Medical Image Analysis

An official journal of the MICCAI Society

Application of Bins Approach with Textural Moments using Machine learning for Binary and Multiclass detection and classification for Lung Image databases

Hrishikesh Telang[1], Kavita Sonawane[2]

[1]*Student, Department of Computer Engineering, St. Francis Institute of Technology, Mumbai, India, 400103*
[2] *Professor, Department of Computer Engineering, St. Francis Institute of Technology, Mumbai, India, 400103*

Abstract

The Coronavirus Disease (COVID-19) is one of the most infectious diseases that was announced a pandemic by WHO in 2020 and has led the entire world yearning in search of diagnosis, prevention and treatment ever since. In the past, several researchers have incorporated the significance of Deep Learning, Transfer Learning, and other biomedical image processing such as CT or ultrasound scan to detect and classify COVID-19 infected images from other lung infected images which has been a great boon to medical science. Currently, COVID-19 detection is performed through what is called the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test. However, the process can prove to be time-consuming owing to manual detection. Apart from that, much of the research in the past has only been limited to Deep Learning frameworks. In this paper, we are employing a unique technique by focusing on color, texture and spatial image features with our novel Bins Approach with four Statistical Moments such as mean, standard deviation, skewness and kurtosis to perform binary as well as multi-class classifications for COVID-19, Normal and Viral Pneumonia Images. To achieve this, we have also introduced our pseudo-coloring approach to first convert the grayscale X-ray images to RGB images as a major image pre-processing step. Finally, these image were reduced into a smaller feature subspace using feature selectors and classified using traditional machine learning classification algorithms as well as ensemble machine learning by applying the techniques of bagging, boosting and stacking ensemble approaches. Our final proposed multi-class distribution technique has reached stable results ranging at accuracies 88.23%, 88.42%, and 88.34%, same performance is reflected in terms of precision and recall parameters. We have done the comparative analysis with the existing algorithms as well. The image dataset was acquired from the Radiography Database acquired from Kaggle. The performance of all these classifiers have been tested and evaluated using precision, recall, F1 score and the AUC along with the analysis of the bins features selected during feature selection analysis.

Keywords: index, 8 bins approach, COVID-19, Normal, Viral Pneumonia, Lungs, Lung Images, statistical moments, accuracy, precision, recall, ROC-AUC, feature engineering, feature selection, dimensionality reduction.

Please cite this article as: First author et al., Article title, Computer Vision and Image Understanding (2017), http://dx.doi.org/10.1016/j.cviu.2017.00.000

1. *Introduction:*

The COVID-19, also known as the "Coronavirus Disease-2019", is an infectious respiratory illness typically caused by the Severe Acute Respiratory Syndrome coronavirus-2 (SARS-CoV-2) [1]. The virus belongs to the family of single-stranded viruses known as *coronaviridae*, which owes this resultant name to crown-like spikes on the surface of the virus [2]. The virus is called SARS-CoV-2 majorly because the genomic sequence structure is almost similar to Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) that appeared in China in February 2002 [1]. The virus later spread across 29 countries, where 8,098 people were affected with SARS, and 774 of them died [3]. Another coronavirus that appeared almost a decade later was the Middle East Acute Respiratory Syndrome coronavirus (MERS-CoV), identified in 2012 in Saudi Arabia and was found to be associated with camels [4]. People infected with the COVID-19 virus experience similar ailments to SARS-CoV-2, with mild to moderate symptoms such as cough, fever, fatigue, and breathlessness; however, most of them recover without special treatment. Older people, especially with comorbidities such as diabetes, hypertension, cardiovascular diseases, or cancer, are more susceptible to developing serious illnesses.

The COVID-19 was found to have originated from the city of Wuhan, at the Hubei Province of China, back in December 2019 [5]. Since then, the virus has been found to have beleaguered 213 countries, making its presence worldwide, and declared a pandemic by WHO on 11th March 2020 [6]. In its most serious cases, it can also cause inflammation of the air sacs in the lungs, given to be known as Pneumonia. The disease can either be viral or bacterial; however, we are only focusing on Viral Pneumonia in this study. To make matters worse, COVID-19 has also evolved over time, making changes within itself what is known as a "mutation"; a multiple of such mutations are also referred to as a "variant" of the original virus [7]. Such variants have been found in the United Kingdom, Denmark, South Africa, South Korea, Japan, Thailand, and India [7].

Owing to its highly infectious nature, it was crucial to detect this disease at an earlier stage. While detection can practically be performed using Reverse Transcription Polymerase Chain Reaction (RT-PCR) test, it can also prove to be time-consuming. It can put a considerable load on medical staff due to manual detection [8]. The test is also susceptible to environmental changes, collection, and operation changes [8]. To expedite this arduous process, various researchers in the past year have performed analysis for COVID-19 detection and classification with Normal Images, typically using Deep Learning models. However, in advanced biomedical domain scenarios, classifications must also be performed with other lung infections for detailed and extended diagnosis. With COVID-19 being so heavily emphasized during research in medical science, the possibility of other lung infections such as Pneumonia is almost overlooked. Certain patients have also been misdiagnosed with COVID-19 or with Pneumonia or other cardiovascular complications in many scenarios, considering the symptoms being common in both infections [9]. This can prolong the treatment of the infection caused in the patient's lungs, which in worst cases, can also lead to severe ailments, or at worst, death.

In this study, we are trying to incorporate our approach towards each of the binary classifications of COVID-19, Viral Pneumonia, and Normal images using our Extended Bins Approach with Statistical Moments. Additionally, we have also performed a multi-class classification of these images to test the level of ambiguity within the classifications, to avoid human errors, while also bearing in mind that the system could be easily deployable. With the modest results thus obtained, we also tried improving the range of accuracy by using Bagging, Boosting, and Stacking Ensemble Learning Classifiers.

2. *Literature Review:*

As mentioned in Section I earlier, various researchers have already worked on classifying COVID-19 X-ray radiography images from Normal lung images using Deep Learning models. Ali Narin performed a multi class classification of COVID-19, Normal and Viral Pneumonia using SVM using feature maps obtained from the ResNet-50 model [10]. It was carried out in 10 replicates using 5-fold cross validation method [10]. Qjidaa *et al.* performed a multiclass classification of the Lung Images, initially by passing through the process of image preprocessing, data augmentation, and developing pre-trained CNN with transfer learning using DenseNet121, VGG16, MobileNet, InceptionV3, Xception, VGG19, and InceptionResNetV2 neural networks, feature extraction and ensemble classification [11]. They have also used Ensemble classification to merge the prediction of these pre-trained neural networks to test the final test accuracy as 98% [11]. Li *et al.* designed a COVID-GATNet neural network model. With the three classes of input images of three disparate sources, processes such as geometric transformation operations

