j Med Imaging(ScholarGen) Volume 4(2021) 01, pp. 25-32 ISSN : 2635-4608

DOI http://dx.doi.org/10.31916/sjmi2021-01-03

Spatial Frequency Analysis by DWT of CXR in the COVIDGR Dataset

[1]Giljae Lee,[2,]Gyehwan Jin, [1]Taesoo Lee*

,Received : 20 July 2021 / Accepted : 15 November 2021 / Published online: 28 December 2021

Abstract In this study, chest X-ray images from the "COVIDGR Dataset" were formalized, and both normal images and disease severity images (mild, moderate, and severe) were extracted by the discrete wave transform (DWT) method. The characteristics of Approximation, Horizontal, Vertical, and Diagonal were analyzed as normal, mild, and vector frequencies. The standardized pixel was a format of 512 x 512 images and was intended to include all lung fields, except the original image did not include the apex of the lung. The standardized images were extracted by DWT for the characteristics of Approximation, Horizontal, Vertical, and Diagonal and stored as COVIDGR-19 Feature Data. Theses data were entered into an Excel file to calculate the sum, and the average value was obtained and confirmed as the "Feature value of images by disease." Experiments showed that images of severe disease had little or no Approximation, high Diagonal content, and similarity between the Horizontal and Vertical characteristics. Normal images had the largest proportion of Approximation because there was no change in frequency due to the absence of disease, while mild and moderate images were similar in Approximation, Horizontal, and Vertical but had a one-sided bias in small areas for Diagonal. It is hoped that the results of this study can be used as an important parameter in an automatic identification system of chest X-ray images to allow timely diagnosis of patients with COVID-19 and other lung conditions.

[1]Giljae Lee,[2,]Gyehwan Jin(✉), [1]Taesoo Lee*

[1]Dept. of Biomedical engineering, Graduate school of Chungbuk National University, Chungbuk, Korea

[2,]Dept. of Radiology, Nambu University, Gwangju, Korea*

Key words: COVID-19, Chest X-ray, DWT, Feature extraction, Machine learning

I. Introduction

The World Health Organization (WHO) declared a global pandemic [1] on March 11, 2020, after the coronavirus disease 2019 (COVID-19) epidemic in China spread beyond a specific regional window to more than two continents [2]. Since establishment of the WHO in 1948, only two other pandemics have been declared: the "Hong Kong Flu" pandemic in 1968, and the "Influenza A virus subtype H1N1" pandemic in 2009 [3].

An unknown respiratory epidemic began in December 2019, in Wuhan, China, and gradually spread to all regions and countries of the world. On January 9, 2020, the WHO declared that it was a new strain of coronavirus officially named COVID-19 [4]. Throughout 2020, the pandemic occurred in developed countries, such as continental Europe and the United States, as well as around the world, with a total of 81,997,998 infections and 1,789,927 deaths as of December 31, 2020. [5]. In 2021, the coronavirus developed mutations named Beta, Gamma, and Delta; Delta mutations became the

predominant strain worldwide in the summer of 2021. Academic organizations are distributing research to overcome the COVID-19 pandemic, and universities, public research institutes, and pharmaceutical and biotechnology companies are working together to develop diagnostic methods, vaccines, and treatments. To date, diagnosis of COVID-19 has been verified using one of the three following methods [7]:

Computed tomography (CT)-based evaluation: Three-dimensional images obtained from various angles can be analyzed, but small medical institutions often do not have these devices available. CT also requires more than 15 minutes of testing and has the disadvantage of radiation exposure.

Reverse transcription polymerase chain reaction (RT-PCR) examination: RT-PCR is a method of detecting viral RNA in sputum or non-pharyngeal swab specimens. Specific materials and equipment frequently are not accessible, and testing requires 12 hours. This method has the limitation of needing an extended time for testing, which is not advantageous when positive patients must be identified as soon as possible. In addition, some studies have shown that RT-PCR results can produce false negatives in the same patient. If the RT-PCR test is positive, additional tests, such as CT scanning, might be required for confirmation.

