Detection Of CT – Scan Lungs COVID-19 Image Using Convolutional Neural Network And CLAHE

Ronaldus Morgan James
Master Program of Informatics Engineering
Universitas Amikom Yogyakarta
Yogyakarta, Indonesia
ronaldus.1188@students.amikom.ac.id

Andi Sunyoto
Faculty of Computer Science
Universitas Amikom Yogyakarta
Yogyakarta, Indonesia
andi@amikom.ac.id

2020 3rd International Conference on Information and Communications Technology (ICOIACT) | 978-1-7281-7356-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICOIACT50329.2020.9332069

***Abstract* — Detecting COVID-19 is a significant task for medical professionals today because of its rapid spread. To overcome this problem, medical professionals have used various techniques and methods to detect to inhibit the proliferation of COVID-19. CT (Computed Tomography) Scan is currently the best method for detecting COVID-19. This diagnostic method is very accurate because it can see organs in three dimensions. However, this method requires a radiologist to detect the disease and requires a long time, which means it will cut valuable time for medical practitioners if a patient is sick. Therefore it is necessary to implement a system to detect the coronavirus automatically as an alternative quickly. This study intends to help medical practitioners to detect computed tomography (CT) Scans of lungs infected with COVID-19. The methods to be used are Limited Adaptive histogram equalization (CLAHE) contrast to improve the quality of CT (Computed Tomography) Scan images of COVID-19 lungs and Convolutional Neural Network (CNN) for the image classification process. The dataset used is 698 RGB images. This study uses three convolutional layers, 3 Max-Pooling layers, and two fully connected layers, resulting in 83.28% accuracy.**

Keywords — Detection, Coronavirus, CT (Computed Tomography) Scan, CLAHE, Convolutional Neural Network.

I. INTRODUCTION .

Coronavirus is a virus from the subfamily Orthocronavirinae in the Coronaviridae family and order Nidovirales. This group of viruses can cause disease in animals and humans. In humans, coronavirus causes respiratory tract infections [1]. In many cases, this viral infection causes mild respiratory infections such as flu, but this virus can also cause severe respiratory infections such as pulmonary infections (Pneumonia). A coronavirus infection called COVID19 (Corona Virus Disease 2019) first discovers in Wuhan City, China, at the end of December 2019. Viruses that have become serious health problems of this society are contagious very rapidly and have spread almost to the world [2], [3]. According to the data that has been issued by the Acceleration Task Force handling COVID-19 Indonesia, the number of cases positively confirm on 15 July 2020 was 78,572 people with a death count of 3,710 people. On the other hand, demographic conditions such as age, gender, and urban parameters such as temperature and humidity affect the prevalence of the disease in Indonesia and the world in the spread of viruses [4], [5]. Such an ordinary occurrence of coronavirus is not the first occurrence of the incident and consumes many casualties. In 2002, Severe Acute Respiratory Syndrome (SARS) was caused by SARS-coronavirus (SARS-Cov) in China. The Middle East Respiratory Syndrome (MERS), which took place in 2012 in Saudi Arabia and the Novel Coronavirus (Covid-19), appeared at the end of December 2019 until now [3].

In Indonesia, efforts have made to anticipate and reduce the number of coronavirus sufferers. Efforts are made one of them by providing policies to limit activities outside the home to cut the transmission of the virus. Another way to find out if someone is affected by COVID-19 is to do a test. Many types of tests can do to detect this coronavirus, but in the opinion of experts that CT (Computed Tomography) Scan can be the best method for detecting COVID-19. CT (Computed Tomography) Scan is a very accurate diagnostic tool because it can see the internal organs in three dimensions. The method of CT (Computed Tomography) Scan is said to be able to identify spots or white patches, so it considered to be a routine testing method to reduce the risk of coronavirus spreading.

After knowing which test is best for detecting COVID-19, another problem arises among medical practitioners. At present, the COVID-19 test is a rather difficult job because of the lack of testing tools [6]. Apart from the lack of availability of tools, to analyze using the method of CT (Computed Tomography) Scan requires a radiologist and takes a long time which means cutting time is very valuable for medical practitioners if there are patients who are sick and need help. Therefore, it is necessary to quickly implement a system for automatic detection of coronavirus as an alternative to assist testing tools and assist medical practitioners in particular in the field of radiology in diagnosing COVID-19 with CT (Computed Tomography) Scan.