were performed [12]. Later, the images are passed through a residual connection in ResNet that helps in achieving good performance [12]. With a concept what is known as skip learning in ResNet, the activation maps generated from the previous layers are effectively passed to the rightmost layers [12]. Nurtiyasari *et al.* also incorporated the use of CNNs to perform COVID X-ray classification from normal images using two different models and testing its accuracy, precision and recall [13]. Berliana *et al.* took input images of CT and chest X-ray images, later resizing them into 64x64 pixels and feature extraction on both sets of images using Gabor Feature [14]. This process is then followed by Stacking Ensemble Learning Model using three base learners namely Random Forest, SVC and KNN [14]. The prediction results promised and accuracy of 99% and 97% accordingly [14]. One of the different approaches made in this domain was made by Thepade *et al.*, who performed COVID-19 identification by computing Luminance Chroma features [15]. The proposed method first converts the input x-ray image into three color spaces: YCrCb, Kekre-LUV, and CIE-LUV [15]. This process is followed by taking different color space combinations, passed through machine learning classifiers using Trees, Bayes, and ensemble models in combination [15]. The results obtained were promising: ExtraTree + RandomForest + SimpleLogistic produced 90.42% accuracy, followed by ExtraTree + RandomForest + NaiveBayes at 87.92% and ExtraTree + RandomForest + SimpleLogistic at 87.08% for classification of COVID-19, Pneumonia, and normal class classification [15]. Umri *et al.* performed a process known as Adaptive Histogram Equalization (AHE) to increase the local contrast of the image, which is later passed to a CNN model [16]. Performance metrics such as Accuracy, Precision, Recall and F1 was calculated with accuracy results nearly ranging at 98% and 99% [16]. A local attention based mechanism was studied by Xu *et al.* with special emphasis on DL approach to distinguish between COVID-19, Viral Pneumonia, and Healthy CT scans [17]. In other experimentations, there were different other image modalities explored by scientists to conclude its significance in COVID-19 detection. Born *et al.* used a lung ultrasound dataset for COVID-19 detection, whereby their DL model POCOCID-Net incorporated the convolution features of VGG16, the denser portion with 64 neurons and subsequent ReLU activation, dropout and normalization layers with Softmax activation function was used to extract image features [18]. With 5-fold cross validation, sensitivity of 0.96 and F1 score of 0.92 was reported [18]. Similarly, Tsiknakis *et al.* have proposed a COVID-19 identification system using the concept of Transfer Learning with a chest X-ray image input dataset [19].

Our research was also influenced by several other concepts and ideas pertaining to the domain of Machine Learning, that is, ensemble learning classifiers such as Bagging and Boosting classifiers [20]. Nai-arun *et al.* majorly focused on data mining techniques for efficiency and predictability in diabetes classification [21]. To achieve this, Naïve Bayes, KNN and Decision Trees were used for classification [21]. Later, ensemble learning such as bagging and boosting were applied with the three classifiers, achieving an accuracy of 95.312% [21]. Nti *et al.* designed a stock price prediction model with the help of machine learning techniques to increase prediction accuracy [22]. A comparative analysis of ensemble techniques such as bagging, boosting, blending and stacking have been performed using DT, SVM and Neural Network [22]. With 25 different ensemble regressors and classifiers, the execution times, accuracy and error metrics were compares using the Ghana, Johannesburg, Bombay and New York Stock Exchange [22]. The study showed that stacking (90-100%) and blending (85.7-100%) offered higher prediction as compared to bagging (53-97.78%) and boosting (52.7-96.32%) [22]. Mienye *et al.* designed an improved heart disease rate prediction using Cleveland and Framingham heart disease datasets [23]. This was followed by a comparative study using machine learning algorithms such as KNN, LR, LDA, SVM, CART, gradient boosting and random forest [23]. On similar lines, Lin *et al.* performed a skin cancer dermoscopy image classification by proposing an ensemble CNN for multi-class classification [24]. A general stacking ensemble flow, CNNs grouping, meta-data concatenating, 1st CNNs ensemble and 2nd meta-classifiers ensemble was performed, deducing accuracy of a maximum of 91% in our ensemble model, compared with DenseNet of 87.3% [24].

Apart from research germane to the domain of COVID-19 detection and other machine learning disciplines, the foundations of our project lay in the algorithm of Bins Approach and its effect towards the statistical moments proposed by Dr. Kavita Sonawane in her dissertation [25]. Thereafter, Kekre *et al.* extracted color contents of the image using this CBIR method proposed in [25] by establishing a relationship using R, G and B planes in the image [26]. Later, bin partitions of 8, 27 and 64 were formed using two partitioning techniques to divide histograms of the image into two, three and four parts respectively [27]. Kekre *et al.* also evaluated this approach using YCbCr color space and performed a comparative study with the RGB color space [28]. This process was computed using mean, standard deviation, skewness and kurtosis [28]. On the basis of such work performed, we had originally performed a malaria parasite classification system by doing a comparative analysis of the CNN results obtained with the count of the pixels obtained using Bins approach [29]. Later, we extracted fusion feature vectors from statistical moments for

the entire image as well as calculated texture features using the GLCM, followed by classification using classifiers such as RNN, KNN and SVM [29]. From this point onwards, we decided to address the challenges faced in COVID-19 detection by performing and experimenting with this algorithm.

3. *Workflow of the System:*

The proposed work presented in this paper emphasizes on the application of the bins approach in medical image processing for COVID-19 and Viral Pneumonia detection and classification. The performance of our approach along with various machine learning techniques is experimented in this research paper using three databases of lung images. Following section explains the workflow of the system which includes preprocessing followed by bins approach and machine learning techniques. A synopsis of the whole system is described in Fig 1, but the detailed description is given below.

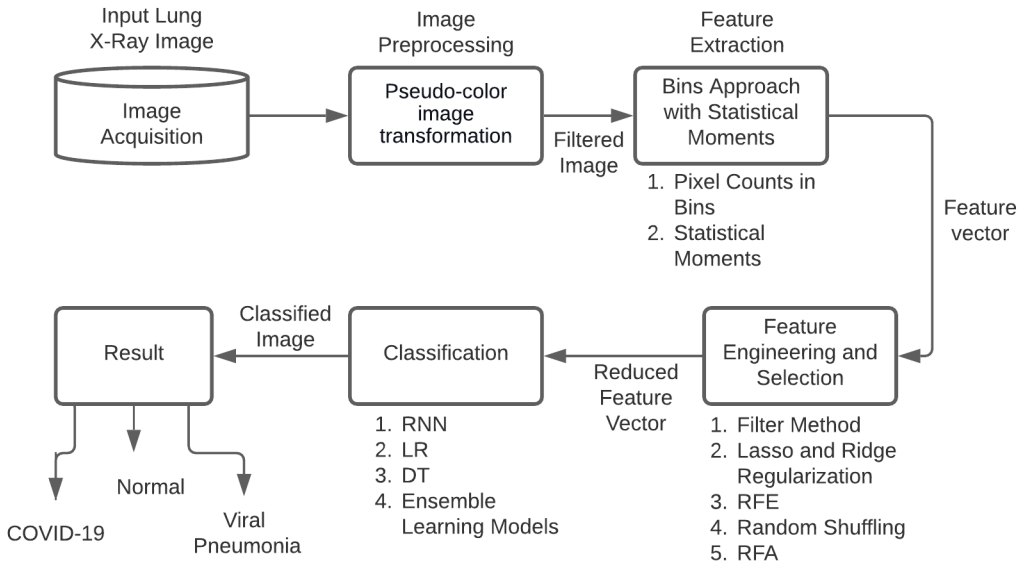


Fig. 1. Workflow of the Lung Image -for Covid-19, Normal and Viral Pneumonia Classification System

3.1. *Image Acquisition:*

Based on the experimental setup of the images mentioned in Section X, a database of all the COVID-19, Normal and Viral Pneumonia images shall be stored in a single image database. Irrespective of the format of the consolidated images, they are applied a standard format through the method of Image Preprocessing.

