73
Run 6	76.1%	1.10	77.6%	38.36	80.1%	10.65	**79.6%**	0.63	77.3%	0.67	77.4%	0.72
Run 7	75.4%	1.12	76.5%	38.47	79.9%	11.17	78.8%	0.60	76.8%	0.67	77.3%	0.76
Run 8	76.2%	1.11	77.7%	39.05	**80.4%**	9.89	78.4%	0.61	77.6%	0.67	77.1%	0.72
Run 9	76.0%	1.12	77.7%	42.87	79.6%	10.31	78.5%	0.61	77.1%	0.81	76.7%	0.75
Run 10	**77.0%**	1.13	**77.7%**	37.52	79.8%	10.77	78.8%	0.62	76.7%	0.73	77.3%	0.73
Average	**76.1%**	**1.12**	**77.4%**	**40.59**	**79.8%**	**11.06**	**78.9%**	**0.63**	**77.0%**	**0.68**	**77.0%**	**0.73**

TABLE IV. Performances of Testing Results

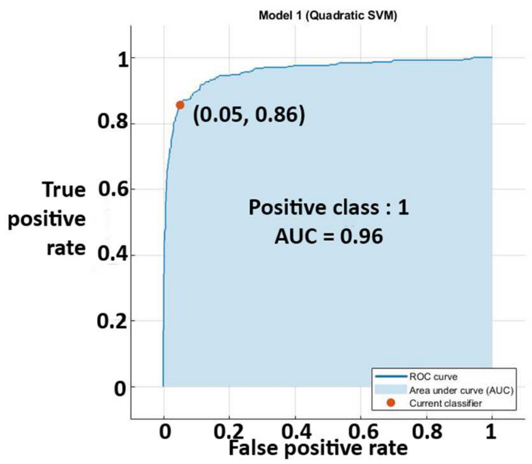
Performance Results of Testing									
Model		**COVID-19**	**Normal**	**Pneumonia**	**Model**		**COVID-19**	**Normal**	**Pneumonia**
Gaussian SVM	Accuracy	92%	86%	87%	Weighted KNN	Accuracy	93%	85%	88%
	Precision	92%	76%	85%		Precision	100%	71%	87%
	Recall	85%	88%	80%		Recall	83%	92%	79%
	Specificity	96%	85%	91%		Specificity	100%	83%	93%
	F-score	89%	81%	82%		F-score	91%	80%	83%
Cubic SVM	Accuracy	89%	85%	85%	Fine KNN	Accuracy	90%	84%	83%
	Precision	94%	72%	78%		Precision	89%	74%	81%
	Recall	76%	86%	79%		Recall	83%	85%	76%
	Specificity	97%	84%	88%		Specificity	94%	84%	88%
	F-score	84%	78%	78%		F-score	86%	79%	78%
Quadratic SVM	Accuracy	89%	83%	86%	Medium KNN	Accuracy	91%	82%	88%
	Precision	89%	70%	85%		Precision	94%	66%	92%
	Recall	80%	85%	77%		Recall	80%	89%	77%
	Specificity	94%	82%	91%		Specificity	97%	79%	96%
	F-score	85%	77%	80%		F-score	87%	76%	84%



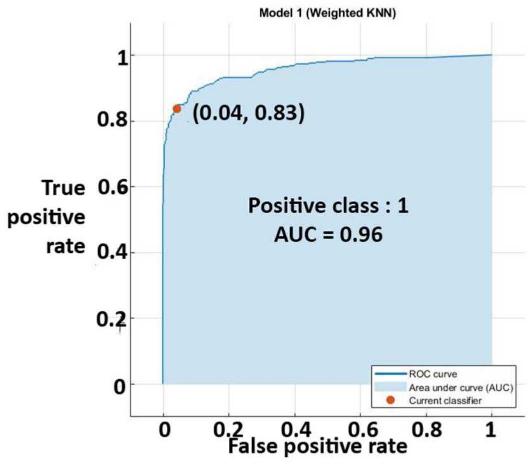
(a)



