Tuijin Jishu /Journal of Propulsion Technology
ISSN:1001-4055
Vol. 44 No. 04 (2023)

Performance Analysis of Gray Level-Based Texture Features Using Chest X-Ray Images

S. Sanjayprabu[1, a)], R. Sathish Kumar[1], S. Jafari[2], R. Karthikamani[3]

[1]*Department of Mathematics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore,India*

[2]*Dr.rer.nat in Mathematics (Graz University of Technology-Graz, Austria) Professor of Mathematics College of Vestsjaelland South Herrestarede 11 and Mathematical Physical Science Foundation 4200, Slaggelse, Denmark*

[3]*Department of ECE, Sri Ramakrishna Engineering College, Coimbatore, India*

Abstract:-A grown-up COVID-19 recognition approach for chest X-ray images is presented in this paper. The COVID-19 virus is still having a terrible impact on people's health and quality of life around the world. The first essential step in the fight against COVID-19 is effectively screening those who are infected, and imaging of the chest represents a few of the main diagnostic methods. This study used chest X-ray images, this study is intended to automatically identify COVID-19 patients while optimizing detection accuracy by employing a gray-level-based method for extracting features and some classifiers. The dataset consists of 196 COVID-19 and 196 ordinary X-rays of the chest images. The initial phase involves pre-processing a dataset of 392 chest X-ray pictures. The preprocessed images were used to extract grey-level-based characteristics in the second step. The third phase presents the KNN classifier to attain 90.1% accuracy in patient classification.

Keywords*: GHF, GLCM, GLRLM, NGTDM, GLSZM.*

1. Introduction

Due to the scale of its global spread, the World Health Organization classified the infectious disease Covid-19 to be a pandemic on the 11th of March in 2020 [1]. The COVID-19 virus has spread over the globe. It is a viral disease that spreads rapidly through one individual in direct touch with the other. Since around December 5, 2020, the world had over 66 million infections worldwide, with over 22 million patients who were physically unwell. Nearly 1.5 million patients worldwide perished as a result. Reverse transcription polymerase chain reaction (RTPCR) tests, which might take longer than 48 hours, are frequently used to identify COVID-19 [2]. This is a significant factor in its severity and quickly spread. Although this method is the one that is most frequently used to test in order to identify COVID-19, it is confusing, exhausting, and costly in terms of time with only a 62% success percentage. So alternatively computer tomography and chest radiograph give the better result and it is more accurate compared with the RTPCR test. The radiographic scan is another significant diagnostic approach for COVID-19 that is quickly accessible [3,4]. By using an X-ray of the chest, the disease covid can be identified somewhat accurately and fairly specific. Hence, excluding COVID-19 may be more useful for a chest CT or ultrasound than separating SARS-CoV-2 infection from other respiratory illnesses. One of the major, quasi-diagnostic adjuncts that are crucial in the initial examination of various pulmonary abnormalities is the X-ray. When the X-ray pictures are analyzed by skilled radiologists to seek contagious diseases associated with COVID-19, it can be used as a secondary screening technique to identify the presence of COVID-19 or to substantiate the related diagnosis [5].

KNN classifiers have been used in a few studies in the field of medical imaging. Our work adds to the body of work by illustrating how KNN can be used to classify chest x-rays. As a result, we think that KNN's advantages will have a significant positive impact on the field of medical image analysis. Hence, we suggest a method for

classifying and detecting chest X-ray images using KNN. We only had a tiny amount of resources to mark the regions on each X-ray image, hence a minimal dataset was collected. The classification model is trained and tested using already-existing MATLAB packages and a data analytics platform. In addition, we examine the model's efficacy, performance, and time execution by observing its loss function. Comparisons with medical representatives are used in performance and timeliness evaluation. Meanwhile, X-ray scans are used to assess how well the features work. We think that our research may advance the state-of-the-art chest X-ray image analysis using KNN and its development.

The structure of the paper is as follows. The Covid-19 and X-ray of the chest image modalities are introduced in Section 1. Section 2 discusses the collection and preprocessing of chest X-ray imaging data. Section 3 describes the attributes of grey level-based extraction of features approaches used on X-ray pictures of the chest. Section 4 provides examples of classifiers in effect. The outcomes of the experiment were covered in Section 5. Section 6 of the study draws a conclusion

2. Material and Methods

2.1. Image Acquisition

The open-source Kaggle website generated 392 X-ray pictures of the chest from the Covid-19 collection used in this analysis. Images from X-rays of the primary patients with normal and Covid-19 are included in this dataset. There are 196 chest X-ray images that are COVID-19 positive and another 196 images that are not out of 392 total images. Figure 1. depicts the overall proposed work