3.2. *Image Preprocessing:*

The X-ray radiography images cannot directly be processed for feature extraction, due to the variability in size, grayscale composition, unclean and noise infiltration present in the image. Thus, to set the images to a standard format, we feed them into the pre-processing stage. It is the set of techniques or algorithms to analyze, enhance and optimize the major characteristics of an image pertaining to its sharpness and contrast, while simultaneously suppressing its unwilling distortions. In this way, it improvises the image data by enhancing its features that accordingly ensure efficient feature extraction. In this stage, we have integrated all of the aforementioned techniques in a single algorithm termed as the Pseudo color image transformation algorithm

3.2.1 *Pseudo color image transformation:*

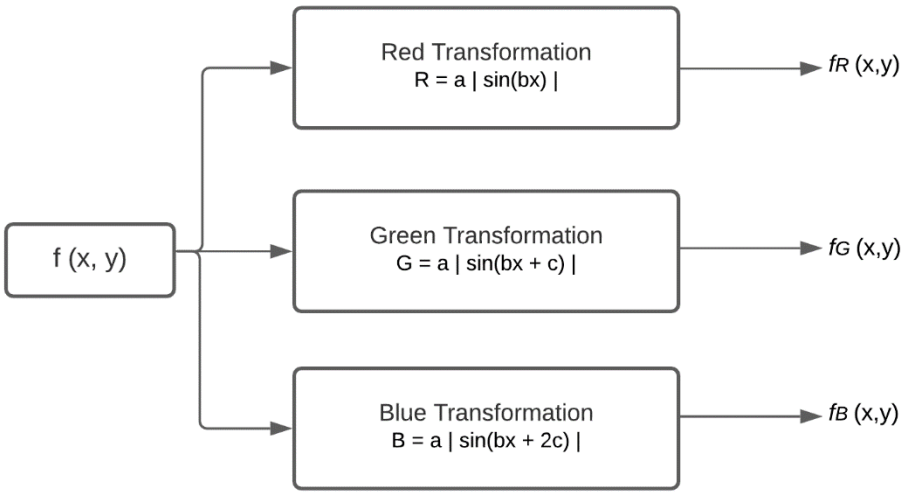


Fig. 2. Sample workflow of a typical pseudo color image transformation

The pseudo color image transformation is a process that is typically employed to convert a grayscale image to a pseudo colored RGB image by performing the process of mapping each intensity value to a color using a lookup table and a function [30]. To achieve this, the grayscale image is first passed through three separate image transformations. The functions used in our project for each of the transformations are deduced in Fig. 2. [31]. Note that a = 255, b = (2 *pi)/255, c = pi/5 are constants.

Based on the algorithm mentioned in Table I, we demonstrate such an example with a sample COVID-19 input image:

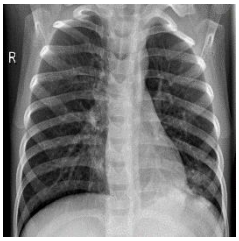
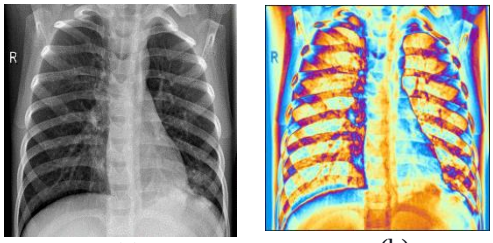


Fig. 3. Sample Viral Pneumonia input image

As per the process of Grayscale to pseudo-RGB transformation explained in Fig. 2, the following image is obtained:

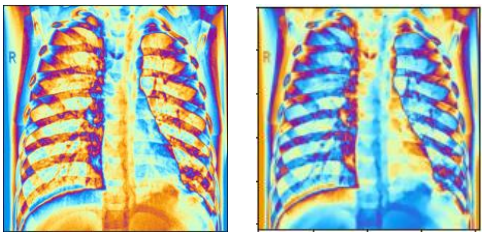


(a) (b)

Fig. 4. (a) (b) Conversion of Grayscale Lung image to pseudo-RGB transformation

As per the process of **RGB to Histogram Equalization**, since the pseudo-color image obtained is high contrast in nature, it is of paramount importance to suppress these intensity values through the Histogram Equalization approach.

Step 1: *Clean and resize the noisy image.*



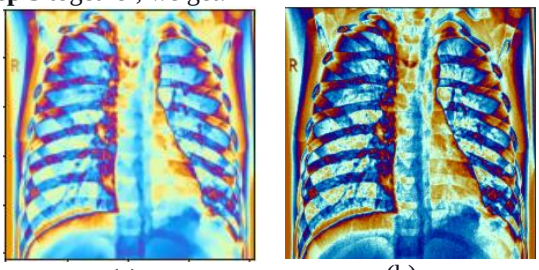
(a) (b)

Fig. 5. (a) (b) Cleaning and resizing process of noisy pseudo-RGB image

Step 2: *Convert the RGB image into YCrCb color space*

Step 3: *Split each of the channels and perform Histogram Equalization on intensity plane Y, then remerge and convert back to RGB image*

On performing **Step 2** and **Step 3** together, we get:



(a) (b)

Fig. 6. (a) (b) Conversion of Grayscale Lung image to YCbCr color space followed by HE and transformation to pseudo-RGB image

3.3. *Feature Extraction using Bins approach:*

Feature Extraction is the most crucial stage of any Content-based Image Retrieval (CBIR) system. Each image in the database undergoes feature extraction technique i.e. bins approach which generates the feature vector of size 8 bin components. In image processing, every color histogram of the RGB plane depicts total tonal distribution that occurs using a set of bins. With a traditional global color histogram approach, an image is usually encoded with its color histogram, and the distance between two images will be determined by the distance between their color histograms. However, it uses 256 bins as a feature vector and increases the size, time, and space complexity required for classification. Although the presence of most of these features can be ignored, it is still a crucial factor to be considered while designing an efficient CBIR system.

Therefore, in our research work, we have resolved this issue by extending this technique used in our previous paper [29] to form these bins so that color details of the image will be separated properly, feature vector size can be reduced through dimensionality reduction without disturbing the features. Using bins approach, we are focusing on the use of image histograms to extract the image features and also trying to reduce the size of the feature vector.

Let, Fig 6 (b) be the input

3.3.1. *Separate the image in R, G and B planes and compute the histograms for each of them as follows:*

In this step we have divided the histogram in 2 partitions in the form of a unique flag. In this case, the flag added to each partition is either a '0' or '1'. The partition threshold is denoted by the Center of Gravity. The idea behind this computation of a CG threshold is to give equal importance to the pixel values ranged at their respective intensities. Each of these intensities is considered as weight of the pixels so that we divide these pixels into two values according to their weights by computing the respective CG of the three planes.