Chest X-ray (CXR): A method that requires less auxiliary equipment, involves lighter equipment, and can be administered using mobile medical devices. In general, CXR testing is an easy way to assess patients prior to RT-PCR and CT-scan testing. CXR is the most efficient method of testing for COVID-19 in terms of both time and cost.

Automatic determination of medical images by machine learning requires images to be acquired in a form that is stereotyped by image acquisition equipment and data mining (i.e., preprocessing) through feature extraction of acquired images [8]. It takes many steps to create models that incorporate pretreated image data (extracted feature values) and clinical data and to build an evidence-based decision tree.

In this study, the CXR images of the “COVIDGR Dataset” were formalized. Normal images and COVID-19 patient images (mild, moderate, and severe) were extracted by discrete wavelet transformation (DWT) to characterize the low-frequency region, horizontal high-frequency region, vertical high-frequency region, and diagonal high-frequency region. It is believed that the results of this study can be used as an important parameter for an automatic identification system of CXR images that will be available in the future.

Ⅱ. Materials and methods

1. COVIDGR Dataset

The COVIDGR Dataset is part of a homogeneous and balanced database, COVIDGR-1.0, that includes positive RT-PCR patients who have been classified with mild, moderate, or severe COVID-19 and was developed in collaboration with the Hospital Universitario San Cecilio in Granada, Spain [8]. COVIDGR-1.0 contains 426 positive posteroanterior (PA) CXR images and 426 negative PA CXR images [7].

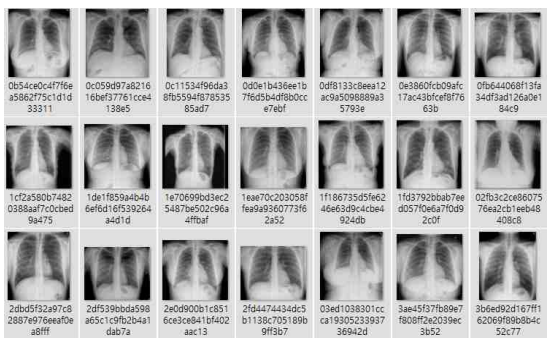

Figure 1. A sample of the negative (normal) images in the COVIDGR-1.0 Dataset

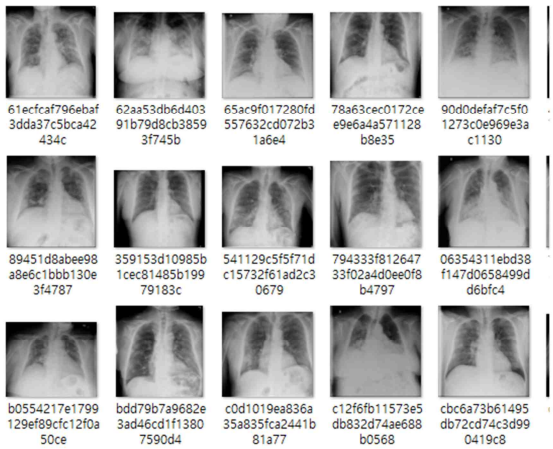

Figure 2. A sample of positive (severe) images in the COVIDGR-1.0 Dataset

2. Extracting characteristics of chest X-ray images

The Matlab wavelet toolbox was used to extract features of CXR images. A major challenge of wavelets is that time-frequency interpretation can be variable at the same time as the characteristics of Fourier's interpretation of signals in the frequency domain. Because Fourier transforms do not have time information and cannot be used for time-frequency analysis, STFTs that introduce the time window concept for Fourier transforms are available. However, in the time-frequency analysis of STFT, there is no coherence of the underlying data, so sensitivity to singularity is reduced. Thus, a method to conduct a time-frequency analysis without compromising the toposociality of the base has been devised, which is the wavelet transformation [8].

The defining characteristics of time-frequency analysis by wavelet transformation are that the time resolution is high in the high-frequency domain and the frequency resolution is high in the low-frequency domain. Therefore, wavelet transformation is more effective in time-frequency analysis than is STFT because the rapidly changing signal's position at the time of the change is important, and the period or frequency of the change is important for the gently changing signal [9].