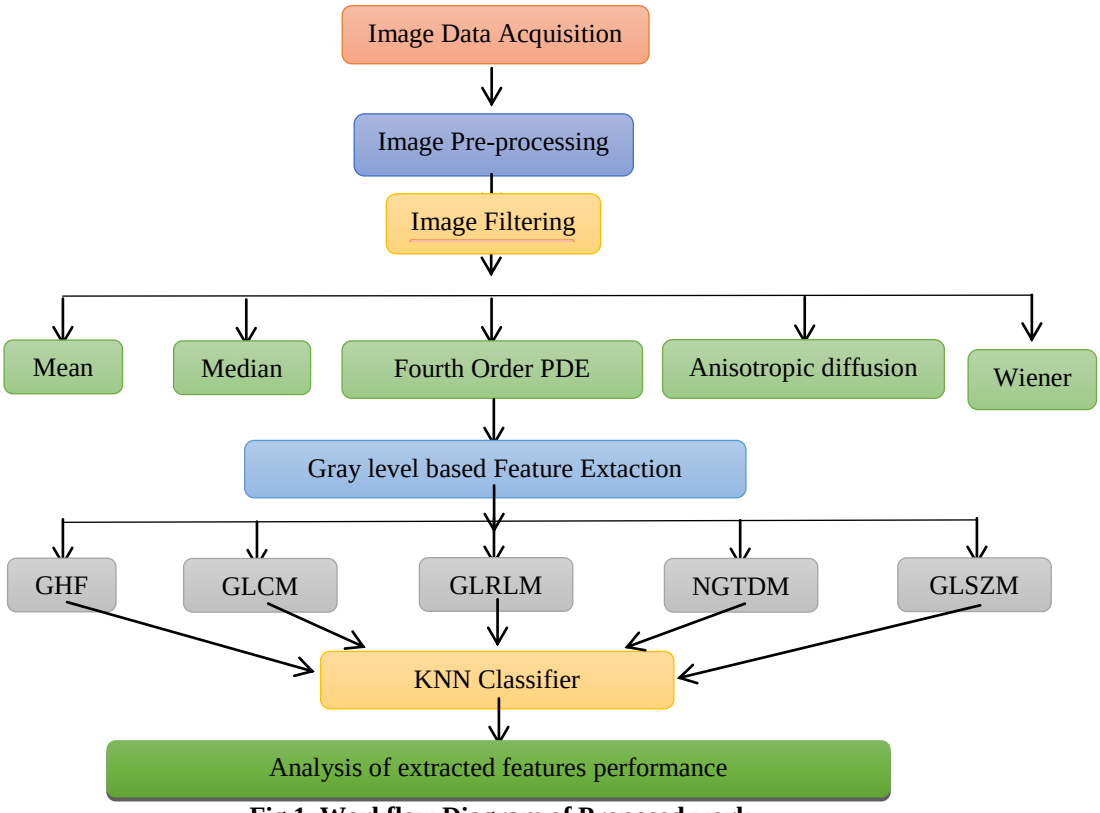


Fig 1. Workflow Diagram of Proposed work

2.2. Image Pre-processing:

The input of the X-ray pictures of the chest is in RGB, and for improved image quality, noise is decreased in this next step of the process where the image is changed into grayscale in accordance with the norms and guidelines of MATLAB. Also, this procedure is essential since it enhances certain image qualities that are important for subsequent processing while also improving the image data. So that the dataset images are converted into a

grayscale image. When the images are converted into grayscale, the images only have 0 to 255-pixel intensities. So we can find the feature extraction values and classification of the images much more easily.

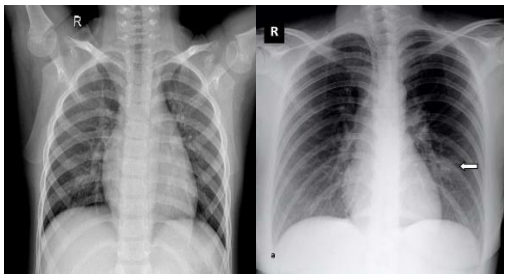


Fig. 2 Normal vs. Covid-19 X-ray image

2.3. Image Filtering:

In the image processing technique, there are many image filtering techniques. Filtering the image is a necessary step in image processing. All the data set images have some noise while scanning and capturing. Filters reduce the noise from the original images. Different filters reduce the different noise appropriately. In this filtering step, some filters are used in the dataset images for reducing noise. In this Mean, Median, Fourth order PDE, Anisotropic diffusion, and Wiener filter techniques are used in the taken images. The MSE, PSNR, and SNR values are found from the filtered images. Based on the result of the MSE, PSNR, and SNR the image filtering technique is chosen for the further process. Because the best filter technique gives better images we analyse the image in a better way. Figure. 3 shows that the Mean, Median, Fourth order PDE, Anisotropic diffusion, and Wiener filter images.

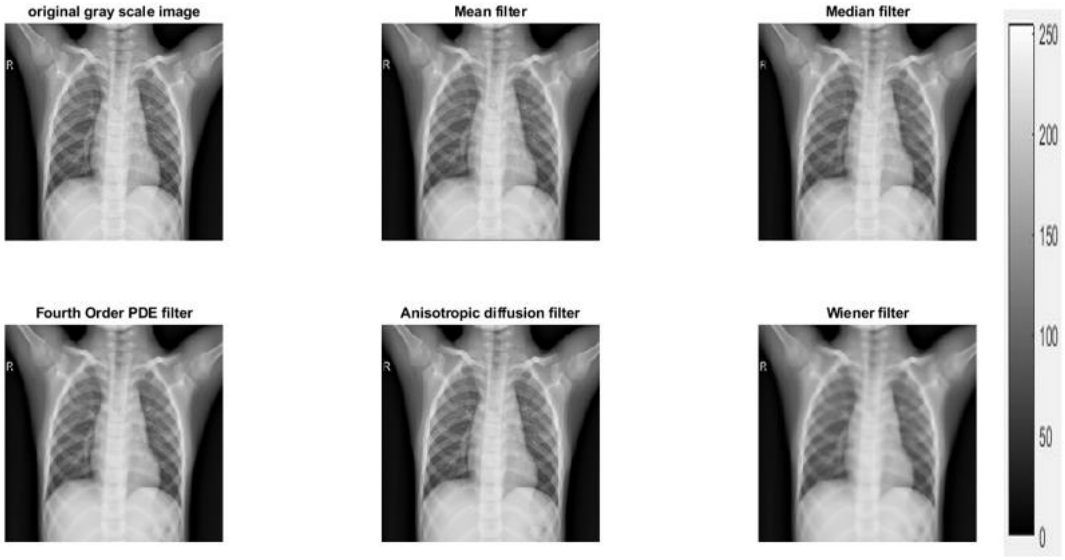


Fig. 3. Original grayscale and Filtered image

Figure 3 and 4 dipicts filtered images and their histogram and Table 1. depicts the performance of the filters in chest X-ray images.

PDEs are excellent image denoising alternatives. Perona and Malik introduced the second-order nonlinear PDE in the 1990s and its different variations are one of the most used PDE-based denoising techniques [6]. The second-order PDEs have the drawback of potentially causing blocky distortions in the image [7,8].