The formula to compute a CG is given by: [32, 33, 34]

$$CG = \frac{(L_i W_i + L_{i+1} W_{i+1} + \cdots + L_n W_n)}{\sum_{i=1}^{n} W_i} \quad (4)$$

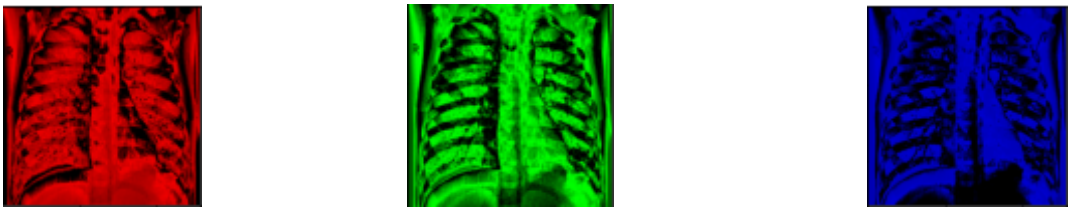


Fig. 7. (a) (b) (c) RGB color space of sample Lung image

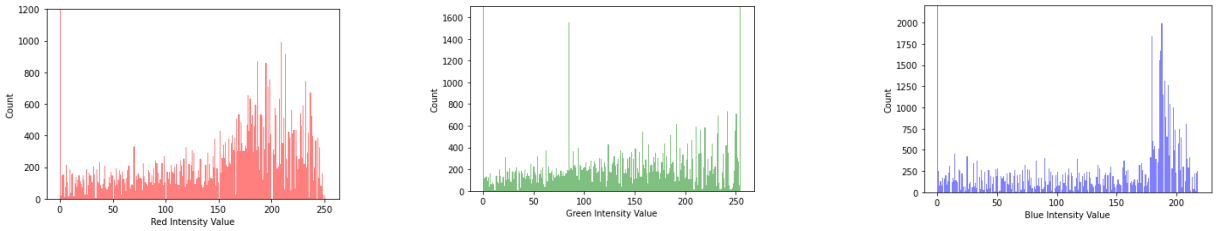


Fig. 8. (a) (b) (c) Histogram values of each of the color space of the Lung Image

3.3.2. *Bins sample of count of pixels:*

The following Fig. 9 show the CG threshold of 0 and 1 from the example input image shown in Fig. 6. In Fig. 9 (a), we can see that the count of pixels from the threshold value of 0 has lesser frequency of intensity than the one at 1. This can similarly be observed in Fig. 9 (b), however, we observe in Fig. 9 (c) that certain pixel counts are the highest at the threshold value of 1 as compared to 0. This goes to show that the pixel intensity values are more concentrated at areas where the threshold value is 1.

To understand this concept better, imagine a virtual sliding box that starts from the topmost left pixel of the image, and slide it across every pixel along the horizontal line. If it reaches the rightmost pixel of that line, start performing the sliding operation from the leftmost pixel just below the one where it started from. Whilst performing this operation, calculate the intensity of the pixel corresponding to the respective R, G, and B plane, as well as compute the CG of the respective channels simultaneously. At every looping iteration, check if the intensity of the pixel of the corresponding channel is lesser than or equal to the respective CG pertaining to that channel. If yes, add a '0', else '1'. This way, we calculate all the counts of each image pixel of each plane as per the 8 bins of corresponding binary values (i.e.: from 000=0 to 111=7).

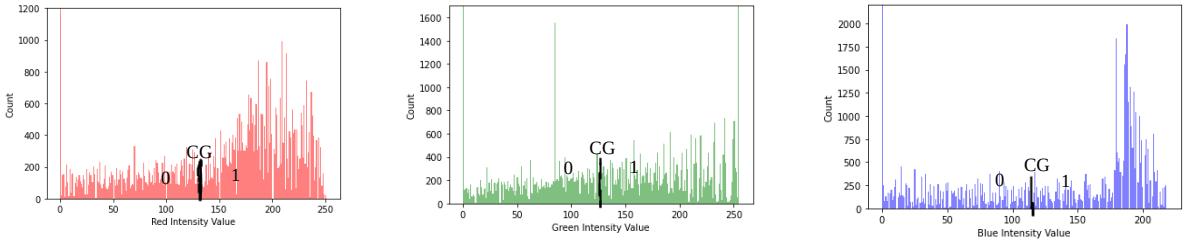


Fig. 9. (a) (b) (c) Histogram CG values of each of the color space of the Lung Image

3.3.3. *Extended Bins Approach using Statistical Moments:*

In the previous method, we only calculated the number of pixels and stored the count of them with the help of the CG values obtained bin values. However, to make efficient use the image contents, statistical properties are computed for each of the pixel intensities counted into the 8 bins. These bins contain each of those pixels that are essentially measured and evaluated using the four moments, that is, mean, standard deviation, skewness and kurtosis. This leads to the generation of 4 moments with 3 color spaces, i.e.: 12 feature vectors, in which each of the feature vector has 8 bins (similar to the calculations obtained in the previous step). They are as follows:

Feature Vectors for R component: $R_{mean}, R_{standard\_Dev}, R_{skewness}, R_{kurtosis}$

Feature Vectors for G component: $G_{mean}, G_{standard\_Dev}, G_{skewness}, G_{kurtosis}$

Feature Vectors for B component: $B_{mean}, B_{standard\_Dev}, B_{skewness}, B_{kurtosis}$

Each of these components has 8 bins for each of the respective statistical components. This forms a total of 96 feature components from this method.

3.3.4. *Final FV databases obtained:*

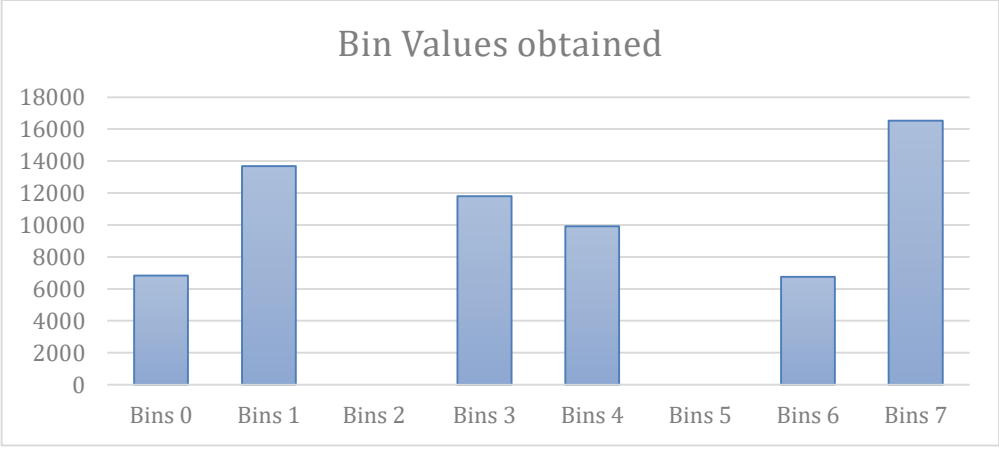


Fig. 10. Bin Values obtained from the Lung Image

3.4. *Feature Engineering and Selection*

Feature Engineering and Selection is one of the fundamental concepts of machine learning that can hugely create a positive or a negative impact on the performance of the classifier model. This is largely contributed by the choice of features that are taken into consideration as input. The process of feature selection essentially helps in reducing overfitting in a model, improves the accuracy and reduces the training time as it reduces algorithm complexity. It also keeps the features that are optimum enough in determining the best fitting model for the data depending on the selection algorithm used. In our system, we used filter methods such as taking selections using the ROC-AUC performance metric, embedded methods such as Lasso (L1) and Ridge (L2) regularization, and Recursive Feature

Elimination (RFE), Feature selection by Random Shuffling, and Recursive Feature Addition (RFA) as hybrid methods. Using these models shall reduce the feature subspace due to dimensionality reduction, which can then be simultaneously fed to the classification models in the next step.