The dwt2 command provided in the Matlab wavelet toolbox performs single-level, two-dimensional wavelet decomposition for specific wavelets or specific wavelet decomposition filters (Lo_D and Hi_D).

[cA,cH,cV,cD] = dwt2 (X, "wname") contains the wavelet name in the input matrix X and the string "wname".

[cA,cH,cV,cD] = dwt2 (X,Lo_D,Hi_D) calculates the two-dimensional wavelet decomposition based on the specified wavelet decomposition filter, as shown above. Here, Lo_D is a decomposition low-pass filter, and Hi_D is a decomposition high-pass filter, and the two are equal in length. Figure 3 shows an example of the wavelet transformation of the input images.

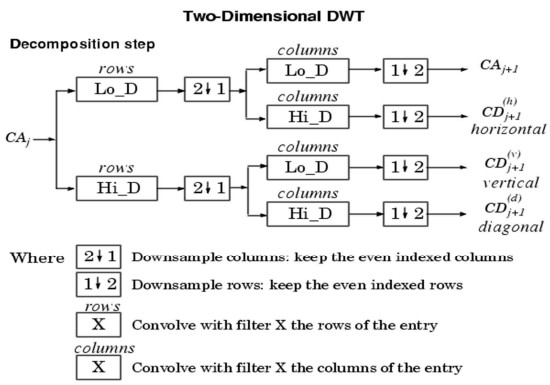

Figure 3. An illustration of 2D DWT

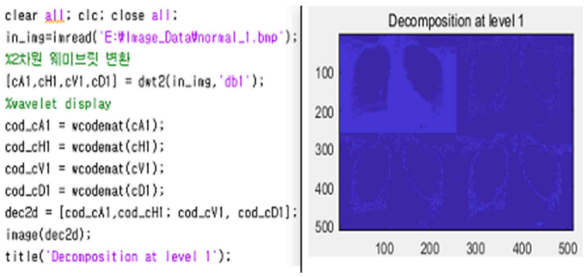

Figure 4. An example of wavelet decomposition

III. Experiment and Results

The final goals of this study were to standardize aspects, to extract features by DWT using the COVID-19 dataset, and to establish a system for automatic determination of the presence or absence of COVID-19 lung disease through mechanical learning combined with clinical data.

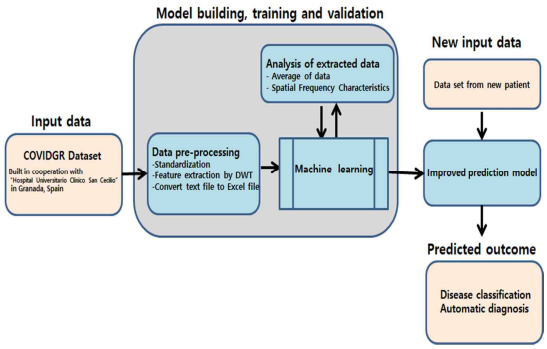

Figure 5. The research concept and final goal

In this paper, we extracted the features shown in Figure 6, stored them as a COVID-19 feature dataset, and analyzed their spatial frequencies.

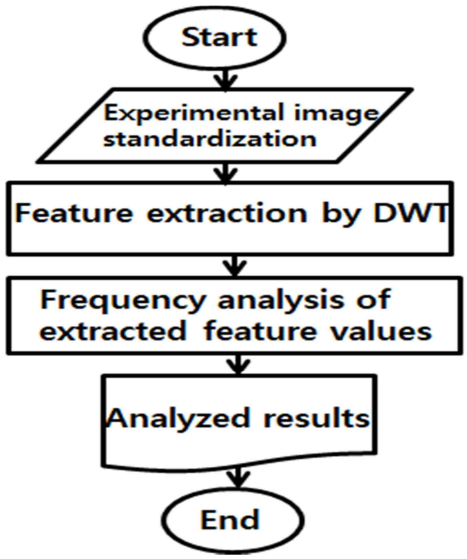

Figure 6. A flowchart of the feature extraction and spatial frequency analysis process

1. *Image standardization*

The COVIDGR Dataset was assembled in collaboration with the Hospital Universitario San Cecilio in Granada, Spain, from the homogeneous and balanced COVIDGR-1.0 Database, which includes images of patients confirmed by RT-PCR as COVID-19-positive and ranked in severity as mild, moderate, or severe. However, many images obtained early in the COVID-19 pandemic were derived from basic CXRs. Therefore, the images in the COVIDGR Dataset were standardized according to the basics of chest imaging. Figure 7 shows examples of original and standardized images.