This study intends to help medical practitioners to detect COVID-19 using CT-SCAN images of the lungs of COVID-19 patients. The study used the Contrast Limited Adaptive Histogram Equalization (CLAHE) method to increase the contrast of the CT-SCAN images of the lungs and the Convolutional Neural Network (CNN) to detect and classify the COVID-19 pulmonary CT-SCAN images. The study used a dataset that had been obtained previously on Github. The image used in this study is a CT (Computed Tomography) scan of a lung with COVID-19 and CT (Computed Tomography) of a normal lung.

II. RELATED LITERATURE

Previous research conducted experiments using Contrast Limited Adaptive Histogram Equalization (CLAHE) To improve image contrast and image classification using the Convolutional Neural Network (CNN) algorithm and produce good accuracy. The research uses the Convolutional Neural Network to classify and detect types of skin diseases. The study uses 11 layers, including a Convolution Layer, an

Activation Layer, Pooling Layer, Fully Connected Layer, and Soft-Max classifier. Datasets used in the database of Dermnet, which consists of four types of skin diseases: acne, keratosis, eczema herpeticum, and urticaria, with each class containing approximately 30 to 60 different samples. The study resulted in an accuracy of 98.6% [7]. Convolutional Neural Network to detect pneumonia disease using X-ray images. Research on the classification of pneumonia divides into two classes, namely pneumonia and non-pneumonia. The study aims to find the highest accuracy by changing the various parameters, Hyperparameters, and the number of convolutional layers. The highest accuracy resulted in 92.31% using three layers of convolutional [8]. In this research study, Convolutional Neural Network (CNN) develop to classify foam images collected from the flotation column of industrial coal operated under various process conditions (airflow rate, frozen dose, a slurry of solid%, froth depth and dosing collector). The study compares the classifications of the Convolutional Neural Network and Artificial Neural Network (ANN). The results showed that the foam classification system based on the Convolutional Neural Network significantly surpassed the Artificial Neural Network classifier in terms of classification accuracy and computational time. CNN can classify images of foam with an overall accuracy of 93.1% [9]. There have also been several studies using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method to increase the contrast of imagery. In the case of stroke disease detection with the Contrast Limited Adaptive Histogram Equalization (CLAHE) method, the contrast of the image using the Convolutional neural network and the number of classification images as many as 15 images resulted in 90% accuracy. [10], subsequent studies to classify microscopic breast imagery using Hybrid CLAHE and CNN architectures. The study used 7909 total images with a size of 700x460 pixels, using LeNet architecture by comparing accuracy when using Contrast Limited Adaptive Histogram Equalization (CLAHE) and which did not apply. Results of the study mentioned that the LeNet architecture using Contrast Limited Adaptive Histogram Equalization (CLAHE) resulted in an enlarged accuracy of 90%. In contrast, the LeNet without a Contrast Limited Adaptive Histogram Equalization (CLAHE) resulted in an accuracy of 86%. [11]

Some studies have also discussed the detection and classification, especially for the same question as to the writer, COVID-19. The first study classified COVID-19 using the X-ray image dataset. The image that is labeled as COVID-19, Normal, and SARS results in an accuracy of 95.12%[12]. Furthermore, research using images amounting to 25 pieces with the same source dataset [13] resulted in 95.38% and 93.8% accuracy [6], [14]

From previous research, there is a difference between the last study and research that will do, in particular, in research with the topic of COVID-19. Previous research using COVID-19 X-ray images for detection and classification while research will be carried out using CT (Computed Tomography) image of COVID-19 lung scans and adding a method of Contrast Limited Adaptive Histogram Equalization (CLAHE) to optimize the image of a CT (Computed Tomography) image of the COVID-19 lung Scan and apply the Convolutional Neural Network (CNN) that has adjusted to be able to measure the accuracy value of the model.

From the research done before, there is a difference between the previous research and the research that will be done next, especially in research with the topic COVID-19. Previous studies used COVID-19 x-ray images for detection and classification while the research that will carry out uses CT (Computed Tomography) COVID-19 lung scans and adds a Contrast Limited Adaptive Histogram Equalization (CLAHE) method to optimize images from CT images (CLAHE) Computed Tomography) COVID-19 lung scan and apply the Convolutional Neural Network (CNN) that has adjusted to being able to measure the accuracy of the model.