3.5. *Classification*

This is the final step of our proposed model. Classification is the process of predicting the class of the given input reduced feature vectors. A common goal of every classification problem is to identify which category or class label would the new data fall under. This is achieved by training the model with given input features through what is called as 'supervised machine learning' system and test the performance with a training set. Our system has a binary classification as well as a multi-class classification. With the base paper proposed in [20], we extend our classification algorithms to other classifiers such as LR, DT, Kernel SVM, and Naïve Bayes. The performance of all these classifiers were tested only once using the AUC score for COVID-19/Normal Image Classification. Based on the results obtained, the redundant classifiers were removed and only the best ones were incorporated from the experimentations that followed. Additionally, we have also used ensemble learning classifiers, with special focus on Bagging and Boosting classifiers such as XGBoost, Extra Trees, AdaBoost, and Bagging Classification using RNN.

For the multi-class classification, apart from regular classifiers mentioned earlier, we have also used various combinations of Stacking ensembles using Boosting classifiers and regular classifiers to deduce which one of them procures the maximum accuracy. All these binary and multi-class classifications have been performed by checking the trade-offs amongst precision, recall and F1 score calculated throughout the course of this step. Additionally, 10-fold cross-validation was used through the entire course of this classification.

4. *Performance Evaluation Parameters*

The various performance evaluation parameters that have been considered in this lung classification system are as follows:

4.1 *Accuracy:* The accuracy determines how often the classifier shows the correct number of predictions made by the system over all kinds of predictions made. The formula is given in Equation (4).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

4.2 *Precision:* It measures the efficiency of being precise, i.e.: it determines what proportion of patients were diagnosed as having COVID-19/Normal/Viral Pneumonia actually had it. The formula is given in Equation (5).

$$\text{Precision} = \frac{TP}{TP+FP} \tag{5}$$

4.3 *Recall:* It measures the probability of how many patients that actually had COVID-19/Normal/Viral Pneumonia were diagnosed by the algorithm as having one of them. The formula is given in Equation (6).

$$\text{Recall} = \frac{TP}{TP+FP} \tag{6}$$

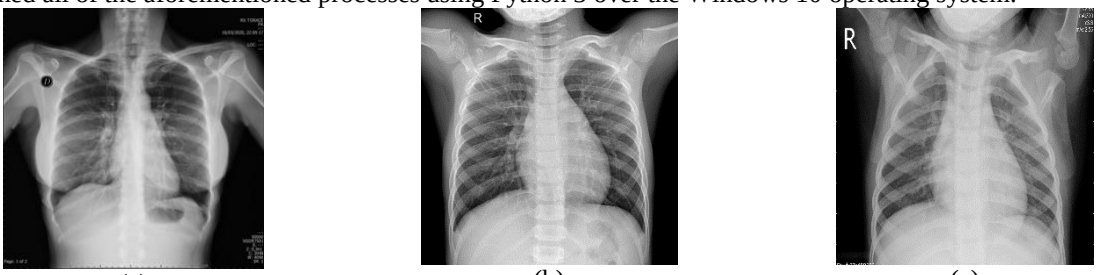
4.4 *F1 Score:* It is the harmonic mean of precision and recall. It tells how precise your classifier is, as well as how robust the classifier is. It maintains a trade-off between the precision and recall values. Higher the F1 score, better the performance of the model. The formula is given in Equation (7).

$$\text{F1 measure} = \frac{2*recall*precision}{recall+precision} \tag{7}$$

4.5 *Area under the Curve (AUC):* It is one of the most widely used performance metrics for evaluation, typically used for binary classification problems. The AUC of a classifier is equal to the probability whether the classifier will adjust the true positive rates more than the false positive rates. It has a range of [0, 1], in which the greater the value, the better the performance of the model.

5. *Experimental Setup:*

In this paper, we procured a 3-class dataset from the Radiography Database of 1143 samples of "COVID-19", 1341 samples of "Normal", and 1345 samples of "Viral Pneumonia" images. This dataset of images is publically available on the Kaggle Website. Each of the image in the dataset has a size of 1024x1024 and is given in Fig. 5 (a) (b) (c). We performed all of the aforementioned processes using Python 3 over the Windows 10 operating system.



(a) (b) (c)

Fig. 5. (a) (b) (c) Sample of Lung Images of COVID-19, Normal and Viral Pneumonia (from left to right)

6. *Results and Discussion:*

6.1 Results obtained through COVID-19/Normal/Viral Pneumonia images:

6.1.1 Classification of COVID-19/Normal Lung Images:

To test the performance of our models, we incorporated the approach of AUC curve to check which classifiers are performing best. This can be observed in Table 1 below. The ROC-AUC performance provides a different scenario: KNN, Linear SVM, and Naive Bayes render an accuracy of 0.5, in some cases lesser, nearly for all algorithms of feature selectors as compared to RNN, LR, Kernel SVM, and DT, peaking at 0.8 and 0.9 AUC score. Thus, we can conclude that not all classifiers are suitable for predicting the class labels given the input feature vectors and are somewhat haphazardly classified. Thus, KNN, Linear SVM and Naïve Bayes have not been ignored in our future experimentations and have been emboldened in yellow.

Table 1: Comparative results with ROC-AUC for selecting the best Classifier for **COVID-19/Normal Lungs detection and classification**

Classifier	Without Feature Selection	Filter Method using ROC-AUC Performance Metric	Lasso Regularization (L1)	Ridge Regularization (L2)	Recursive Feature Elimination	Feature Selection by Random Shuffling	Recursive feature addition
RNN	0.9705	0.9701	0.9703	0.9729	0.9688	0.9698	0.9764
Logistic Regression	0.9738	0.9745	0.9067	0.9158	0.9478	0.9712	0.9582
Kernel SVM	0.8947	0.8206	0.5384	0.8537	0.5021	0.8992	0.7604

Decision Tree	0.8837	0.9018	0.8946	0.8776	0.8748	0.8846	0.8903

6.1.2 Classification of Normal/Viral Pneumonia Lung Images

The following Table 2 describes the comparative accuracy results obtained using several feature selectors for Normal/Viral Pneumonia Lung Image Classification.