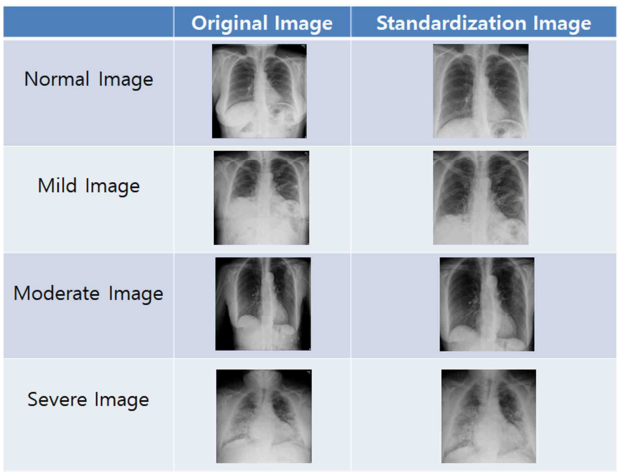

Figure 7. Examples of standardization images

2. *Feature extraction*

The standardized images were extracted by DWT, and the average feature values obtained by disease severity by summing the feature values of the extracted images are shown in Tables 1 through 4.

Table 1. Feature extraction values of normal images

Approximation	-0.10938	-0.07842	-0.05878	-0.03612	-0.02478	-0.01041	0.007	0.023581	0.028853	0.037144	0.049475	0.058997	0.073559	0.085291	0.097234	0.109373
Horizontal	-0.1028	-0.10452	-0.10561	-0.10687	-0.1075	-0.1083	-0.10927	-0.11019	-0.11048	-0.11094	-0.11157	-0.11216	-0.11291	-0.11362	-0.11428	-0.11496
Vertical	-0.10162	-0.10134	-0.10115	-0.10095	-0.10085	-0.10071	-0.10055	-0.1004	-0.10035	-0.10028	-0.10017	-0.10007	-0.09995	-0.09984	-0.09973	-0.09962
Diagonal	1.23E-06	6.25E-07	-1.9E-06	-1.3E-06	-1.9E-06	-1.9E-06	-1.3E-06	-1.3E-06	-6.3E-07	-1.3E-06	-1.3E-06	-1.3E-06	-2.5E-06	-2.5E-06	-1.3E-06	-1.3E-06

Table 2. Feature extraction values of mild images

Approximation	-0.05	-0.03585	-0.02687	-0.01651	-0.01133	-0.00476	0.0032	0.01078	0.01319	0.01698	0.02216	0.02697	0.03317	0.03899	0.04445	0.05
Horizontal	-0.04879	-0.04896	-0.04907	-0.04919	-0.04926	-0.04934	-0.04943	-0.04952	-0.04955	-0.0496	-0.04966	-0.04972	-0.0498	-0.04987	-0.04993	-0.05
Vertical	-0.04857	-0.04858	-0.04858	-0.04859	-0.0486	-0.0486	-0.04861	-0.04861	-0.04861	-0.04862	-0.04862	-0.04862	-0.04863	-0.04863	-0.04864	-0.04864
Diagonal	0	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002

Table 3. Feature extraction values of moderate images

Approximation	-0.5	-0.3585	-0.2687	-0.1651	-0.1133	-0.0476	0.032	0.1078	0.1319	0.1698	0.2216	0.2697	0.3317	0.3899	0.4445	0.5
Horizontal	-0.45801	-0.45748	-0.4571	-0.45671	-0.45651	-0.45625	-0.45593	-0.45564	-0.45555	-0.45541	-0.4552	-0.45502	-0.45478	-0.45455	-0.45434	-0.45412
Vertical	-0.45023	-0.44893	-0.44811	-0.44718	-0.44669	-0.44609	-0.44536	-0.44467	-0.44444	-0.4441	-0.44362	-0.44318	-0.44263	-0.44209	-0.44158	-0.44108
Diagonal	0.00001	0.00001	0.00001	0.00001	0	0.00001	0.00001	0.00001	0.00001	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002