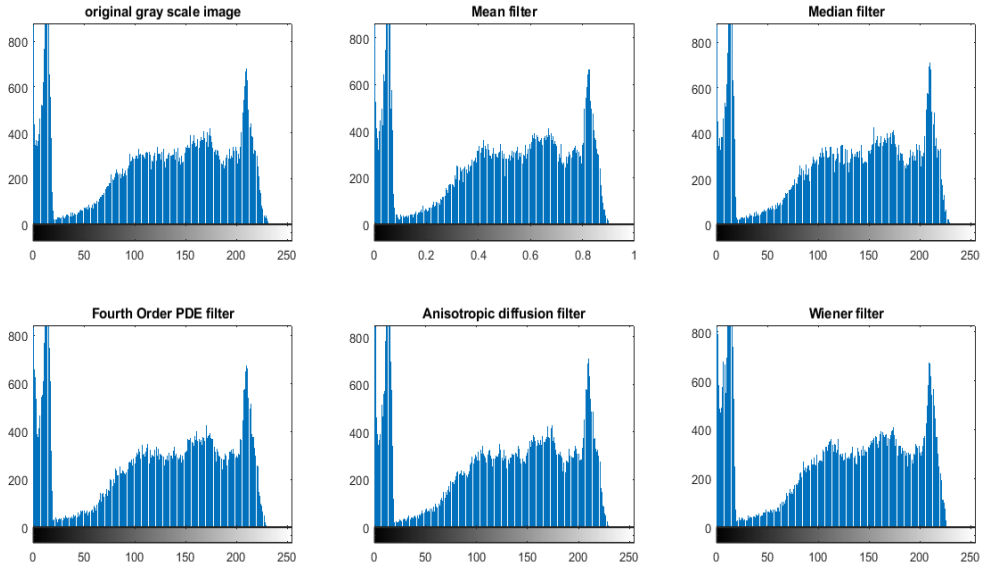


Fig. 4. Histogram of Original and Filtered image

FILTERS	PSNR	SNR	MSE
Mean	5.33727425	1.90E+04	2.48E+02
Median	36.252845	15.40976	184
Fourth order PDE	42.0276935	4.076721	8
Anisotropic diffusion	38.5135693	9.156372	17
Wiener	35.7364783	17.35532	28

Table 1. Filters performance in X-ray images of the chest

3. Material and Methods

The procedure of feature extraction keeps the original information in the dataset images while converting raw data into effective features that are numerical. Compared to using the raw data directly, it produces better outcomes. By extracting features from the input data, feature extraction improves the accuracy of learned models. This basic concept stage lowers the data's dimensionality by eliminating redundant data. Of course, it speeds up inference and training.

In this work, the features from Gray Level Histogram Features (GLHF), Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray Level Size Zone Matrix (GLSZM) are extracted from the given images.

3.1. Gray level Histogram Feature (GLHF):

These textural characteristics offer details about the homogeneity, smoothness, flatness, and contrast of the image's intensity level distributions. The probabilistic distributions of the levels of intensity in the histogram bins are used to calculate the statistical textural properties of mean, variance, skewness, kurtosis, energy, and smoothness [9,10].

a. Mean:

$$Mean = \frac{1}{\mathcal{N}} \sum_{i} \sum_{j} p(i - \bar{i}, j - \bar{j})$$

b. Variance:

$$Variance = \frac{1}{\mathcal{N}} \sum_{i} \sum_{j} [p(i - \bar{i}, j - \bar{l}) - M]^2$$

c. Skewness:

$$Skewness = \frac{1}{\mathcal{N}} \sum_{i} \sum_{j} [p(i - \bar{i}, j - \bar{l}) - M]^3$$

d. Kurtosis:

$$Kurtosis = \frac{1}{\mathcal{N}} \sum_{i} \sum_{j} [p(i - \bar{i}, j - \bar{l}) - M]^2 - 3$$

e. Energy:

$$Energy = \frac{1}{\mathcal{N}} \sum_{i} \sum_{j} p(i - \bar{i}, j - \bar{l}) \, . \, p(i - \bar{i}, j - \bar{l})$$

d. Smoothness:

$$Smoothness = \frac{1}{\mathcal{N}} \sum_{i} \sum_{j} [p(i - \bar{i}, j - \bar{l}) \, . \, log2p(i - \bar{i}, j - \bar{l})]$$

The discrete image is represented as $p(i, j)$, where $\mathcal{N}$ is the chest region's overall pixel count.

Features/Images	Normal	Abnormal
Mean	4.3276	6.9094
Variance	4.2155	3.0074
Skewness	-0.2276	-0.8655
Kurtosis	1.9169	3.7387
Energy	0.1445	0.1878
Smoothness	1.9817	1.8197

Table 2. Feature values of Gray Level Histogram Features

The above Table 2. Depicts the Normal and Covid-19 images of the gray level histogram feature values.

3.2. Gray Level Cooccurance Matrix (GLCM)

A co-occurrence matrix is a matrix built through an image to describe the allocation of co-occurring image pixels at a specific offset. It is a texture analytical method with a wide range of uses, including the analysis of medical images. [11]. The second-order statistical texture analysis method is called GLCM. It estimates the regularity with which a group of pixels occurs together in an image at a particular range θ and direction d. It also analyses the coordinates between pixels.In the Grey level Coocurance Matrix feature extraction technique some of the features are calculated using the following equations:

a. Auto Correlation:

$$autocorrelation = \sum_{i=1}^{N_m} \sum_{j=1}^{N_m} (i.j) P_{ij}$$

b. Contrast:

$$contrast = \sum_{i=1}^{N_m} \sum_{j=1}^{N_m} (i - j)^2 P_{ij}$$

c. Correlation:

$$correlation = \sum_{i=1}^{N_m}\sum_{j=1}^{N_m}\frac{(i.j)P_{ij-}\mu_x\mu_y}{\sigma_x\sigma_y}$$

d. Energy:

$$energy = \sum_{i=1}^{N_m}\sum_{j=1}^{N_m}P_{ij}^2$$

e. Entropy:

$$entropy = -\sum_{i=1}^{N_m}\sum_{j=1}^{N_m}P_{ij}\ log\ P_{ij}$$

f. Homogenity:

$$homogenity = \sum_{i=1}^{N_m}\sum_{j=1}^{N_m}\frac{1}{1+(i-j)^2}P_{ij}$$

g. Sum average:

$$sum\ average = \frac{1}{N_m.N_m}\sum_{i=1}^{N_m}\sum_{j=1}^{N_m}[i.P_{ij}+j.P_{ij}]$$

h. Dissimilarity:

$$dissimilarity = \sum_{i=1}^{N_m}\sum_{j=1}^{N_m}|i-j|\ P_{ij}$$

Here, N_m is the maximum discrete intensity level in the picture. P_{ij} be considered the normalized co-occurance matrix, genrealized for any direction d and angle θ.