Table 2: Comparative results for Bins approach with Different classifier for Normal/Viral Pneumonia Lung Images for detection and classification (Accuracy %)

Classifier	Without Feature Selection	Filter Method using ROC-AUC Performance Metric	Lasso Regularization (L1)	Ridge Regularization (L2)	Recursive Feature Elimination	Feature Selection by Random Shuffling	Recursive feature addition
RNN	81.31%	74.23%	81.09%	82.00%	80.67%	81.36%	81.04%
Logistic Regression	86.7%	71.30%	83.09%	85.41%	85.3%	79.71%	75.72%
Kernel SVM	65.01%	72.95%	74.33%	62.35%	67.8%	63.68%	60.54%
Decision Tree	76.78%	65.87%	77.31%	76.52%	76.04%	76.14%	77.84%

It can be observed that Logistic Regression performs the best accuracy, ranging at 86.7% and 85.3%, followed by RNN, DT and the worst accuracy ranges being Kernel SVM. Moreover, RNN has continually shown excellent performance across all feature selectors. Overall, we observed that LR, and RNN getting higher accuracy than Kernel SVM (0.5). Due to this reason, we are excluding Kernel SVM from our future experimentations as well. Another observation that can be made is over the feature selectors: the best feature selector is Ridge Regularization, whereas the worst is the filter method using ROC-AUC Performance metric. This is followed by RFE, Random Shuffling and RFA. Most of these noticeable changes can be observed in Logistic Regression.

6.1.3 Classification of COVID-19/Viral Pneumonia Lung Images

The following Table is pivotal to our research work, essentially because classification amongst COVID-19 and Viral Pneumonia is paramount for efficient diagnosis and expedited treatment of several patients suffering from this disease.

Table 3: Comparative results for Bins approach with Different classifier for COVID-19/Viral Pneumonia detection and classification

Classifier	Without Feature Selection	Filter Method using ROC-AUC Performance Metric	Lasso Regularization (L1)	Ridge Regularization (L2)	Recursive Feature Elimination	Feature Selection by Random Shuffling	Recursive feature addition
RNN	84.03%	75.05%	83.40%	91.69%	80.67%	84.90%	92.33%
Logistic Regression	92.45%	71.47%	91.70%	95.27%	84.35%	86.92%	91.00%
Decision Tree	81.21%	70.83%	81.39%	89.10%	76.04%	82.30%	89.50%

Table 3 clearly shows that with LR, we are obtaining the best accuracy amongst other classifiers, with RNN and DT getting accuracies in the range of 91.69% and 89.10% for Ridge Regularization. The results obtained from the Table 3 are also promising. In contrast, some results provided lesser accuracy as compared to the process without feature selection, Ridge Regression, Random Shuffling, and Recursive Feature Addition. Ridge Regression is recommended to be used as it shows a substantial increase in accuracy and ROC-AUC metric performance, however, Random Shuffling and RFA are not equally bad enough either.

6.1.4 Feature Selection Analysis and Inferences:

In this section, we have analyzed the significance of each color component that was a part of the specific bin and also the statistical moment revealing the image details which are giving better performance in detection and classification. Each of the analysis has been explained from the Tables mentioned above.

As per the features selected by each classification model in **Table 1**, the following are the deductions made: It was generally perceived that Bins 2 and 7 were dominant across most moments selected during the ROC-AUC performance metric, with Bins 0, 3, 4, and an occasional 6 used interchangeably amongst them. During Lasso, bin values of 0, 5, 6, 7 of the red plane, bins 0, 6, 7 of the green plane, and Bins 0, 3, 5 of the blue plane were selected. Similarly, for Ridge regression, bins 3, 4, 6, and 7 were dominantly selected across all three planes. Recursive Feature Elimination took moments of the mean and standard deviation of all three planes along with dominant bin values 4 and 5. Recursive Feature Addition selects dominant bin values 0, 1, 3 from the planes, whereas Feature Shuffling selects Bins 5, 6, 7. Overall, the Red and green plane features of the lung images are dominant in predicting the class labels, especially with the moments of mean and standard deviation.

Based on analytical study of results in **Table 2** we found that RFA, Random Shuffling, Lasso, and Ridge Regression predominantly chose more red features than blue or green. RFE, on the other hand, chose more blue features than red or green plane ones. There is no standard majority of which feature moments are predominant across all models, as some color moments are preferred more over others during feature selection. However, Lasso chose Bins moments typically 0, 2, 4, 5, and 7, whereas Ridge selected Bins 0, 3, 4, and 6. RFE haphazardly chose Bins 0 to 7 for a few moments and 0, 2, 3, 4, 6, and 7 in myriad combinations, whereas RFA chose bins of 0, 3, 4, and 6. Finally, Feature Shuffling predominantly chose Bins 0, 1, 4, 6, and 7.

Detailed analysis of **Table 3** states that red and blue color moments are selected more than green ones, and mean and skewness is preferred over the rest of the two moments. Feature selection using ROC-AUC metric typically chose bins of 2, 3, 5, 6, and 7; Lasso chose 2, 4, 5, 6, 7, and Ridge chose 2, 3, 4, 7. Overall, bins 2, 3, 4, 6 and 7 have been chosen throughout the course of our experimentations. Based on the results obtained and the detailed analysis done we still wanted to improve it further since we already know that in the medical domain, we cannot compromise with the accuracy. With this hope, the study of several ML techniques motivated us to apply Bagging, Boosting and Stacking Classifiers as performed by Thepade *et al.* [15], as they create more accurate solutions than a single model would, by combining several classifiers into one.

In the following tables, we tried experimentation on what if we consider all the feature components instead of going for selecting the few. We were motivated to learn how the overall dataset of Bins features extracted from the Lungs Dataset impacts the overall performance of the classification using Bagging and Boosting algorithms which is discussed below with results.

6.1.5 Using Bagging and Boosting Classifiers:

We incorporated the use of XGBoost, Extra Trees, AdaBoost, and Bagging Classification using RNN with 10-fold cross-validation. The results obtained are shown in the following Table 4.

Table 4: Comparative results using Bagging and Boosting classifiers for binary classification

Type of Classification	Performance Metrics	XGBoost	Extra Trees	AdaBoost	Bagging Classifier with RNN
COVID-19 and Normal	**Accuracy**	95.67%	94.23%	96.40%	92.10%
	Precision	96.24%	95.64%	97.50%	93.90%

	Recall	95.87%	94.17%	96.50%	92.92%
	F1-score	96.01%	94.78%	96.50%	91.90%
Normal and Viral Pneumonia	**Accuracy**	87.96%	86.79%	84.60%	83.24%
	Precision	87.52%	86.02%	84.60%	83.24%
	Recall	88.39%	87.10%	85.60%	84.30%
	F1-score	87.91%	86.44%	84.60%	83.45%
COVID-19 and Viral Pneumonia	**Accuracy**	92.79%	91.18%	91.99%	85.32%
	Precision	93.00%	90.73%	92.32%	84.80%
	Recall	93.89%	92.62%	93.32%	86.80%
	F1-score	93.43%	91.98%	92.43%	84.76%

From Table 4, it is clear that XGBoost has provided the best accuracy, precision, recall, and F1 score amongst all three types of classifiers, with over 96% for classification between COVID-19 and Normal Images, 88% for Normal and Viral Pneumonia images, and over 93% for COVID-19 and Viral Pneumonia Images. The classifier showing the most modest accuracy is the Bagging Classifier with RNN, with ranging accuracies between 85% and 92%. Therefore, these improvised results are promising, thereby adopting these classifiers to classify all three images to test the overall accuracy.

6.1.6 Classification of COVID-19/Normal/Viral Pneumonia Lung Images:

Table 5 describes the accuracy, precision, recall and F1 score values obtained by the best classifiers along with respective bagging and boosting classifiers.