Table 4. Feature extraction values of severe images

Approximation	-0.4	-0.2868	-0.21496	-0.13208	-0.09064	-0.03808	0.00256	0.08624	0.10552	0.13584	0.17728	0.21576	0.26536	0.31192	0.3556	0.4
Horizontal	-0.39633	-0.40273	-0.40677	-0.41145	-0.4138	-0.41677	-0.42037	-0.4238	-0.42486	-0.42658	-0.42893	-0.43109	-0.4339	-0.43653	-0.43889	-0.44152
Vertical	-0.39422	-0.3946	-0.39484	-0.39512	-0.39525	-0.39542	-0.39563	-0.39584	-0.39591	-0.396	-0.39614	-0.39626	-0.39641	-0.39658	-0.39674	-0.39687
Diagonal	0.00002	0.00002	0.00002	0.00004	0.00005	0.00005	0.00007	0.00008	0.00007	0.00009	0.00009	0.00009	0.00009	0.0001	0.0001	0.00012

3. *Spatial frequency analysis*

The average feature values extracted by disease were analyzed to determine the Approximation, Horizontal, Vertical, and Diagonal frequency characteristics. Figure 8 shows the low-frequency region of Approximation by disease severity. Normal images cover a large area; the mild and moderate images are similar, and the severe images do not appear at all.

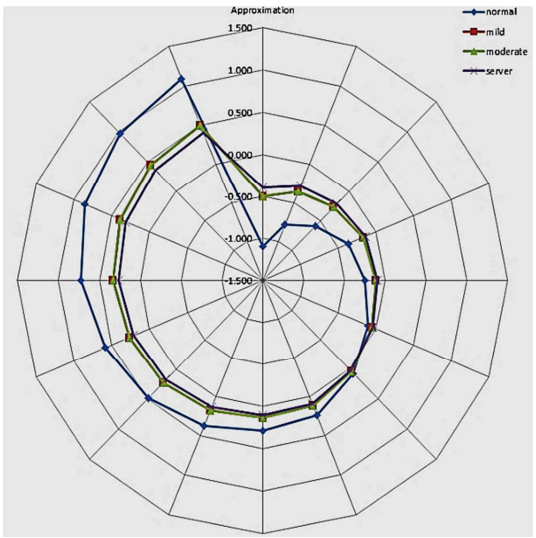

Figure 8. Spatial frequency analysis of the approximation data

Figure 9 shows the Horizontal section by disease. The normal images cover a large area. The mild and moderate images are similar, and the severe images represent very small areas.

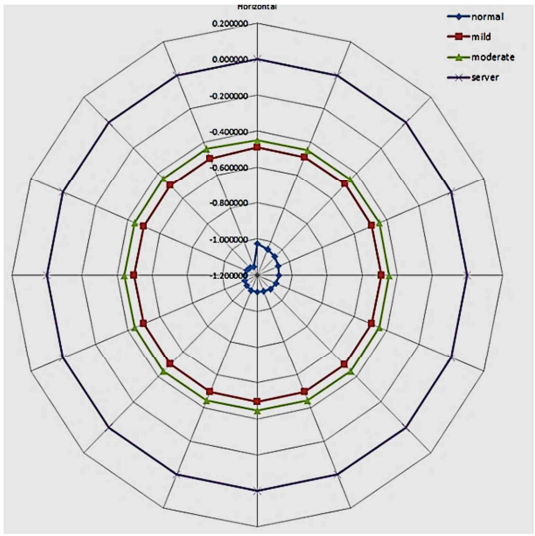

Figure 9. Spatial frequency analysis of the Horizontal data

Figure 10 shows the Vertical section by disease severity. The normal images and the mild and moderate severity images are similar, and severe images represent very small areas.

