Chest Image Based COVID Prediction Using Genetic Cluster and Histogram Feature

Priyanka Singh

Research Scholar, Computer Science &Engineering, RNTU, Bhopal

Dr S. Veenadhari

Professor, Computer Science &Engineering, RNTU, Bhopal

Abstract

Medical emergency was clearly visible during corona virus pandemic in 2020 and 2021 globally. Such services load were need to be reduce by the researchers at different level like, pharmaceutical, medical equipments, diagnosis, etc. This paper has work in field of chest image diagnosis for the prediction of COVID or not. For the analysis histogram color feature was extract from gray scale image and genetic cluster image values were used. Genetic clustering was done by butterfly algorithm. Extracted features were blocked into fix size vectors for the the training of neural network. Trained neural network classify the chest image into infected and non-infected class. Real X-Ray chest images were used for the experiment purpose. Results were compared with existing techniques of chest image classification.

Index Terms— Image Data analysis, image processing, Genetic Algorithm, visual processing.

DOI Number: 10.14704/nq.2022.20.9.NQ44161 **Neuro Quantology 2022; 20(9):1400-1408**

I. INTRODUCTION

Findings related to opacity have been discovered in the COVID-19 radiographic reports [1]. According to the findings of one study [2], their patients exhibited both bilateral and unilateral ground-glass opacity. This essential quality may prove useful in the creation of a deep learning model, which would facilitate the screening of large volumes of radiograph images for possible COVID-19 cases.

Researchers have an increased ability to sense how to analyse the common variations that will lead to disease [3] thanks to machine learning (ML) and deep transfer learning methods. Conventional algorithms such as support vector machines (SVMs), decision trees (DT), random forests (RF), logistic regression (LGR), and k-nearest neighbours (KNN) are included in these methods. In addition, deep learning algorithms such as convolutional neural networks are also a part of these methods (CNNs). A training dataset is converted to a higher dimension when using the support vector machine (SVM), which is a classification method. It does this by optimising a hyperplane [5] in order to separate the two classes with as few mistakes in classification as possible. The DT generates a model in the form of a tree to define the connections that exist between the various features and the various class labels [6]. The RF is a DT ensemble algorithm that generates multiple trees via a re-sampling procedure known as bootstrap aggregation [7]. LGR models calculate the probability that a set of data points belongs to a particular class based on the value of the features independently considered. After that, it applies this model to determine whether or not a particular data point belongs to a particular category [8]. The KNN is a type of classifier that learns by analysing the relationship between a specific unlabeled data point and the dataset that it is being taught on [9].

The quality of the data is of the utmost significance for the accomplishment of deep learning. The idiomatic expression "garbage in, garbage out" is applicable to deep learning applications in general just as much as it is applicable to deep learning in chest radiography. There has been research done in the past that suggests that interpretive errors made by radiologists can originate from both internal and external sources [10]. The terms "search,"

"recognition," "decision," and "cognitive errors" are examples of the former types of sources, whereas "fatigue," "workload," and "distraction" are examples of the latter types of sources. If the labels that are used to train the deep learning architecture are inaccurate, the resulting models will perform poorly.

This paper is broken up into a few different sections. To begin, there were the introductions, which outlined an essential component of this study. The second section provides a synopsis of the research that various academics have conducted to identify covid-19 through image diagnosis. In a subsequent paper, the proposed model of SIFCD (Selected Image Features for COVID Detection) was broken down into its component parts, complete with a flow chart and an explanation of the algorithm. Additional research has revealed a variety of ways to compare parameter values for the COVID 19 model's predictions. The work is concluded in the final paper with a variety of outcomes and potential future applications.

