

EVALUATING THE PERFORMANCE OF INVARIANT MOMENTS AND SHAPE FEATURES TO CLASSIFY NORMAL AND COVID SUBJECTS

Dushyanth Balasai Jakkam^{1,*} and Ramesh Munirathinam²

Received: April 20, 2021; Revised: October 10, 2021; Accepted: December 14, 2021

Abstract

Alteration in lung shape and texture are considered as the significant pathological cores for analysing the severity of COVID-19 disease. The main objective of this analysis is to estimate the shape deformation of lung in CT images due to the incidence of COVID-19. Totally 176 samples are considered for the analysis. Comparison between invariant moments and shape features is carried out to analyse the geometric variations of lungs. The extracted features are used to classify normal and COVID subjects using neural network classifier. From the results, it is observed that feature values obtained using invariant moments are found to be statistically significant ($p < 0.05$) compared to shape features. Inter (normal and COVID) subjects mean and STD value, for invariant moments M5 feature is $(0.76 \pm 0.07, 0.02 \pm 0.10)$ and for shape features extent feature is $(0.61 \pm 0.09, 0.54 \pm 0.14)$. Invariant moments obtained 0.993 of accuracy and 0.992 of F1 score using neural network classifier. In this study, it is found that invariant moments perform better in comparison with shape features demonstrating significant differences between normal and COVID CT scan images.

Keywords: COVID-19, 2D shape analysis, Invariant moments, Shape feature

Introduction

A novel coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) caused an outbreak of COVID-19 (Al-Quteimat and Amer 2020). Fever, cough, shortness of breath and muscle pain are common symptoms of the disease. Broadly, as of March 2021, 109,594,835 confirmed COVID-19 cases including 2,424,060 deaths are reported to WHO (Öztürk *et al.*, 2021). Medical imaging techniques such as Ultrasound imaging, X-ray and Computed Tomography (CT) are main diagnostic methods in diagnosing COVID-19. Due to high radiation

exposure, usage of X-rays for COVID diagnosis is subtle. CT scans are capable of detecting COVID-19 complications (Li 2020). CT becomes an important imaging modality in diagnosing patients and reports on radiological appearances of COVID-19 pneumonia are emerging.

Dogantekin presented a technique based on the invariant moments using an adaptive network based fuzzy inference system classifier for recognising human parasite eggs. Hu invariant moment exhibits better differentiation of human parasitic eggs (Dogantekin *et al.*, 2008). Ramteke

¹ Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Thandalam, Chennai, Tamilnadu, Pincode: 602105. E-mail: dushyanth7jakkam@gmail.com

² Department of Biomedical Engineering, Karpagam Academy of Higher Education, Coimbatore- 641021. E-mail: rameshmunirathinam@gmail.com

* Corresponding author

analysed the Invariant moments based feature extraction for the recognition of devanagari vowels by using Fuzzy Gaussian membership function for the classification (Ramteke, 2010). Asodekar used shape features and machine learning algorithms for detecting the brain tumour in the brain using MRI scans. The classification is made using the decision tree and Multi Layer Perceptron (MLP) algorithms (Asodekar *et al.*, 2019).

Materials and Methods

For sample size calculation G power software is used. Effect size, standard error rate, algorithm power are the pre-required factors which are given as 0.3, 0.05, 0.80 respectively. Allocation rate is expressed as N_2/N_1 , where N_1 is the normal subject and N_2 is the abnormal subject (Faul *et al.*, 2009). For the given inputs, sample size of 176 per each group is obtained. The images (normal and COVID-200) used in this work are acquired from 'Kaggle', a publicly available open access data repository (Ma Jun *et al.*, 2020) the dataset comprises normal and COVID CT scan images of 20 subjects each. Each subject contains 300 slices each. 10 images, from each subject, are extracted for a sample test by choosing the middle slice information to obtain maximum lung region. The selected slices are grouped based on its category since the analysis of normal and covid CT scans is prominent.

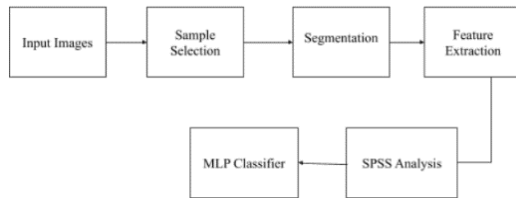


Figure 1. Block diagram representation of the analysis

The hardware requirement for this analysis is an x64 based processor of Intel with RAM of 8GB and 64-bit OS. The software requirements are SPSS and MATLAB. The workflow of this analysis is represented in Figure 1. The lung images are segmented using level set method. The feature extraction for geometric analysis is made with these segmented images by using Invariant moments and shape features in MATLAB software. The feature values are being analysed and statistical values are obtained. Graphs are plotted in the SPSS software using feature values which are represented in Figure 4 and Figure 5. From the extracted feature

values the classification is performed by using Neural Network classifier.