III. Proposed Method

In this study, the authors used qualitatively. Exploring and focusing on objectives is the aim of this research. Then the suitable research method used in qualitative research is experimental.

A. Dataset

In this case, COVID-19 lung CT scan images will be used as research. The image of the study obtained on the Kaggle website [15]. Not many people use this dataset because there was an initial debate about the authenticity of the images and sources of this data. Still, the usefulness of this dataset has been confirmed directly by senior radiologists at Tongji Hospital, Wuhan, China, who have diagnosed and treated a large number of COVID-19 patients during this disease outbreak between January and April. Therefore, the authors also believe and use this dataset. Lung CT-Scan image dataset consists of 2 labels with a total of 746 images with a jpg extension. The first label contains imagery lung CT scan image as much as normal with the number 349, and the second label includes a CT scan image of the lungs COVID-19 with a total of 397 pieces. So that research can be balanced, the authors use 349 each image for each label. for each label. Details on the number of datasets can be seen in Table 1

TABLE I. Detail Dataset

Samples	Number	Repository
CT CT scan of the COVID-19 lung	349	Kaggle (COVID-19 Lung CT Scans)
Normal CT images of the lungs	349	
Total	698	

All images in this dataset will be resized to 120x120 pixels. Here are Figures 1 and Figure 2, COVID-19 lung CT scan images, and normal lung scan images.

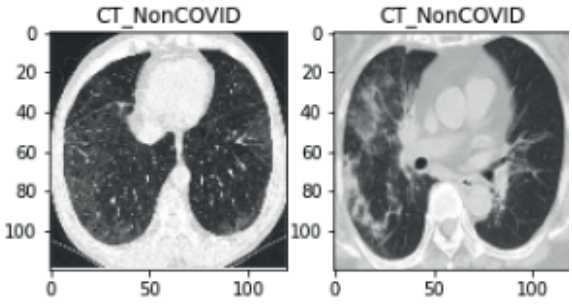


Fig 1. Lung CT scan normal

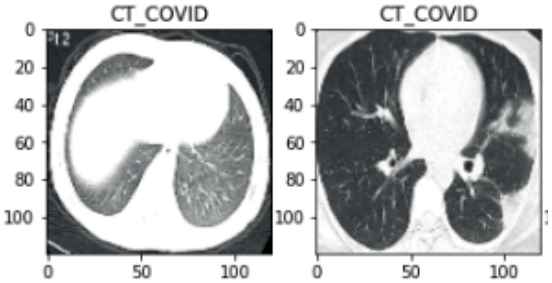


Fig 2. Lung CT scan COVID-19

B. Analysis Process

This section will show the entire process of research to carry out on detecting COVID-19 using a CT CT scan of the lungs. In the first step, after collecting the dataset, a preprocessing process will be carried out to extract the image. The second process undertaken is to develop a scenario for those who are looking for the best. In this research, the writer wants to use the number of epochs and dataset ratios as parameters. After the scenario was made, the data training process will be carried out. After the training process is complete, the data will continue with the data collection process to find which scenario is best. Each scenario will evaluate five times to get its accuracy, then the evaluation results from 5 trials will be taken an average value and made an accuracy value from the scenario. The following complete process can see in Figure 3 :

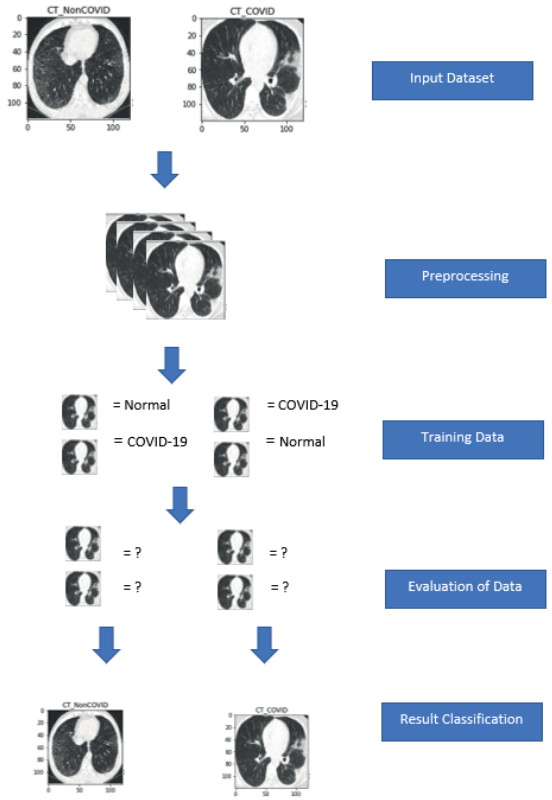


Fig 3. Classification Process

C. Preprocessing

Preprocessing is the stage where the stage to prepare the data will be carried out after the dataset is obtained. The purpose of data preparation is to be maximized when the classification process is carried out. The process that will be carried out is extracting the image using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method and changing the image's size to be uniform.