II. Related Work

An automated method for the diagnosis of COVID-19 from chest X-Ray images was proposed by K. K. Singh et. al. in [11]. This method was proposed. The chest X-ray images can be analysed using an improved depthwise convolution neural network, which is presented by the method. The wavelet decomposition method is utilised in order to successfully incorporate multiresolution analysis into the network. The frequency sub-bands that were determined to best identify the disease are taken from the input images and fed into the network. The objective of the network is to determine which of the three classes—normal, viral pneumonia, and COVID-19—best fits the input image. The diagnosis is performed by combining the model's predicted output with the visualisation provided by Grad-CAM. In addition, a comparative study with the currently utilised approaches is carried out.

MHA-CoroCapsule is the name given to the capsule network model with multi-head attention routing algorithm that was proposed by F. Li et al. in [12] in order to provide rapid and accurate diagnostics for COVID-19 diseases based on chest X-ray images. Convolutional layers, two capsule layers, and a non-iterative, parameterized multi-head attention routing algorithm are the components that make up the MHA-CoroCapsule. This algorithm is used to quantify the relationship that exists between the two capsule layers. Experiments are carried out on a combined dataset that is comprised of two different datasets that are available to the public. These datasets include normal, non-COVID pneumonia images as well as COVID-19 images.

In their paper [13], Q. Yan and colleagues present a feature variation block that can automatically adjust the global properties of the features in order to segment COVID-19 infection. The FV block that has been proposed has the potential to improve the capability of feature representation both effectively and adaptably for a wide variety of cases. We handle the complex infection areas that have a variety of appearances and shapes using a method that we've developed called Progressive Atrous Spatial Pyramid Pooling. This method fuses features at different scales. The performance of the suggested approach is comparable to that of the most recent innovations.

In the paper [14], written by Zhang et al., an attempt is made to improve the classification accuracy using domain shift between datasets, specifically COVID-DA. This is done within the context of a semi-supervised framework, which means that both labelled and unlabeled data are used for the purpose of learning the model. The training set contains a total of 8154, 2306, and 258 CXR images, respectively, categorised as normal, pneumonia, and COVID-19.

Before conducting a comprehensive study on several widely used models like VGG, ResNet, DenseNet, Inception-ResNet, Inception-V3, and MobileNet-V2, El Asnaoui and Chawki [15] divide pneumonia into bacterial pneumonia and viral pneumonia. This is done so that the results of the study can be more accurately compared. The dataset that was used includes 231 COVID-19 cases, 1583 normal cases, 2780 bacterial cases, and 1493 viral cases. Eighty percent of the samples were used for training, and the remaining twenty percent were used for testing. Finally, they have determined that the InceptionResNet-V2 is superior to all other models with an overall accuracy and F1-score that are, respectively, 92.18% and 92.07%.

Liang et al. [16] proposed a multimodal Deep Belief Network (DBN) as a means of conducting an unsupervised learning strategy-based analysis of cancer data. The DBN is able to identify cancer subtypes based on genomic data as well as clinical data, and it can capture intra-modality correlations as well as cross-modality correlations. If the dataset used for training is relatively small, however, their method might not have the capacity for generalisation if it does not employ higher-level feature fusion. In addition, the works described above do not address the issue of the missing modality in situations involving multiple modalities. In the research that we conducted, CT images and EMR data were utilised to make severity assessments of the COVID-19 patients. Particularly in the case that a sample has a problem with a missing modality (for example, there are no available CT images), other available modalities can be a complement.

The findings that Chang Min and colleagues reported in [17] The authors of this paper propose a model for the detection of chest X-ray outliers that makes use of dimension reduction and edge detection as a potential solution to the problems that have been identified. Using the proposed method, an X-ray image is scanned while a window of a certain size is kept in view throughout the process. The AND operation is then used to extract edge information in a binary format after difference imaging of adjacent segment-images has been performed. In order to convert the extracted edge, which is a form of visual information, into a series of lines, it is computed in convolution with the detection filter that has a coefficient of 2n. The lines are then sorted into one of sixteen distinct types according to their characteristics. After that, the data that was previously reduced is put into the RNN-based learning model as an input. This is accomplished by counting the data that has been converted, which leads to the production of a one-dimensional array with a size of 16 for each image segment. Counting the data also helps ensure that the correct values are used. In addition, in order to evaluate how well the proposed model works, a number of experiments were carried out as part of the research. The COVID-chest X-ray dataset was used as the foundation for these experiments.