Table 5: Comparative results using Bagging and Boosting classifiers for multiclass classification

Classifier	Accuracy	Precision	Recall	F1 Score
RNN	73.02%	73.02%	73.02%	73.17%
Logistic Regression	83.80%	83.80%	83.80%	83.80%
Decision Tree	71.14%	71.04%	71.04%	70.34%
XGBoost	88.45%	88.45%	88.45%	88.45%
Extra Trees	83.56%	83.56%	83.56%	83.56%
AdaBoost	79.60%	79.60%	79.60%	79.60%
Bagging Classifier with DT	80.04%	82.30%	82.30%	82.30%

It can be outlined from Table 5 that XGBoost has provided the highest accuracy, ranging at around 88%. The poorest classifier performance has been RNN, with accuracy ranging at 73.02%. LR, ET and Bagging Classifier with DT work modestly, with accuracy around the range of 80%. The algorithm was robust indeed, as it provided such high accuracy without the feature selection approach. This is followed by Extra Trees, with an accuracy of around 85% for Lasso Regularization.

6.1.7 Using Ensembles:

Table 6: Comparative results using Stacking Ensemble classifiers for multiclass classification

Ensembles	Accuracy	Precision	Recall	F1 Score
Extra Tree + RNN + Logistic Regression	84.08%	85.80%	85.80%	85.80%
Extra Tree + RNN + Logistic Regression + AdaBoost	84.12%	86.50%	86.50%	86.50%
XGBoost + AdaBoost + Logistic Regression	**88.23%**	**87.40%**	**87.40%**	**87.40%**
XGBoost + AdaBoost + Extra Tree + RNN	**88.42%**	**87.90%**	**87.90%**	**87.90%**
XGBoost + AdaBoost + Extra Tree	**88.34%**	**87.42%**	**87.72%**	**87.42%**

This final Table 6 concludes our result section. With the ensemble learning techniques briefly discussed in the proposed method and the results obtained from Table 6, we wanted to experiment how ensemble models impact overall performance. Our method focused on a stacked ensemble machine learning approach, whereby we combine the respective accuracies of well-performing classifiers. Unlike Bagging, they are all different classification models, and unlike boosting, it does not correct the predictions of models which were placed in sequence earlier. We used the Boosting classifiers with regular classification algorithms, later performing 10-fold cross-validation [15]. In this way, we procured an accuracy better than the results obtained by Thepade *et al.* [15]; XGBoost + AdaBoost + Extra Tree + RNN gives an accuracy of 88.42%, XGBoost + AdaBoost + Logistic Regression of 88.23% and XGBoost + AdaBoost + Extra Tree of 88.15% as compared to two of their best classifiers proposed by them (87.92%, 87.08%, 90.42%). In the previous approaches as also mentioned in the literature review, most of the researchers have predominantly focused on deep learning models.

We can confirm from our findings that by taking a different approach to performing feature extraction using bins, our technique is one of the fewest to have incorporated a color and texture of the image to achieve an accuracy as much commensurate as, if not more insightful, than other past research papers, especially focusing on results obtained by Thepade *et al.* [15]. Thus, we propose the emboldened text in Table 6 as our COVID-19/Viral Pneumonia/Normal X-ray classification technique.

Based on the comparative analysis with respect to accuracy, precision and recall parameters of various proposed methods with other existing algorithms experimented on same data sets we have got the following findings. In binary classification we have achieved (Accuracy-96.40%, Precision-97.50%, Recall -96.50%) and (Accuracy-91.99%, Precision-92.32%, Recall -93.32%) as best results for Covid normal and Covid Pneumonia respectively for AdaBoost classifier. Binary classification done by COVID-GATNet with [12] achieved (Accuracy-94.3%, Precision-98.9%, Recall -91.9%). Another approach CLAHE and CNN [16] achieved 98% accuracy but experimentation [16] done only with 100 images for all 2 classes.

Similarly comparing the results for multiclass classification with the ensemble learning (XGBoost + AdaBoost + Extra Tree + RNN) as given Table 6, the best results are (Accuracy - 88.42%, Precision- 87.90% and Recall-87.90%) and multiclass classification by CNN [10] shows with SVM CUBIC they achieved (Accuracy – 99.35% and 94.30%) (recall – 95.89% and 95.54%) for COVID-19 and viral pneumonia respectively. The overall comparison with existing shows that proposed models in [10] and [12] are performing better than bins approach with ML classifiers.

7. *Conclusion*

Proposed application of bins approach in medical domain is experimented in this study with COVID-19, Normal and Viral Pneumonia dataset. Simplicity is best feature of bins approach which is encouraging us to apply it in various image processing-based retrieval or classification applications. Bins approach combined with Machine learning classifiers speeds up the process of classification for huge databases and specially in medical domain as well [29][35]. With this hope same approach is applied here for classifications of COVID-19, Viral Pneumonia and Normal image datasets. This entire study covered rigorous experimentation and detailed analysis with respect to different performance evaluation parameters, different types of bins-based feature vectors for binary as well as multiclass classification for all three datasets. Comparing this work of bins approach for its performance with other medical applications like malaria [29] and Alzheimer [35] states that bins is performing better for both the domains in terms of accuracy i.e 96% plus and 93% respectively. So, its application for COVID-19 also proved to be better.

Proposed application of bins with different classifiers indicates Logistic regression, bagging and boosting classifiers namely XGBoost and AdaBoost are giving better performance for binary classification. Additionally, we have also performed a multi-class classification of these images to test the level of ambiguity within the classifications, to avoid human errors, while also bearing in mind that the system could be easily deployable. In multiclass classification we tried to apply ensemble learning techniques and we found combination (XGBoost + AdaBoost + Extra Tree + RNN) performs better among all combinations.

Based on overall results and discussion we found that, novel application of bins with ML are doing better for both binary and multiclass classification for all three datasets which reflects that we could achieve the objective set for checking its performance in medical domain specially in the rising COVID-19. Comparative analysis with other existing methods contributed by [10] and [12] with similar datasets shows that still there is scope to improve performance our bins approach. However as we can notice that the precision and recall both for bins approach with AdaBoost are holding the same and higher values which clearly reflects that not only the accuracy(precision) but also same completeness(recall) is achieved in the classification which is always desirable in medical domain.

References

[1] Pal M, Berhanu G, Desalegn C, Kandi V. Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update. Cureus. 2020;12(3):e7423. Published 2020 Mar 26. doi:10.7759/cureus.7423

[2] Fang Li, Structure, Function, and Evolution of Coronavirus Spike Proteins, Annual Review of Virology 2016 3:1, 237-261, Volume publication date September 2016, https://doi.org/10.1146/annurev-virology-110615-042301

[3] National Center for Immunization and Respiratory Diseases, Division of Viral Diseases, SARS Basics Fact Sheet. Retrieved December 6, 2017, from https://www.cdc.gov/sars/about/fs-sars.html

[4] World Health Organization (WHO), Middle East respiratory syndrome coronavirus (MERS-CoV), Fact Sheet, Retrieved March 11, 2019 from https://www.who.int/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-(mers-cov)

[5] World Health Organization (WHO), Novel Coronavirus (2019-nCoV) Situation Report – 1, Data as reported by: 20 January 2020

[6] Cucinotta, D., & Vanelli, M. (2020). WHO Declares COVID-19 a Pandemic. Acta bio-medica : Atenei Parmensis, 91(1), 157–160. https://doi.org/10.23750/abm.v91i1.9397

[7] World Health Organization (WHO), SARS-CoV-2 Variants, Data as reported by: 31 December 2020, https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/

[8] Kameswari. S; M.P. Brundha; D. Ezhilarasan. "ADVANTAGES AND DISADVANTAGES OF RT- PCR IN COVID 19". European Journal of Molecular & Clinical Medicine, 7, 1, 2020, 1174-1181.