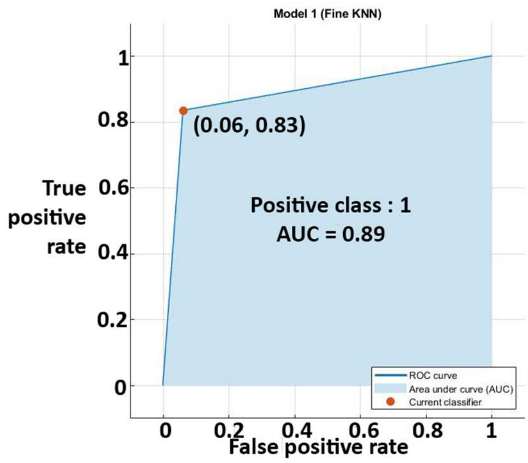
(b)



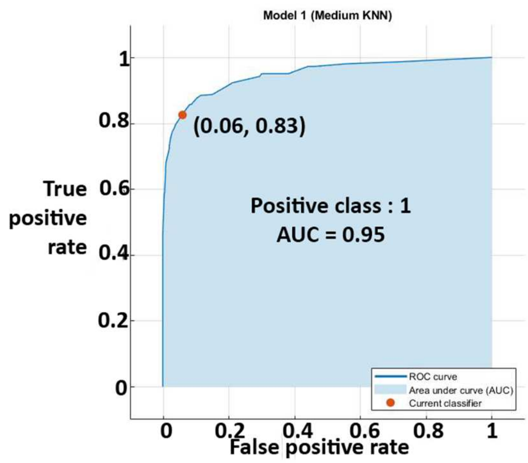
(c)



(d)



(e)



(f)

Fig. 4. ROC classification results of (a) Gaussian SVM, (b) Cubic SVM, (c) Quadratic SVM, (d) Weighted KNN, (e) Fine KNN, and (f) Medium KNN

B. Classification Results

The classification applied six models, including Gaussian SVM, Cubic SVM, Quadratic SVM, Weighted KNN, Fine KNN, and Medium KNN. Each model was run ten times to discover the comparison of results, both from its accuracy and processing time.

The overall classification results are summarized in Table III, describing accuracy and processing time. The overall accuracy for each model was not much different, ranging from 75% to 80%. However, the average processing time of KNN models was faster, with a range time of less than one second, as depicted in Table III. Weighted KNN acquired the fastest processing time of these six models. Nevertheless, the highest accuracy rate of the entire models was obtained by Quadratic SVM, with an average of 79.8%, as illustrated in Table III.

Receiver Operating Characteristics (ROC) was another classification result intended to review the performance of the classification results listed in Fig. 4; if the Area Under Curve (AUC) in the ROC has a scale of 0 to 1 with a full visual display of acres when the value is maximum. The results of ten runs on six models were 0.95 for Gaussian SVM, 0.95 for Cubic SVM, 0.96 for Quadratic SVM, 0.96 for Weighted KNN, 0.98 for Fine KNN, and 0.95 for Medium KNN. The AUC values were all close to 1, indicating a good classification.

C. Testing Results

Testing was conducted to check data shared using the algorithm stored from the training. This testing utilized 10% of the entire data, resulting in 119 images being extracted and used. Table IV describes the testing results encompassing accuracy, precision, recall, specification, and f-score. Weighted KKN acquired the highest accuracy of 93% in the testing.

IV. Conclusion

The classification for COVID-19 detection was performed on 1,065 X-ray images with Haar wavelet transformation and SVM and KNN. Quadratic SVM obtained the highest accuracy of 80.4% during training. Meanwhile, Weighted KNN acquired the highest accuracy of 93% during testing. As for other models, the accuracy range was not much different, ranging from 89% to 93%.

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] Barstugan, M., Ozkaya, U., and Ozturk, S.: ‘Coronavirus (COVID-19) classification using ct images by machine learning methods’, arXiv preprint arXiv:2003.09424, 2020

[2] Elaziz, M.A., Hosny, K.M., Salah, A., Darwish, M.M., Lu, S., and Sahlol, A.T.: ‘New machine learning method for image-based diagnosis of COVID-19’, Plos One, 2020, 15, (6), pp. e0235187

[3] Mohammad-Rahimi, H., Nadimi, M., Ghalyanchi-Langeroudi, A., Taheri, M., and Ghafouri-Fard, S.: ‘Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review’, Frontiers in cardiovascular medicine, 2021, 8, pp. 185

[4] Zhang, F.: ‘Application of machine learning in CT images and X-rays of COVID-19 pneumonia’, Medicine, 2021, 100, (36)