III. Proposed Methodology

This section brief proposed SIFCD (Selected Image Features for COVID Detection) model. Fig. 1 shows the block diagram of covid-19 prediction model steps. In order to increase the understanding of the SIFCD explanation of each block was also doen in the section with all set of equations. Proposed work uses some notations and meaning of each was done in table 1.

Table 2 Notation used in work.

Notations	Meaning
XI	X-Ray Image Input
XIP	X-Ray Image Pre-Processed
HF	Histogram Feature
B	Butterfly
S	Sensitivity
C1	Cognitive Parameter
C2	Social Parameter
m	Number of row
n	Number of Column
α	Constriction
b	Block size
GC	Genetic Cluster Feature
TF	Training Feature
TCDM	Trained Covid Detection Model

Input X-Ray Pre-Processing

Input X-Ray XI images were pre-processed before the feature extraction. As input image dimension need to be reset as per the working environment. Further image was transformed into gray format. Transformation into specific matrix was done by the eq.1, where

$$X \leftarrow RGB\_to\_Gray\_Format(XI)$$

$$XIP = XI_R + XI_G + XI_B \text{-----Eq. 1}$$

In order to extract histogram feature form the image gray format reduces the calculation cost. Hence work has transformed RGB format into gray scale.

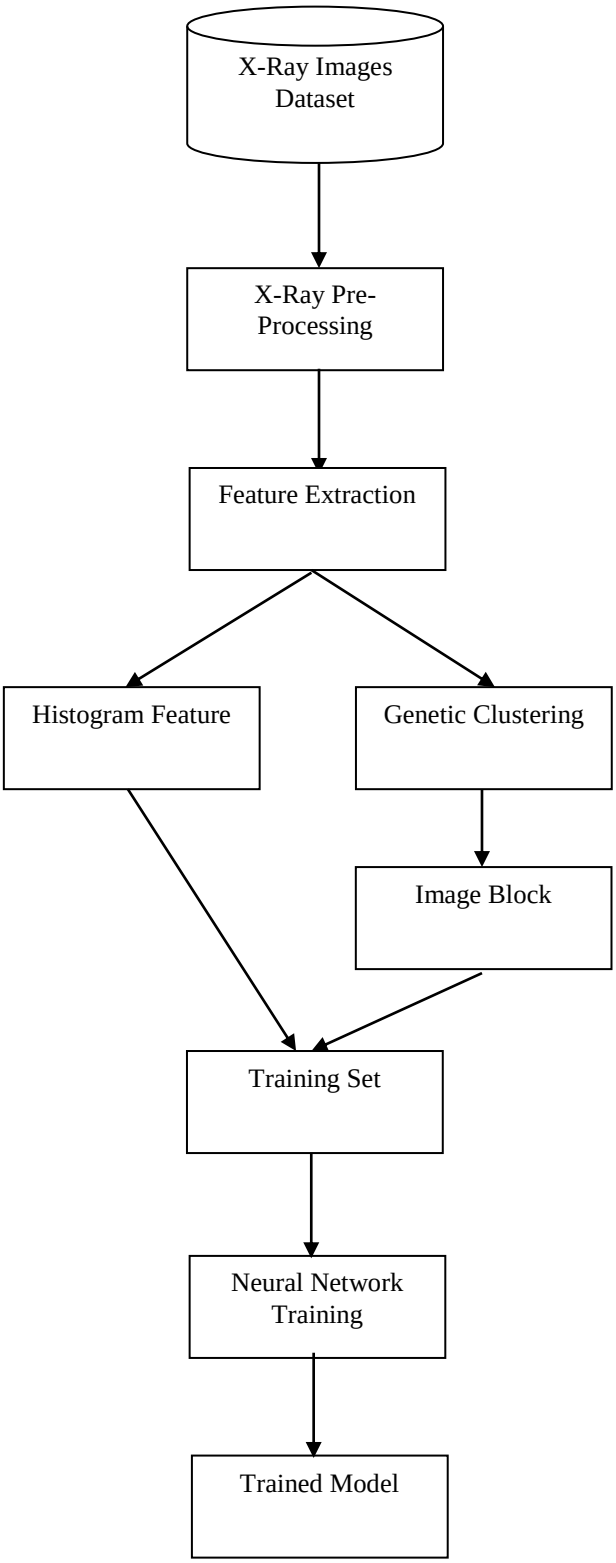


Fig. 1 Block diagram Selected Image Features for COVID Detection from Chest X-ray.

Feature Extraction

This work has extract color histogram feature from the input X-Ray pre-processed images. For these 16 bins of 0-255 gray range was developed and counted. Extracted feature was commonly used in the training of the proposed model.

$$HF \leftarrow Hist(XIP, b) \text{---Eq. 2}$$

Hist Function
Input: XIP, b
Output
Loop 1:b
Bin_Count=0
    Loop 1:m
Loop 1:n
        If b==XIP[m,n]
            Bin_Count= Bin_Count+1
EnfIf
EndLoop
EndLoop
HF[b]= Bin_Count
EndLoop

Genetic Clustering

Butterfly Optimization Genetic algorithm was used in the work for clustering the image XIP. This clustering includes steps shown in fig. 2.

Generate Butterfly

In order to get clustered image, paper need cluster center. Hnece random set of cluster center is termed as butterfly. Collection more than one butterfly in the model is population of butterfly. So Eq. 2 shows the butterfly population matrix B, having c number of clusters.

$$B \leftarrow Gaussian\_Funciton(c, XIP) \text{---Eq. 3}$$

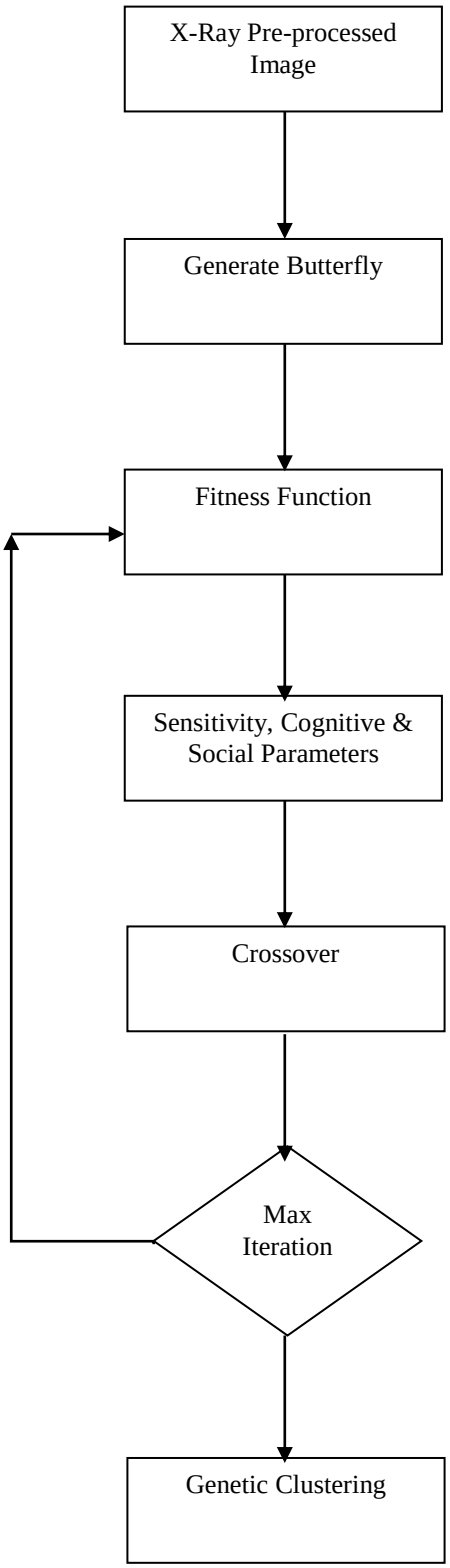


Fig. 2 Genetic clustering steps.