Features / Images	**Normal**	**Abnormal**
Autocorrelation	18.87412	42.4283
Contrast	0.345457	0.171214
Correlation (Corrm)	0.954167	0.96735
Correlation (Corrp)	0.954167	0.96735
Cluster Prominence	397.7761	405.0116
Cluster Shade	-6.42803	-33.2191
Dissimilarity	0.262134	0.112251
Energy	0.095888	0.188967
Entropy	2.659113	2.065455
Homogeneity (homom)	0.879694	0.949143
Homogeneity (homop)	0.1679	0.948337
Maximum probability	0.1679	0.298367

Sum of sqaures: Variance	18.88334	42.24982
Sum average	7.817529	12.63224
Sum variance	44.12851	124.0465
Sum entropy	2.395278	1.967552
Difference variance	0.345457	0.171214
Difference entropy	0.629655	0.353479
Information measure of correlation1	-0.61944	-0.77885
Informaiton measure of correlation2	0.952398	0.963639
Inverse difference (INV) is homom	0.988033	0.988033
Inverse difference normalized (INN)	0.997684	0.997684

Table 3. Feature values of GLCM

Table 3. gives the GLCM extracted values of the Normal and Covid-19 images.

3.3. Gray Level Run-Length Matrix (GLRLM)

Greater-order statistics characteristics of texture to be extracted, it used in the GLRLM technique. A line of pixels in a specific direction with the same intensity value is known as a grey-level run. The quantity is known as the run-length, and the number of these pixels determines the value of a grey level's run length. The grey-level run length can be calculated using the blow equations [12]:

$$N = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij}, \;\; 1 \le N_z(\theta) \le N_p$$

a. Short Run Emphasis:

$$SRE = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} \frac{P_{ij}}{j^2} / \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij}$$

b. Long Run Emphasis:

$$LRE = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} j^2 P_{ij} / \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij}$$

c. Gray Level Non uniformity:

$$GLN = \sum_{i=1}^{N_m}\left(\sum_{j=1}^{N_n} P_{ij}\right)^2 / \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij}$$

d. Run Percentage:

$$RP = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij} \; / N$$

e. Run Length Non uniformity:

$$RLN = \sum_{j=1}^{N_m}\left(\sum_{i=1}^{N_n} P_{ij}\right)^2 / \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij}$$

f. Low Gray level Run Emphasis:

$$LGRE = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} \frac{P_{ij}}{i^2} / \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij}$$

g. High Gray level Run Emphasis:

$$HGRE = \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} i^2 P_{ij} \,/\, \sum_{i=1}^{N_m}\sum_{j=1}^{N_n} P_{ij}$$

Where, N_m and N_n are the numbers of the pixel value and run length, respectively, in the image.

Features/Images	Normal	Abnormal
SRE	0.76677	0.67065
LRE	240.96	124.998
GLN	0.00673	0.00917
RP	0.5451	0.41713
RLN	0.58989	0.38049
LGRE	0.00117	0.00017
HGRE	16201.9	30921.4

Table 4. Feature values of GLRLM

Table 4. shows the GLRLM extracted values of the Normal and abnormal images.

3.4. Neighboring Gray Tone Difference Matrix (NGTDM)

The NGTDM calculates the total deviations among a pixel's or voxel's mean gray level and its neighboring pixels within a specified distance. NGTDM complexity, activity, and coarseness are important characteristics. An ROI madeup of larger regions with relatively uniform grey levels will have a greater coarseness value. Coarseness tends to reflect the gray-level variation between the input image or voxel and its neighborhood and thus describes the spatial rate of variations in gray-level intensities. Contrarily, busyness is reflected by quick variations in the gray-level among the core pixel's or voxel's and the surrounding areas making an ROI made up of numerous small areas with noticeably varied grey levels busier [13].

The total of the grey level distances d between voxels with intensity value i and their close neighbors mean intensities $\mathcal{A}_i$ are represented by the i^{th}value of the $s(i|d)$.

a. Coarseness:

$$Coarseness = [\epsilon + \sum_{n=1}^{N_g} p(i)s(i)]^{-1}$$

b. Contrast:

$$Contrast = \left(\frac{1}{N_p(1-N_p)}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i)p(j)(i-j)^2 . (\frac{1}{N}\sum_{i=1}^{N_g} s(i)\right)$$

c. Busyness:

$$Busyness = \frac{\sum_{n=1}^{N_g} p(i)s(i)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} |ip(i) - jp(j)|}$$

d. Complexity:

$$Complexity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} |i-j| . \frac{p(i)s(i) + p(j)s(j)}{N(p(i)+p(j))}$$

e. Strength:

$$Strength = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\{p(i)+p(j)(i-j)^2}{\epsilon+\sum_{n=1}^{N_g}s(i)}$$

The below table 5. dipicts the extracted feature values of the normal and Covid-19 images.