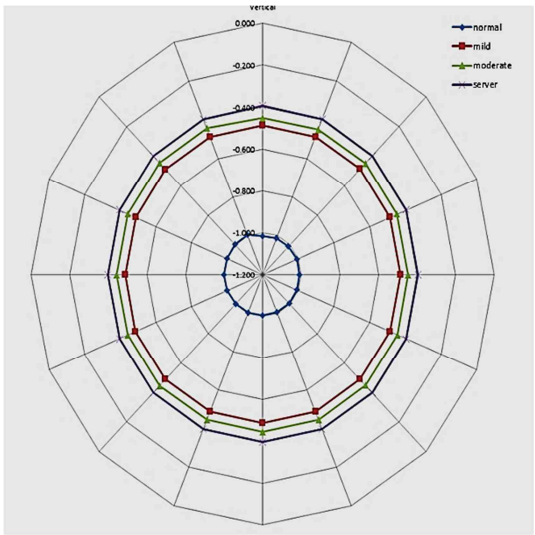

Figure 10. Spatial frequency analysis of Vertical data

Figure 11 shows the Diagonal characteristics by disease severity. The severe images represent a wide range, while the normal and moderate images are

similar. The mild images represent small areas and are skewed to one side.

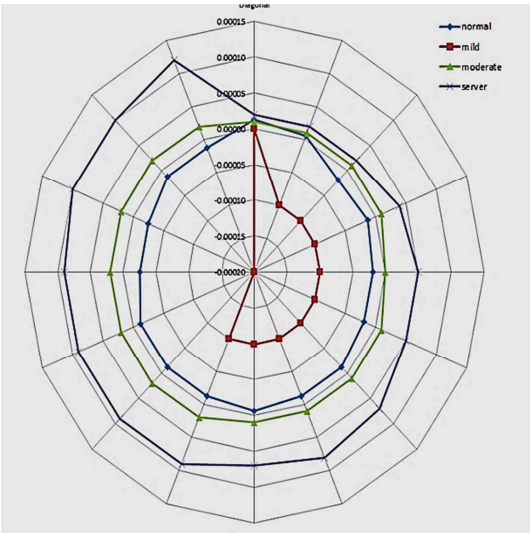

Figure 11. Spatial frequency analysis of the Diagonal data

IV. Discussion

Since the COVIDGR Dataset was obtained in urgent situations, there was a departure from the basic methods of breast imaging. Therefore, the images in the COVIDGR Dataset have been standardized to adhere to the basic methods of chest imaging (Figure 7). The standardized pixel format included 512 x 512 images and was intended to include all lung fields, except that the original image did not include the lung apex part. The standardized images were saved by DWT as the characteristics of Approximation, Horizontal, Vertical, and Diagonal are extracted and COVIDGR-19 Feature Data[10][11]. The COVIDGR-19 Feature Data were entered into an Excel file to calculate the sum, and the average value was obtained and confirmed as the "Feature value of images by disease.[11][12]" A frequency (Approximation, Horizontal, Vertical, Diagonal) analysis was performed [10].

Approximation analysis revealed that normal images occupy a large area; mild and moderate images are similar, and severe images are not shown at all. Severe images are likely due to severe pneumonia, and most of the characteristic values are high-frequency (video-changing) components.

The Horizontal images have a large area. The mild and moderate images are similar, and the severe images represent very small areas, resulting in similar results to the Approximation analysis.

The Vertical images had similar results as the normal images, and the mild and moderate images represented very small areas.

The Diagonal shows a wide range of severe images. The normal and moderate images are similar, while the mild images represent a small area, and the image appears to be skewed to one side.

V. Conclusion

In this study, the CXR images in the COVIDGR Dataset were formatted, and the features of the normal images and disease images (mild, moderate, and severe) were extracted by DWT in an approximation, horizontal, vertical high frequency region (Vertical) and a diagonal post-frequency region (Diagonal).