Fitness Function

Each butterfly fitness was evaluate by the summation of minimum pixel value difference from cluster centers.

$$BFV \leftarrow SUM(\min(B, XIP))$$

Sensitivity of Butterfly

$$S = e^{-(M_r - C_r)/M_r} \text{---- } 4$$

Where S is sensitivity of r^{th} iteration where M_r is maximum number of iteration takes place and Cr is current iteration of this BF-PSO algorithm.

Cognitive and Social parameters

$$C_1 = y * \left(\frac{C_r}{M_r} + x\right) \text{-----Eq. 5}$$

$$C_2 = x * \left(\frac{C_r}{M_r}\right) \text{-----Eq. 6}$$

Where x, y are constant

Crossover

As per the sensitivity constriction factor was evaluate by the work by eq. 5.

Constriction Factor C_{eq}

$$\alpha = C_1 + C_2 + S \text{-----Eq. 7}$$

$$C_{eq} = 1 - \alpha - \sqrt{\alpha^2 - 4\alpha} \text{---- } 8$$

This constriction factor finds the number of positions c in the butterfly to modify. Best butterfly perfrom crossover operation with other existing butterflies.

Max Iteration

Check for maximum iteration count. If max iteration not reach then move to fitness function step, otherwise current best butterfly of max iteration is final cluster centers of image.

Blocked Image

Genetic cluster center gives c number of values to the pixels present in the image. Now whole image was break into bxb blocks. Blocks were moved from left to right and top to bottom. Cluster values from the blocked image were transform into single dimension vector and merge with histogram feature to the training feature .

GC←Block(HIP, B, b) -----Eq. 9

TF←[GC HF] -----Eq. 10

Training of Neural Network

- A layer neural network is assume which have three layers.
- Input layer neuron were identified by I, while hidden layer neuron were identified by j. Output layer neuron is identified by k.
- Weights between neuron is represent by $w_{ij,}$ where i and j are neuron layers.
- Eq. 11 shows output of neuron as per weight and biasing value b_j:

$$X_j = \sum x_i \, . \, w_{ij} \, b_j \qquad 11$$

where, $1 \leq i \leq n$; n is the number of inputs to node j, and b_j is the biasing for node j. Hence network will learn the weights between layers. This error need to be correct by adjusting the weight values of each layer. So estimation of error was done by eq. 8 [22].

Proposed SIFCD Algorithm

Input: XI//X-Ray Images

Output: TCDM// Trained Covid Detection Model

1. X←RGB_to_Gray_Format(XI)
2. HF←Hist(XIP, b)
3. B←Genetic_Cluster(XIP,c)
4. GC←Block(HIP, B, b)
5. Loop 1:nb // Number of Blocks
6. TF←[GC HF]
7. Endloop
8. TCDM ←Initialize(TF)
9. Loop 1: epochs
10. EBPNN←Train(TCDM, TF, D)
11. EndLoop