Features/Images	Normal	Abnormal
Coarseness	0.00158	0.00148
Contrast	0.28534	0.10645
Busyness	0.02532	0.01078
Complexity	48301.1	37697
Strength	36.716	76.5903

Table 5. NGTDM features and their values of image of X-ray

3.5. Gray Level Size Zone Matrix (GLSZM)

The volume of associated, homogenous zones of a particular size and intensity within the volume is described by a grey-level size-zone matrix. The (m,n)th item of the GLSZM p(m,n) indicates the quantity of interconnected gray level m and size n regions. In order to describe heterogeneity at a local level, GLSZM characteristics thus describe affected homogenous regions within the affected volume [14]. Let:

- in the current GLSZM, $p(m,n)$ is the $(i,j)^{th}$element.
- $N_g$the count of image's distinct pixel intensities,
- $N_z$the extent of the most substantial, homogenous area within the amount of interest,
- $N_s$the overall number of homogeneous areas (zones), where $N_s = \sum_{m=1}^{N_g}\sum_{n=1}^{N_z}p(m,n)$
- $p_z$the distribution of all the zones with sizes j, added together, here $p_z(n) = \sum_{m=1}^{N_g}p(m,n)$
- $p_g$the distribution of grey levels mover total amount of zones, where $p_g(m) = \sum_{n=1}^{N_z}p(m,n)$
- $N_p$the quantity of image voxels, where $N_p = \sum_{n=1}^{N_z}np_r$
- $\mu_r$average zone size, where$\mu_r = \sum_{m=1}^{N_g}\sum_{n=1}^{N_z}jp(m,n|\theta)$
- μ_g the average gray-level, where $\mu_g = \sum_{m=1}^{N_g}\sum_{n=1}^{N_z}ip(m,n|\theta)$

a. Small Zone Emphasis (SZE)

$$SZE = \sum_{n=1}^{N_z}\frac{p_z}{n^2}$$

b. Large Zone Emphasis (LZE)

$$LZE = \sum_{n=1}^{N_z}n^2p_z$$

c. Gray Level Non-uniformity (GLN)

$$GLN = \sum_{m=1}^{N_g}p_g^2$$

d. Zone-Size Non-uniformity (ZSN)

$$ZSN = \sum_{m=1}^{N_g}p_z^2$$

e. Zone Percentage (ZP)

$$ZP = \frac{N_s}{N_p}$$

f. Low Gray-Level Zone Emphasis (LGZE)

$$LGZE = \sum_{m=1}^{N_g} \frac{p_g}{m^2}$$

g. High Gray-LevelZone Emphasis (HGZE)

$$HGZE = \sum_{m=1}^{N_g} m^2 p_g$$

h. Small Zone High Gray-Level Emphasis (SZLGE)

$$SZLGE = \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} \frac{p(i,j)}{m^2 n^2}$$

i. Large Zone Low Gray-Level Emphasis (LZLGE)

$$LZLGE = \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} \frac{p(m,n)n^2}{m^2}$$

j. Large Zone High Gray-Level Emphasis (LZHGE)

$$LZHGE = \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} p(m,n)n^2 m^2$$

k. Gray-Level Variance (GLV)

$$GLV = \frac{1}{N_g X N_Z} \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} \left(ip(m,n) - \mu_g\right)^2$$

l. Zone Size Variance (ZSV)

$$ZSV = \frac{1}{N_g X N_Z} \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} (np(m,n) - \mu_z)^2$$

Features / Images	**Normal**	**Abnormal**
SZE	0.76677	0.67983
LZE	240.96	625.64
GLN	0.00673	0.0062
ZSN	0.5451	0.42852
ZP	0.58989	0.35265
LGZE	0.00117	0.00028
HGZE	16201.9	27944.5
SZLGE	0.0006	0.00016
SZHGE	12320.6	17669.1
LZLGE	235.47	0.02942

LZHGE	103378	3.8E+07
GLV	8.69E-05	0.00014
ZSV	1.46E-08	6.16E-08
SZE	0.76677	0.67983
LZE	240.96	625.64
GLN	0.00673	0.0062
ZSN	0.5451	0.42852
ZP	0.58989	0.35265
LGZE	0.00117	0.00028
HGZE	16201.9	27944.5

Table 6. GLSZM features and their values of X-ray images

Table. 6 gives extracted values of the GLSZM feature values of the X-ray pictures.

4. Classification

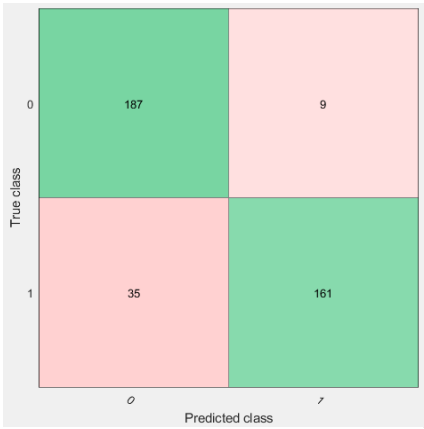
4.1. K Nearest Neighbors (KNN)

K-Nearest Neighbors, sometimes called as KNN, is a well-known supervised machine learning algorithm for resolving regression defets and classification. The basic goal of the KNN method is to anticipate a new sample point's categorization using data points that have been divided into various classifications.Due to its exceptionally accurate predictions, the Proposed technique can compete with the most precise models. As a result, the KNN method can be used for applications that need great accuracy but don't need a model that can be read by humans. The distance measurement affects how accurate the projections are.

The KNN classifier uses K patterns to get around issues and increase the approach's resilience. As the KNN classifier doesn't imply any assumptions about the statistical makeup of the data, it is regarded as a non parametric classifier [15].The K-nearest neighbor classifier is carried out using a training data set that includes both the inputs and the target output, and the standard set to the testing data, which only comprises the input data.by achieving a clear majority, classification values of the most nearby reference points, Its classification is determined by the distance from the undetermined to the K closest neighbors [16].

5. Experimental Results

The database used to create the input images is made up of a total of 392 pictures of the chest.Those images are all separated into two categories: Covid-19 and Normal Images. The KNN classifier in this experiment receives input from five different types of gray level based feature extraction values. After that, the performance of the feature approaches is assessed using common parameters.