1) Contrast Limited Adaptive Histogram Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a method that aims to reduce noise and also increase image contrast. CLAHE is also a better version than the previous process, AHE (Adaptive Histogram Equalization). CLAHE determines the kernel matrix and works to replace the intensity value of each pixel of the input image with the average cost of the kernel weighting for each neighboring pixel and the pixel itself. In the process of reducing noise, kernel size also affects obtaining good image results. The kernel size used in this study is 3x3. In increasing contrast, the CLAHE image gives the boundary value on the histogram. The limit value is called the clip limit. How to calculate the clip limit on a histogram can be seen from equation (1)

$$\beta = \frac{M}{N}\left(1 + \frac{\alpha}{100}(s_{max} - 1)\right) \quad (1)$$

Variable M represents the area size, N represents the grayscale value (256), and α is a clip factor that states the addition of a histogram that is between 0 to 100.

Following is the CT scan of the lungs; it can be seen that the comparison of Figure 4 is the image before CLAHE and figure 5 after the CLAHE process.

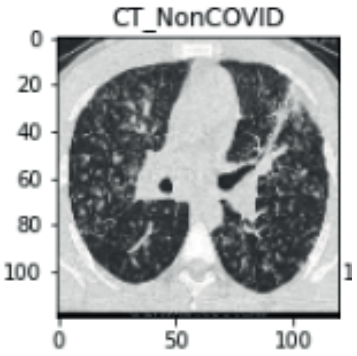


Fig 4. Image before CLAHE

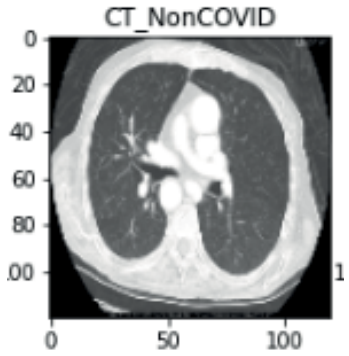


Fig 5. Image after CLAHE

2) Resize

The resize process is done because the data collected does not have the same size and image size that is too large. Resize also aims to adjust the size of the image to the size of the system input data. In this study, the image will resize to 120x120 pixels.

D. Classification

In this classification process, several preparations were made, starting from the creation of a Neural Network Convolutional (CNN) architecture, the process of training data to evaluating data. In the installation architecture, the authors look for which architecture is right by changing several parameters, such as the number of epoch and ratio dataset, to find which architecture is the best. Continue to enter the training and evaluation process. At the architecture installation stage, a number of parameters have changed for the experiment and some parameters that have set the same. Following is the architecture that will use in the training and test process can see in Figure 6.

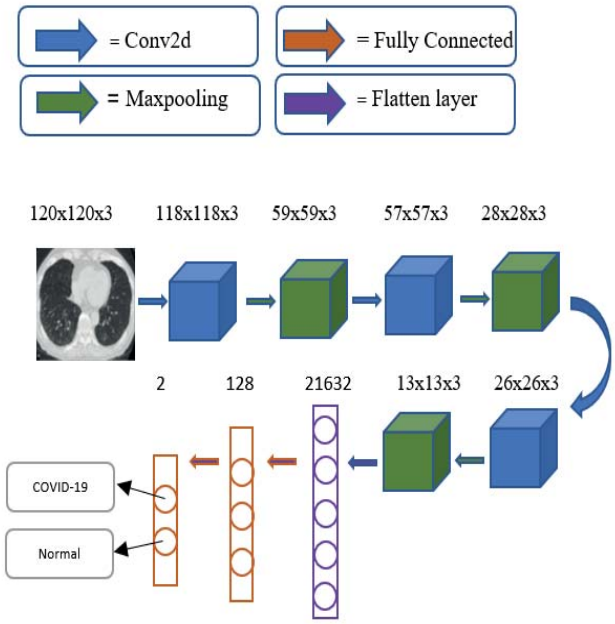


Fig 6. Classification architecture

After designing the architecture that is considered the best, the next research carries out a test process. This testing process is carried out to train the architecture with CT-Scan image data that has labeled before. This process also aims to find out how well the accuracy is generated before conducting the test process. In the test process itself, several experiments will be carried out by combining fixed parameters and parameters that will be changed to find the best architecture that produces the best accuracy for detecting COVID-19 with pulmonary CT-Scan images.