[9] Yousefzai, Rayan, and Arvind Bhimaraj. "Misdiagnosis in the COVID-19 Era: When Zebras Are Everywhere, Don't Forget the Horses." JACC. Case reports vol. 2,10 (2020): 1614-1619. doi:10.1016/j.jaccas.2020.04.018

[10] A. Narin, "Detection of Covid-19 Patients with Convolutional Neural Network Based Features on Multi-class X-ray Chest Images," 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299289.

[11] M. Qidaa et al., "Early detection of COVID19 by deep learning transfer Model for populations in isolated rural areas," 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), 2020, pp. 1-5, doi: 10.1109/ISCV49265.2020.9204099.

[12] J. Li, D. Zhang, Q. Liu, R. Bu and Q. Wei, "COVID-GATNet: A Deep Learning Framework for Screening of COVID-19 from Chest X-Ray Images," 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 1897-1902, doi: 10.1109/ICCC51575.2020.9345005.

[13] D. Nurtiyasari and D. Rosadi, "COVID-19 Chest X-Ray Classification Using Convolutional Neural Network Architectures," 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2020, pp. 667-671, doi: 10.1109/ISRITI51436.2020.9315499.

[14] A. U. Berliana and A. Bustamam, "Implementation of Stacking Ensemble Learning for Classification of COVID-19 using Image Dataset CT Scan and Lung X-Ray," 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020, pp. 148-152, doi: 10.1109/ICOIACT50329.2020.9332112.

[15] S. D. Thepade, P. R. Chaudhari, M. R. Dindorkar and S. V. Bang, "Covid19 Identification using Machine Learning Classifiers with Histogram of Luminance Chroma Features of Chest X-ray images," 2020 IEEE Bombay Section Signature Conference (IBSSC), 2020, pp. 36-41, doi: 10.1109/IBSSC51096.2020.9332160.

[16] B. K. Umri, M. Wafa Akhyari and K. Kusrini, "Detection of Covid-19 in Chest X-ray Image using CLAHE and Convolutional Neural Network," 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), 2020, pp. 1-5, doi: 10.1109/ICORIS50180.2020.9320806.

[17] Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Qin Ni, Yanfei Chen, Junwei Su, Guanjing Lang, Yongtao Li, Hong Zhao, Jun Liu, Kaijin Xu, Lingxiang Ruan, Jifang Sheng, Yunqing Qiu, Wei Wu, Tingbo Liang, Lanjuan Li, "A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia", Engineering, Volume 6, Issue 10, 2020, Pages 1122-1129, ISSN 2095-8099, https://doi.org/10.1016/j.eng.2020.04.010.

[18] J. Born, G. Brändle, M. Cossio et al., "POCOVID-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS)," 2020, https://arxiv.org/abs/2004.12084.

[19] N. Tsiknakis, E. Trivizakis, E. Vassalou et al., "Interpretable artificial intelligence framework for COVID-19 screening on chest X-rays," Experimental and Therapeutic Medicine, vol. 20, no. 2, pp. 727–735, 2020.

[20] F. Huang, G. Xie and R. Xiao, "Research on Ensemble Learning," 2009 International Conference on Artificial Intelligence and Computational Intelligence, 2009, pp. 249-252, doi: 10.1109/AICI.2009.235.

[21] Nai-arun, Nongyao & Sittidech, Punnee (2014). "Ensemble Learning Model for Diabetes Classification", Advanced Materials Research, 931-932, 1427-1431. 10.4028/www.scientific.net/AMR.931-932.1427.

[22] Nti, I.K., Adekoya, A.F. & Weyori, B.A. "A comprehensive evaluation of ensemble learning for stock-market prediction", J Big Data 7, 20 (2020). https://doi.org/10.1186/s40537-020-00299-5.

[23] Mienye, Domor & Sun, Yanxia & Wang, Zenghui. (2020). An improved ensemble learning approach for the prediction of heart disease risk. Informatics in Medicine Unlocked. 20. 100402. 10.1016/j.imu.2020.100402.

[24] T. -C. Lin and H. -C. Lee, "Skin Cancer Dermoscopy Images Classification with Meta Data via Deep Learning Ensemble," 2020 International Computer Symposium (ICS), 2020, pp. 237-241, doi: 10.1109/ICS51289.2020.00055.

[25] Kavita Vinay Sonawane , “Bins Approach To Content Based Image Retrieval”, Ph. D dissertation, Department of Technology Management, Narsee Monjee Institute of Management Studies, NMIMS, Mumbai, Maharashtra, India, 2014.
[26] H. B. Kekre, Kavita Sonawane, “ Image Retrieval Using Histogram Based Bins of Pixel Counts and Average of Intensities” (pp. 74-79, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 10, No. 1, 2012
[27] H. B. Kekre, Kavita Sonawane, “Comparative Performance of Linear and CG Based Partitioning Of Histogram for Bins Formation in CBIR”, International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN : 2278-800X, www.ijerd.com Volume 5, Issue 7 (January 2013), PP. 75-83
[28] H. B. Kekre, Kavita Sonawane, “Performance Evaluation of Bins Approach in YCbCr Color Space with and without Scaling”, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-3, Issue-3, July 2013.
[29] H. Telang and K. Sonawane, "Effective Performance of Bins Approach for Classification of Malaria Parasite using Machine Learning," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 427-432, doi: 10.1109/ICCCA49541.2020.9250789.
[30] T. LI and H. ZHU, "Research on Color Algorithm of Gray Image Based on a Color Channel," 2020 Chinese Control And Decision Conference (CCDC), 2020, pp. 3747-3752, doi: 10.1109/CCDC49329.2020.9164375.
[31] Selvapriya, B. & Raghu, B.. (2018), “A color map for pseudo color processing of medical images,” International Journal of Engineering and Technology(UAE), 7. 954-958.
[32] H. B. Kekre, Kavita Sonawane, “Bins Formation using CG based Partitioning of Histogram Modified Using Proposed Polynomial Transform ‘Y=2X-X2’for CBIR”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 5, 2011.
[33] H. B. Kekre, Kavita Sonawane, “Histogram Partitioning for Feature Vector Dimension Reduction in Bins Approach for CBIR”, International Journal of Electronics Communication and Computer Engineering Volume 3, Issue 6, ISSN (Online): 2249–071X
[34] H. B. Kekre, Kavita Sonawane, “Partitioning of Modified Histograms to Generate 27 Bins Feature Vector to Improve Performance of CBIR”, (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-4, April 2013.
[35] M. Alva and K. Sonawane, "Hybrid Feature Vector Generation for Alzheimer’s Disease Diagnosis Using MRI Images," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-6, doi: 10.1109/I2CT45611.2019.9033826.