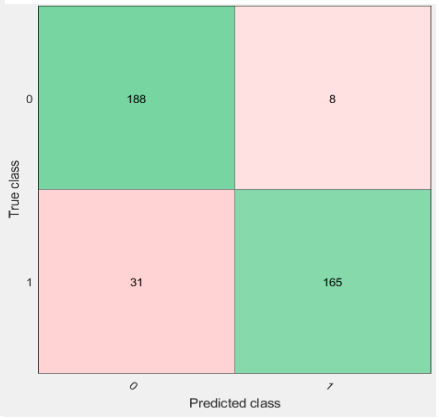
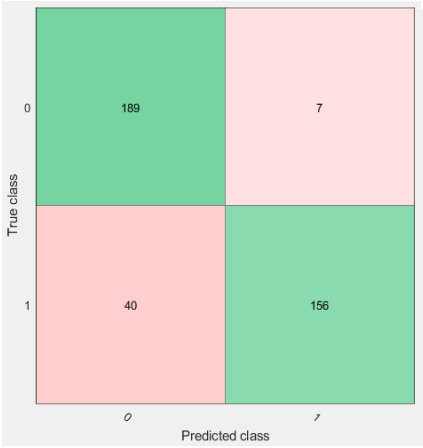


Fig. 5 Cofusion Matrix for GLHF **Fig. 6 Cofusion Matrix for GLCM**



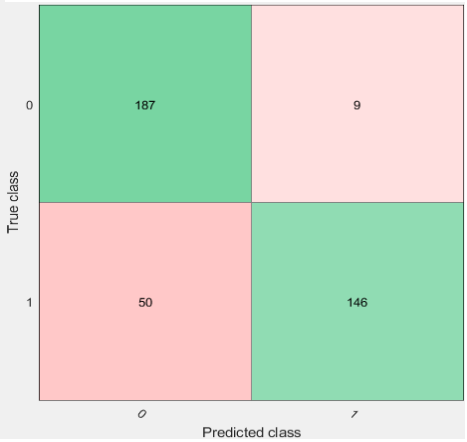


Fig. 7 Cofusion Matrix for GLRLM **Fig. 8 Cofusion Matrix for NGTDM**

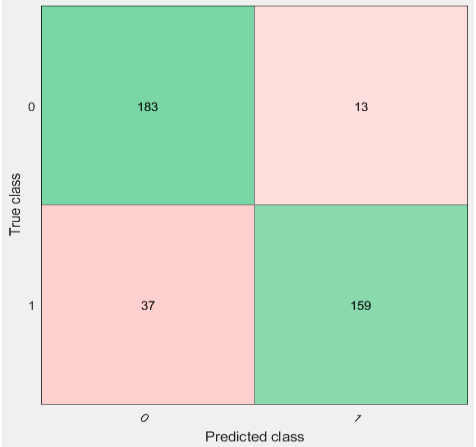


Fig. 9 Cofusion Matrix for GLSZM

Using data from Figure. 5, 6, 7, 8, and 9, the standard parameters like sensitivity, specificity, accuracy are calculated for all the feature extraction techniques. These measures are in the Table 7. The following some formulas are used for to find the standard parameters.

A test's ability to accurately differentiate between patients and healthy instances is a measure of its accuracy. To determine a test's accuracy, it is important to calculate the total of actual positive and negative cases across all examined cases. The same can be mathematically represented.

$$\text{Sensitivity} = \frac{TP}{TP+FN} * 100$$

$$\text{Specificity} = \frac{TN}{TN+FP} * 100$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

$$\text{Error rate} = \frac{FN+FP}{TP+TN+FN+FP} * 100$$

$$\text{Precision} = \frac{TP}{TP+FP} * 100$$

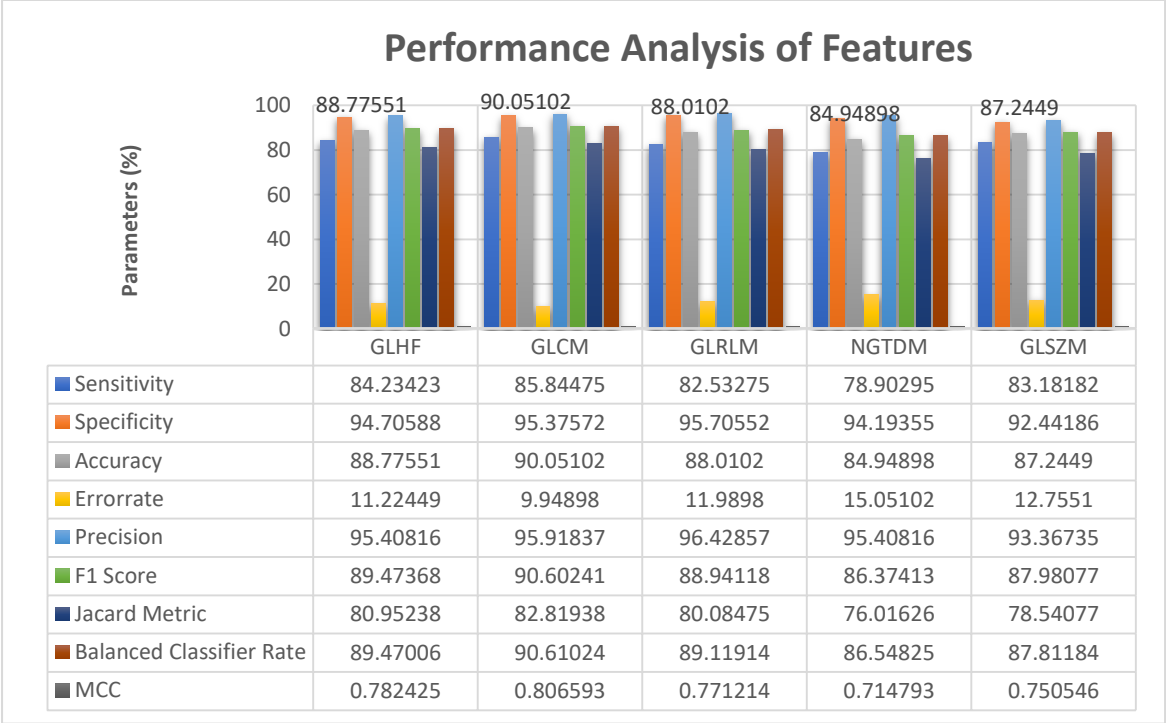
$$\text{F1 Score} = \frac{2 * TP}{(2 * TP) + FN + FP} * 100$$

$$\text{Jacard Metric} = \frac{TP}{TP + FN + FP} * 100$$

$$\text{Balanced Classifier Rate} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