IV. Result and Evaluation

In this research, the writer will test the architecture that has made. Test data has extract using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method, and a classification architecture creates using the Convolutional Neural Network (CNN). The parameters that will try in the scenario are the epoch's size and changes in the ratio of the dataset to the training and testing process. In addition to testing and finding the best accuracy in classification, the authors also want to determine whether the epoch and dataset ratio will affect the accuracy. The parameter that will be changed and tested is epoch with a size of 25, 50, and 75. Then the ratio of the dataset will be adjusted based on the number of epochs. The ratio of the dataset to be tried is 80:20 and 70:30. The following detailed scenarios that will try to find the best accuracy value will display in Table 2

TABLE 2. DETAIL OF SCENARIO

Scenario	Epoch	Dataset Ratio
S 1	25	80:20
S 2	25	70:30
S 3	50	80:20

S 4	50	70:30
S 5	75	80:20
S 6	75	70:30

After making six scenarios, further research will go into the training data. The process of preparing for the sake of training each architecture in the scenario so that the evaluation installs produces good results. The following are the results of the training data which can see in Figure 7:

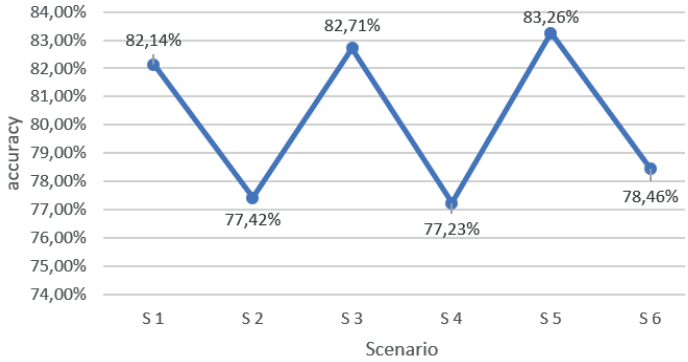
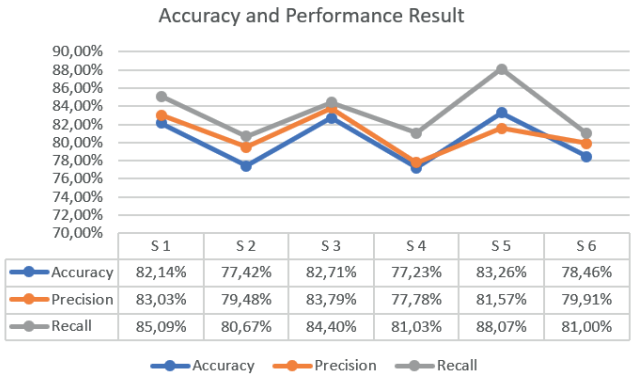


Fig 7. Training Accuracy

From the results of the training, scenario 5, with the number of epochs is 75, and the dataset ratio is 80:20 to get the maximum results. Perskenario is carrying out five times each training process to get accuracy. The study looked for the average value of the five accuracy values from each scenario to be the final result for training accuracy.

After conducting the training process, architecture can now evaluate. The evaluation process is also carried out five times per scenario and will take an average of accuracy. This evaluation process aims to find out how well a scene can distinguish COVID-19 CT images from normal. The higher the accuracy, the better a model can automatically detect image datasets. In addition to searching for accuracy, research also looks for performance from the results of detection. To find the best performance, researchers use the Confusion Matrix. This method provides information comparing the results of the classification of the model with the actual classification. To measure the performance of the confusion matrix model, several performance models are usually used. In this study, precision and recall will use. The following results from the calculation of accuracy and performance results can see in Figure 8:



	S 1	S 2	S 3	S 4	S 5	S 6
Accuracy	82,14%	77,42%	82,71%	77,23%	83,26%	78,46%
Precision	83,03%	79,48%	83,79%	77,78%	81,57%	79,91%
Recall	85,09%	80,67%	84,40%	81,03%	88,07%	81,00%

Fig 8. Evaluation and performance result

Based on experiments that have done can show that the scenario in scenario 5 with the number of epoch 75 and dataset ratio of 80: 20 produces the highest test compared to another scenario. The results obtained for scene 5 are 83.28% accuracy, 81.57% precision, and 88.07% recall.