The standardized format used 512 x 512 images and was intended to include all lung fields, except that the original image did not include the apex of the lung. The standardized images were extracted by DWT for the characteristics of Approximation, Horizontal, Vertical, and Diagonal and were stored as COVIDGR-19 feature data. The COVIDGR-19 feature data were entered into an Excel file to calculate the sums, and the average values were obtained and confirmed as the "Feature value of images by disease."

A frequency (Approximation, Horizontal, Vertical, and Diagonal) analysis of feature extraction values by disease severity produced the following results.

Approximate: Normal images occupy a large area. Mild and moderate images are similar, and severe images are not present.
Horizontal: Regular images cover large areas. Mild images and moderate images are similar, and severe images are very small.
Vertical: Normal images and mild and moderate images are similar, while the severe images occupy very small areas.
Diagonal: The severe images were broad in scope. The normal images and the moderate images were similar, and the mild images showed a small, one-sided area.
Studies have shown that severe images have little or no approximation, a high diagonal content, and similar horizontal and vertical values. Normal images traditionally have the largest proportion of approximation because there is little change in frequency due to the absence of disease. The mild images and moderate images were similar in approximation, horizontal, and vertical but produced a one-sided bias in small areas in diagonal.

The results of this study can be used for development of an automatic identification system of CXR images.

[References]

[1] Jung Yeon Heo, “Clinical and Epidemiology Characteristics of Coronavirus Disease 2019 in the Early Stage of Outbreak”, Korean J Med. (2020) Vol. 95, No. 2, PP. 67-73

[2] Cristina Mesa Vieira, Oscar H. Franco, Carlos Gómez Restrepo, et. Al. “COVID-19: The forgotten priorities of the pandemic”, Maturitas, Vol. 136, June (2020), pp. 38-41.

[3] Coleen A. Boyle, Michael H. ScD, Susan M. Navercamp, Jennifer Zubler, “The public health response to the COVID-19 pandemic for people with disabilities”, Disability and Health Journal, Vol. 13, Issue 3, July(2020), 100943.

[4] Marco Ciotti, Massimo Ciccozzi, Alessandro Terrinoni, et. al., “The COVID-19 pandemic”, Critical Reviews in Clinical Laboratory Sciences, Vol. 57, Special issue:COVID-19 Pandemic and the Critical Role of the Clinical Laboratory, PP. 365-388

[5] Ki-il Lee, Dong Kyu Kim, Ju-Hun Mo, “Clinical Reviews of COVID-19 for Otorhinolaryngologists ”, J Rhinol, (2021)Vol. 28, No. 1, PP. 1-13

[6] So-Hee Hong, Hyo-Jung Park, Jae-Hwan Nam, “Lessons Learned from SARS-CoV: Preparation for SARS-CoV-2 induced COVID-19”, The Korean Society for Mocrobiology, (2020) Vol. 50, No. 2, pp. 76-96

[7] S. Tabik, A. Gomez-Rios, J. L. Martin-Rodriguez, et al, “COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images”, IEEE Journal of Biomedical and Health Informatics, (2020), Vol. 24, Issue 12, PP. 3595-3605

[8] Christopher Torrence, Gilbert P. Compo, “A Practical Guide to Wavelet Analysis”, Bulletin of the American Meteorological Society, Vol. 79, Issue 1(1998), PP. 61-78.

[9] E. P. Serrano, M.A. Fabio, “Application of the wavelet transform to acoustic emission signals processing”, IEEE Transactions on Signal Processing, Vol. 44, Issue 5(1996), pp. 1270-1275.

[10] Sang-bock Lee, Hwunjae Lee, V. R. singh, “Determining the Degree of Malignancy on Digital Mammograms by Artificial Intelligence Deep Learning”, j Med Imaging(ScholarGen)(2020), Vol. 03, No. 1, PP. 17-32

[11] Junhaeng Lee, “Analysis of the DWT Characteristics of MR Molecular Images Obtained by Using MNPCA”, j Med Imaging(ScholarGen), Vol. 01, No. 01(2018), PP. 1-15.

[12] R. Vijayarajan, S. Muttan, : Discrete wavelet transform based principal component averaging fusion for medical images”, AEU-International Journal of Electronics and Communications, Vol. 69, Issue 6, June(2015), PP. 896-902.