IV. Experiment and Results

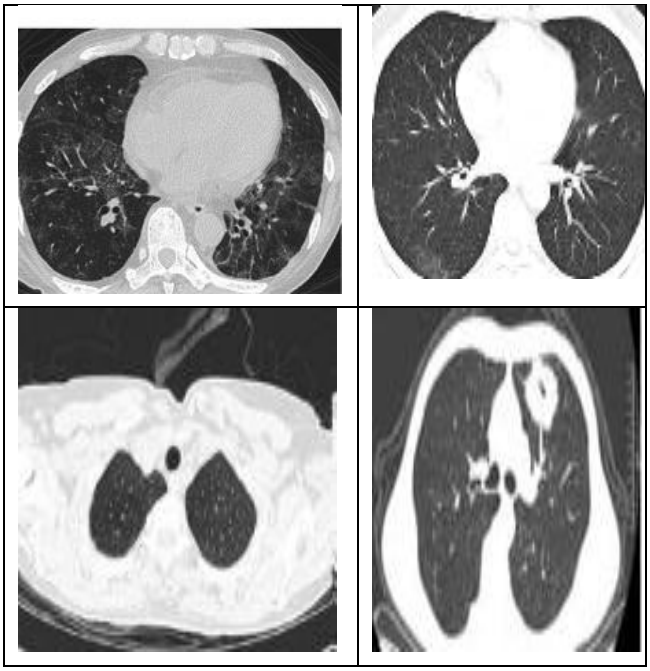
Implementation of X-ray chest images classification was done on MATLAB. Experimental machine have configuration of I3 6th generation processor with $GB RAM.. Experiment dataset was taken from the [20]. Detail description of the dataset was done in table 2.

Table 2 Detail description of dataset.

Features	Details
Class	2 (Covid, Non Covid)
Number of Images Non Covid	397
Number of Images Covid	347
Dimension Range Covid	1637x1225 to 148x61
Dimension Range Covid	512x450 to 236x184

Table 3 Dataset images.

Non Covid	Covid
	
	



Evaluation parameters

Prediction of image class models were evaluate on following parameters.

$$\text{Precision} = \frac{\text{True\_Positive}}{\text{True\_Positive} + \text{False\_Positive}}$$

$$\text{Recall} = \frac{\text{True\_Positive}}{\text{True\_Positive} + \text{False\_Negative}}$$

$$F\_Score = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$Accuracy = \frac{Correct\_Classification}{Correct\_Classification + Incorrect\_Classification}$$

Results

Table 4 Chest image classification models precision values.

Testing Images	SIFCD	Previous Model
25	1	0.9011
50	1	0.9105
100	0.99	0.9274
200	0.9874	0.8725
250	0.9766	0.8456

Precision values in table 4 shows that proposed SIFCD model has higher values as compared to previous model in [17] at different testing images set. It was found that proposed model has improved the work performance by the used of genetic cluster features. In previous model use of blocked edge feature has not perform well to predict correct class of image.

Table 2 Recall value based image comparison.

Testing Images	Proposed Model	Previous Model
25	1	1
50	1	1
100	1	1
200	1	1
250	1	1

Chest image classification models recall values in table 5 shown that proposed SIFCD model has almost same values as compared to previous model. Use of neural network for learning has increases the performance of the work. Histogram feature and genetic cluster feature improves the learning of the model.

Table 3 F-measure value based image comparison.

Testing Images	Proposed Model	Previous Model
25	1	0.948
50	1	0.9531
100	0.995	0.9624
200	0.9937	0.9319
250	0.9882	0.9163

F-measure values in table 4 shows that proposed SIFCD model has higher values as compared to previous model in [17] at different testing images set. It was found that proposed model has improved the work performance by the used of genetic cluster features. In previous model use of blocked edge feature has not perform well to predict correct class of image.

Table 4 Accuracy value based image comparison.

Testing Images	Proposed Model	Previous Model
10	100	95.0489
20	100	95.5210
30	99.952	96.3707
40	99.1208	93.6252
50	98.985	92.3092

Chest image classification models recall values in table 5 shown that proposed SIFCD modle has improved it by 5.056% as compared to previous model. Use of neural network for learning has increases the performance of the work. Histogram feature and genetic cluster feature improves the learning of the model.