V. CONCLUSION

Based on research, implementation, and testing, the following conclusions can draw:

In the research that has done, it can be concluded that image extraction using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method and the Neural Network Convolutional (CNN) algorithm uses 3 convolutional layers, 3 layer pooling, and 2 fully connected layers can detect CT-Scan images of the COVID lungs -19 pretty well. This study also seeks to know whether the number of epochs and dataset ratios can affect the detection results. That research conduct using six different scenarios based on the number of epochs and datasets. This study concludes that the number of epochs and dataset ratios can affect the results of accuracy. Scenarios that use a dataset ratio of 80:20 and an epoch of 75 get the best results with an accuracy of 83.28%, precision 81.57%, and recall 88.07%.

REFERENCES

[1] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics Med. Unlocked*, vol. 19, p. 100360, 2020, doi: 10.1016/j.imu.2020.100360.

[2] L. Yan *et al.*, "Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan," *medRxiv*, p. 2020.02.27.20028027, 2020, doi: 10.1101/2020.02.27.20028027.

[3] 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.

[4] B. Pirouz, S. S. Haghshenas, S. S. Haghshenas, and P. Piro, "Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of Coronavirus) through a binary classification using artificial intelligence and regression analysis," *Sustain. (United States)*, vol. 12, no. 6, 2020, doi: 10.3390/su12062427.

[5] B. Pirouz, S. S. Haghshenas, B. Pirouz, S. S. Haghshenas, and P. Piro, "Development of an assessment method for investigating the impact of climate and urban parameters in confirmed cases of COVID-19: A new challenge in sustainable development," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, 2020, doi: 10.3390/ijerph17082801.

[6] P. Kumar and S. Kumari, "Detection of coronavirus Disease (COVID-19) based on Deep Features," *Https://Www.Preprints.Org/Manuscript/202003.0300/V1*, no. March, p. 9, 2020, doi: 10.20944/preprints202003.0300.v1.

[7] T. Shanthi, R. S. Sabeenian, and R. Anand, "Automatic diagnosis of skin diseases using convolution neural network," *Microprocess. Microsyst.*, vol. 76, p. 103074, 2020, doi:

10.1016/j.micpro.2020.103074.

[8] R. Jain, P. Nagrath, G. Kataria, V. Sirish Kaushik, and D. Jude Hemanth, "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," *Meas. J. Int. Meas. Confed.*, vol. 165, p. 108046, 2020, doi: 10.1016/j.measurement.2020.108046.

[9] M. Zarie, A. Jahedsaravani, and M. Massinaei, "Flotation froth image classification using convolutional neural networks," *Miner. Eng.*, vol. 155, no. April, p. 106443, 2020, doi: 10.1016/j.mineng.2020.106443.

[10] J. T. Marbun, Seniman, and U. Andayani, "Classification of stroke disease using convolutional neural network," *J. Phys. Conf. Ser.*, vol. 978, no. 1, 2018, doi: 10.1088/1742-6596/978/1/012092.

[11] A. Vidyarthi, J. Shad, S. Sharma, and P. Agarwal, "Classification of Breast Microscopic Imaging using Hybrid CLAHE-CNN Deep Architecture," *2019 12th Int. Conf. Contemp. Comput. IC3 2019*, no. September, pp. 1–5, 2019, doi: 10.1109/IC3.2019.8844937.

[12] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," 2020, [Online]. Available: http://arxiv.org/abs/2003.13815.

[13] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 Image Data Collection: Prospective Predictions Are the Future," no. February, 2020, [Online]. Available: http://arxiv.org/abs/2006.11988.

[14] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0276-2.

[15] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, and P. Xie, "COVID-CT-Dataset: A CT Scan Dataset about COVID-19," pp. 1–14, 2020, [Online]. Available: http://arxiv.org/abs/2003.13865.