[5] Hussain, L., Nguyen, T., Li, H., Abbasi, A.A., Lone, K.J., Zhao, Z., Zaib, M., Chen, A., and Duong, T.Q.: ‘Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection’, Biomed Eng Online, 2020, 19, (1), pp. 1-18

[6] Kadry, S., Rajinikanth, V., Rho, S., Raja, N.S.M., Rao, V.S., and Thanaraj, K.P.: ‘Development of a machine-learning system to classify lung CT scan images into normal/COVID-19 class’, arXiv preprint arXiv:2004.13122, 2020

[7] Sharma, S.: ‘Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients’, Environmental Science and Pollution Research, 2020, 27, (29), pp. 37155-37163

[8] Somasekar, J., Kumar, P.P., Sharma, A., and Ramesh, G.: ‘Machine learning and image analysis applications in the fight against COVID-19 pandemic: Datasets, research directions, challenges and opportunities’, Materials Today: Proceedings, 2020

[9] Tuli, S., Tuli, S., Tuli, R., and Gill, S.S.: ‘Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing’, Internet of Things, 2020, 11, pp. 100222

[10] Chakraborty, S., and Mali, K.: ‘SUFMACS: A machine learning-based robust image segmentation framework for COVID-19 radiological image interpretation’, Expert Syst Appl, 2021, 178, pp. 115069

[11] Hasoon, J.N., Fadel, A.H., Hameed, R.S., Mostafa, S.A., Khalaf, B.A., Mohammed, M.A., and Nedoma, J.: ‘COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images’, Results in Physics, 2021, 31, pp. 105045

[12] Kassania, S.H., Kassanib, P.H., Wesolowskic, M.J., Schneidera, K.A., and Detersa, R.: ‘Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning based approach’, Biocybern Biomed Eng, 2021, 41, (3), pp. 867-879

[13] Khan, M.A.: ‘An automated and fast system to identify COVID - 19 from X - ray radiograph of the chest using image processing and machine learning’, Int J Imag Syst Tech, 2021, 31, (2), pp. 499-508

[14] Kwekha-Rashid, A.S., Abduljabbar, H.N., and Alhayani, B.: ‘Coronavirus disease (COVID-19) cases analysis using machine-learning applications’, Applied Nanoscience, 2021, pp. 1-13

[15] Montazeri, M., ZahediNasab, R., Farahani, A., Mohseni, H., and Ghasemian, F.: ‘Machine Learning Models for Image-Based Diagnosis and Prognosis of COVID-19: Systematic Review’, Jmir Med Inf, 2021, 9, (4), pp. e25181

[16] Rasheed, J., Hameed, A.A., Djeddi, C., Jamil, A., and Al-Turjman, F.: ‘A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images’, Interdisciplinary Sciences: Computational Life Sciences, 2021, 13, (1), pp. 103-117

[17] Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A.I., Etmann, C., McCague, C., and Beer, L.: ‘Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans’, Nature Machine Intelligence, 2021, 3, (3), pp. 199-217

[18] Wu, Z., Li, L., Jin, R., Liang, L., Hu, Z., Tao, L., Han, Y., Feng, W., Zhou, D., and Li, W.: ‘Texture feature-based machine learning classifier could assist in the diagnosis of COVID-19’, Eur J Radiol, 2021, 137, pp. 109602

[19] Zargari Khuzani, A., Heidari, M., and Shariati, SA: ‘COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images’, Sci Rep-Uk, 2021, 11, (1), pp. 1-6

[20] Shiri, I., Salimi, Y., Saberi, A., Pakbin, M., Hajianfar, G., Avval, A.H., Sanaat, A., Akhavanallaf, A., Mostafaei, S., and Mansouri, Z.: ‘Diagnosis of COVID-19 Using CT image Radiomics Features: A Comprehensive Machine Learning Study Involving 26,307 Patients’, medRxiv, 2021

[21] Saygılı, A.: ‘A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods’, Appl Soft Comput, 2021, 105, pp. 107323

[22] Khanday, A.M.U.D., Rabani, S.T., Khan, Q.R., Rouf, N., and Mohi Ud Din, M.: ‘Machine learning based approaches for detecting COVID-19 using clinical text data’, International Journal of Information Technology, 2020, 12, (3), pp. 731-739

[23] Muhammad, L., Algehyne, E.A., Usman, S.S., Ahmad, A., Chakraborty, C., and Mohammed, I.A.: ‘Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset’, SN computer science, 2021, 2, (1), pp. 1-13