$$\text{MCC} = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} * 100$$

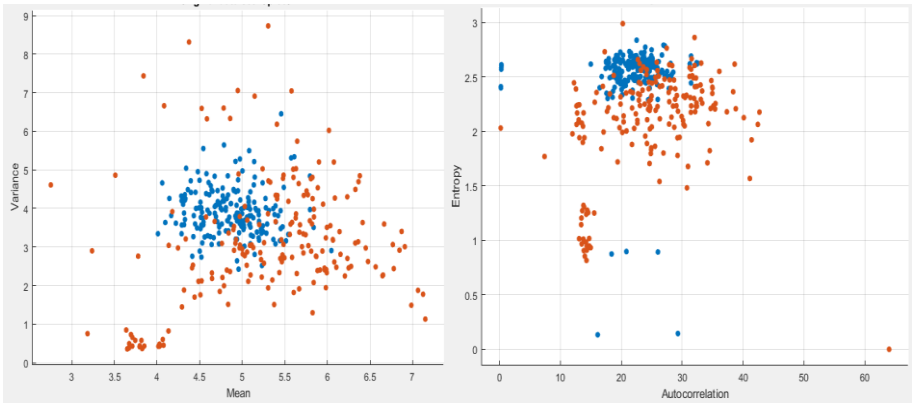
Where,True Positive (TP) - Patients who have been properly diagnosed as ill, False positives (FP) occur when healthy people are mistakenly classified as ill.Healthy individuals were appropriately classified as such by True Negative, (TN). False Negative, or FN, is when sick people are mistakenly classified as healthy [17].



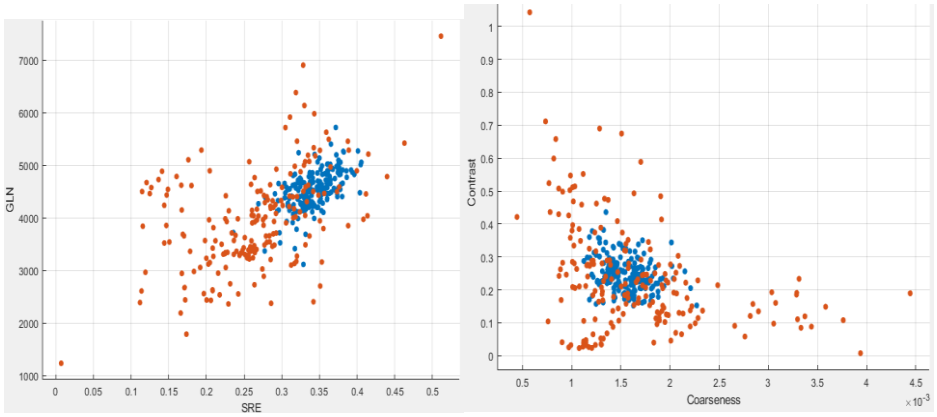
	GLHF	GLCM	GLRLM	NGTDM	GLSZM
Sensitivity	84.23423	85.84475	82.53275	78.90295	83.18182
Specificity	94.70588	95.37572	95.70552	94.19355	92.44186
Accuracy	88.77551	90.05102	88.0102	84.94898	87.2449
Errorrate	11.22449	9.94898	11.9898	15.05102	12.7551
Precision	95.40816	95.91837	96.42857	95.40816	93.36735
F1 Score	89.47368	90.60241	88.94118	86.37413	87.98077
Jacard Metric	80.95238	82.81938	80.08475	76.01626	78.54077
Balanced Classifier Rate	89.47006	90.61024	89.11914	86.54825	87.81184
MCC	0.782425	0.806593	0.771214	0.714793	0.750546

Table 7. Perfromance of the Gray Level based features using KNN classifier

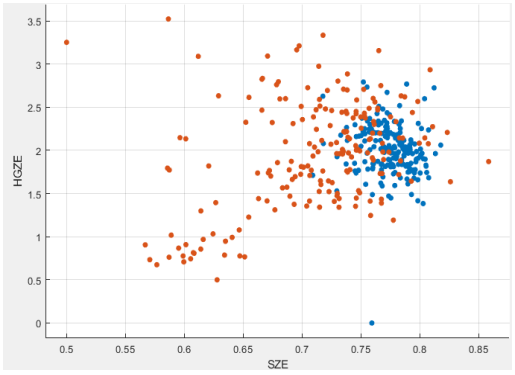
The evaluation measurements have shown in Table 7. In that KNN classifier with five fearures shows, GLCM features gives the better result of 90.1% classification rate and NGTDM gives the lower classification rate of 84.9%.



(a) (b)



(c) (d)

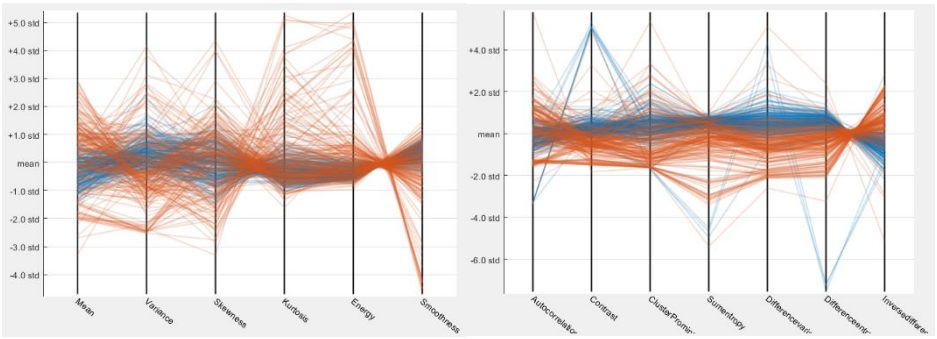


(e)