V. Conclusions

Global pandemic teaches us a lot and proofs that technology reduces the losses. To contribute the medical image diagnosis this paper has proposed a model that predicts the Covid disease from the input X-Ray chest image. It was found that use of histogram color feature for the training with image cluster proposed model has increases the work efficiency of the correct class detection. Experiment work was performed on real X-Ray chest images on different testing dataset size. Result shows that use of neural network and training feature combination in SIFCD model has improved the precision value 10.03%, f-measure by 5.32%. In future scholars can adopt other learning model for the training of covid prediction.

References

[1] AAbbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest x-ray images using detrac deep convolutional neural network," Applied Intelligence, vol. 51, 2020.

[2] L. Li, "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy," vol. 296, no. 2, 2020.

[3] Rezaeijo SM, Ghorvei M, Alaei M (2020) A machine learning method based on lesion segmentation for quantitative analysis of CT radiomics to detect covid-19. In: 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). Mashhad: IEEE, pp 1–5.

[4] 24. Fatima M, Pasha M (2017) Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl 9(01):1–16.

[5] 25. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297.

[6] Quinlan JR (1987) Simplifying decision trees. Int J Man-Mach Stud 27(3):221– 234.

[7] Ho TK (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20(8):832–844

[8] Peng C-YJ, Lee KL, Ingersoll GM (2002) An introduction to logistic regression analysis and reporting. J Educ Res 96(1):3–14.

[9] Keller JM, Gray MR, Givens JA (1985) A fuzzy k-nearest neighbor algorithm. IEEE Trans Syst Man Cybern 14(4):580–585

[10] Rehman, S. Naz, A. Khan, A. Zaib, and I. Razzak, Improving Coronavirus (Covid-19) Diagnosis Using Deep Transfer Learning, MedRxiv, Cold Spring, NY, USA, 2020.

[11] K. K. Singh and A. Singh, "Diagnosis of COVID-19 from chest X-ray images using wavelets-based depthwise convolution network," in Big Data Mining and Analytics, vol. 4, no. 2, pp. 84-93, June 2021.

[12] F. Li, X. Lu and J. Yuan, "MHA-CoroCapsule: Multi-Head Attention Routing-Based Capsule Network for COVID-19 Chest X-Ray Image Classification," in IEEE Transactions on Medical Imaging, vol. 41, no. 5, pp. 1208-1218, May 2022.

[13] Q. Yan et al., "COVID-19 Chest CT Image Segmentation Network by Multi-Scale Fusion and Enhancement Operations," in IEEE Transactions on Big Data, vol. 7, no. 1, pp. 13-24, 1 March 2021.

[14] Y. Zhang et al., "COVID-DA: Deep domain adaptation from typical pneumonia to COVID-19," 2020. [Online]. Available: arXiv:2005.01577.

[15] K. El Asnaoui and Y. Chawki, "Using X-ray images and deep learning for automated detection of coronavirus disease," J. Biomol. Struct. Dyn., vol. 39, no. 10, pp. 3615–3626, 2021.

[16] M. Liang, Z. Li, T. Chen and J. Zeng, "Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach", IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 12, no. 4, pp. 928-937, Jul. 2015.

[17] Chang Min Kim, Ellen J. Hong, Roy C. Park. "Chest X-Ray Outlier Detection Model Using Dimension Reduction and Edge Detection". IEEE Access, June 22, 2021.

[18] Priyanka Joshi, Asst. Prof. Ashwini Bade. "LSB Modification Techniques of Audio Steganography for Secure Communication". volume-10-issue-3, IJSET, 2022.

[19] T. Kobayashi, "BFO Meets HOG: Feature Extraction Based on Histograms of Oriented p.d.f. Gradients for Image Classification," 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 747-754.

[20] Annesha Malakar, Joydeep Mukherjee. "Image Clustering using Color Moments, Histogram, Edge and K-means Clustering". International Journal of Science and Research (IJSR), Volume 2 Issue 1, January 2013.

[21]]https://github.com/UCSD-AI4H/COVID-CT/tree/master/Images-processed.