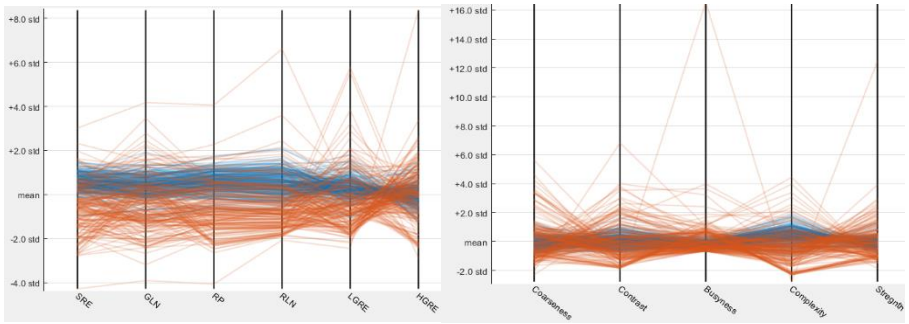
Fig. 10 Scatter-plot of the Gray Level based Feature extraction techniques (a) Scatterplot for GLHF,

(b) Scatterplot for GLCM, (c) Scatterplot for GLRLM, (d) Scatterplot for NGTDM, (e) Scatterplot for GLSZM

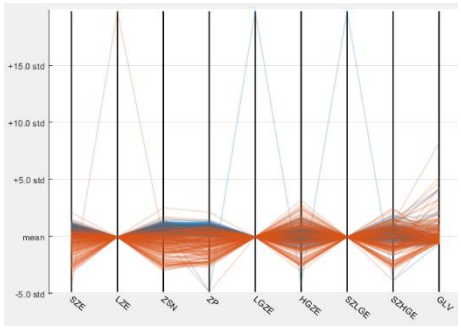
Figure 10. gives the scatter plot of the five gray level based feature extraction techniques using KNN classifiers. Figure 11. show the parallel coordinate plots of the five feature extraction techniques.



(a) (b)



(c) (d)



(d)

Fig. 11 Parellel coordinates plot for (a) GLHF,(b) GLCM, (c) GLRLM, (d) NGTDM, (e) GLSZM

6. Conclusion

In this work, it was suggested that X-ray images of chest may be accurately classify and identify abnormalities. Using a KNN classifier, the input X-ray image was classified as Normal or Covid-19. Five extracted feature techniques—GLHF, GLRLM, GLCM, NGTDM, and GLSZM—were proposed using a KNN classifier, and GLCM generated the most effective results, with a classification rate of 90.1%. The GLCM with KNN classifier algorithm is a better algorithm, with 90.1% of correct classification, relatively higher than the other four technique, according to the results obtained. An incorrectly classified image is processed further. This GLCM with KNN classifier approach will be improved by adding the ability to detect abnormalities in X-rays of the chest image datasets.

References

[1] WHO. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-openin g-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

[2] Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA. 2020;323(18):1843–1844. doi:10.1001/jama.2020.3786
[3] Xu XT, Chen P, Wang JF, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for the risk of human transmission. Sci China Life Sci. 2020 doi: 10.1007/s11427-020-1637-5.
[4] Ksiazek, Thomas G., et al. "A novel coronavirus associated with severe acute respiratory syndrome." New England journal of medicine 348.20 (2003): 1953-1966.
[5] Chandra T.B., Verma K. Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In: Chaudhuri B.B., Nakagawa M., Khanna P., Kumar S., editors. Proceedings of third international conference on computer vision & image processing. Springer; Singapore: 2020. pp. 21–33.
[6] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Machine Intell., vol. 12, pp. 629–639, July 1990.
[7] Yu-Li You and M. Kaveh, "Image enhancement using fourth order partial differential equations," Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers (Cat. No.98CH36284), Pacific Grove, CA, USA, 1998, pp. 1677-1681 vol.2, doi: 10.1109/ACSSC.1998.751611.
[8] Y. Wang, Y. Zhang and T. Yu, "Noise Removal Using Edge-Preserving Fourth-Order Partial Differential Equations," 2009 2nd International Congress on Image and Signal Processing, Tianjin, China, 2009, pp. 1-5, doi: 10.1109/CISP.2009.5304166.
[9] Alnihoud J., "Content-Based Image Retrieval System Based on Self Organizing Map, Fuzzy Color Histogram and Subtractive Fuzzy Clustering," International Arab Journal of Information Technology, vol. 9, no. 5, pp. 452- 458, 2012.
[10] S. Sergyan, "Color histogram features based image classification in content-based image retrieval systems," 2008 6th International Symposium on Applied Machine Intelligence and Informatics, Herlany, Slovakia, 2008, pp. 221-224, doi: 10.1109/SAMI.2008.4469170.
[11] Sivakumar, Sanjayprabu & Kumar, R. & Somasundaram, K. & Ramamoorthy, Karthikamani. (2023). Brain tumor detection using GLCM features and classifiers. AIP Conference Proceedings. 2725. 020010. 10.1063/5.0125260.
[12] S. Sanjayprabu, R. Sathish Kumar, K. Somasundaram and R. Karthikamani, "Mathematical Model for Anisotropic diffusion Filter and GLRLM Feature Extraction to Detect Covid-19 from Chest X-Ray Images," 2022 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2022, pp. 1-5, doi: 10.1109/STCR55312.2022.10009430.
[13] Chen, S.; Harmon, S.; Perk, T.; Li, X.; Chen, M.; Li, Y.; Jeraj, R. Using neighborhood gray tone difference matrix texture features on dual time point PET/CT images to differentiate malignant from benign FDG-avid solitary pulmonary nodules. Cancer Imaging 2019, 19, 56.
[14] Chu, A., Sehgal, C. M., & Greenleaf, J. F. (1990). Use of gray value distribution of run lengths for texture analysis. Pattern Recognition Letters, 11(6), 415-419. https://doi.org/10.1016/0167-8655(90)90112-F.
[15] Selvaraj, Henry, et al. "Brain MRI slices classification using least squares support vector machine." International journal of intelligent computing in medical sciences & image processing 1.1 (2007): 21-33.
[16] Hiremath, P. S., Parashuram Bannigidad, and Manjunath Hiremath. "Automated Identification and classification of rotavirus-A particles in digital microscopic images." IJCA, Special Issue on RTIPPR 1 (2010): 16-20.
[17] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification," in IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973, doi: 10.1109/TSMC.1973.430931.