Seven Invariant Moments:

The invariant moments are invariant under translation, scaling, rotation and reflection. Seven invariant moments are being used in classical mechanics and statistical theory. In image processing the moments are used as feature vectors for classification, image texture attributes and shape descriptors of objects in image processing (Huang and Leng, 2010).

The two dimensional moments of order $(u+v)$ with image intensity distribution, $f(x,y)$ is expressed as:

$$m_{sv} = \iint x^s y^v f(x, y) \cdot \partial x \cdot \partial y \quad (1)$$

where u, v are positive integers.

Invariant moments are variant to any distortions, so the central moments are defined as,

$$\mu_{sv} = \iint (x - \bar{x})^s (y - \bar{y})^v f(x, y) \cdot \partial (x - \bar{x}) \cdot \partial (y - \bar{y}) \quad (2)$$

where $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$

The central moments are invariant under translation. The central moments work to normalize the moments with respect to size. This condition can be done by using the area (μ_{00}). The normalized central moments are obtained from

$$n_{sv} = \frac{\mu_{sv}}{\mu_{00}^r} \quad (3)$$

With these normalized moments, seven invariants moments of central moments are expressed as follows

$$\varphi_1 = \eta_{20} + \eta_{02} \quad (4)$$

$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (5)$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2 \quad (6)$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2 \quad (7)$$

$$\varphi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (8)$$

$$\varphi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (9)$$

$$\varphi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (10)$$

Moreover, mean and standard deviation are obtained for each feature.

Shape Features:

The binary image of the lungs have a number of properties for the MATLAB function region props (Dmitriev *et al.*, 2019). In this paper thirteen shape properties are used which includes area, circularity, filled area, convex area, Euler number, equivalent diameter, solidity, extent, perimeter, minimum feret angle, maximum feret angle, minimum feret diameter, maximum feret diameter (Mingqiang *et al.*, 2008).

Classification

Accuracy, precision, recall, F1 score and AUC are used in evaluating the classifier performance in differentiating normal and abnormal subjects using extracted Invariant and Shape features. Using a neural network classifier the performance metrics are measured as represented in Table 5 and Table 6.

- Accuracy checks the closeness of the measurement to a specific value. It shows whether FNs or FPs are more common.

where TP=true positives, TN=true negatives

FN=false negatives, FP=false positives

- Recall is the proportion of known positives which are predicted correctly.
- Precision defines the predicted positives proportion that are correct. It doesn't capture the TNs and FNs.
- The F score uses β parameter to control the recall balance and precision. With $\beta=1$, the most commonly used F1 score is reduced.

Results and Discussion

The CT scan images of normal and COVID subjects are considered for the analysis. Figure 2(a) and Figure 2(b) shows the CT scan images of COVID subjects. To analyse the shape of the lungs, only the lung region from the CT scan images are segmented using level set method. The segmented images are converted into binary format. The segmented lung binary CT scan images of normal subjects are represented in Figure 2(c) and Figure 2(d).

When the normal subjects are multiplied with an infected lung mask, the binary images of COVID subjects are obtained. The segmented lung binary CT images of COVID subjects are represented in Figure 2(e) and Figure 2(f).

Table 1 and Table 2 represents the significant normal and normalised feature values of normal and COVID CT images for Invariant moments. Through the feature mean it is observed that mean value for

normal and COVID lungs are (0.23, 0.18). For the feature M5 the mean value of inter subjects are (0.76, 0.02). Moreover, most of the features have high mean value for normal subjects. This statistical difference between normal and COVID subjects indicate loss of shape in the lung due incidence of COVID. The obtained significant value is less than 0.05 for most of the features.

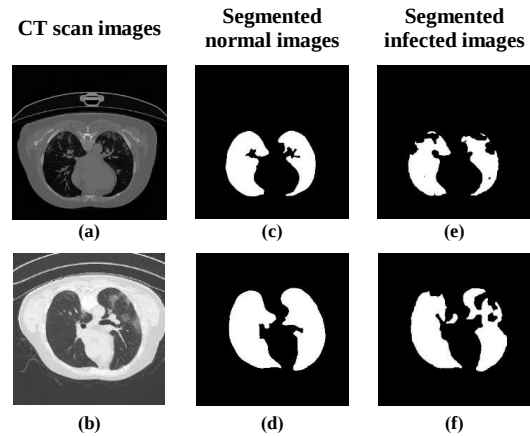


Figure 2. Representation of (a-b) normal CT scan images, corresponding (c- d) segmented lung binary CT scan images of normal subjects and (e-f) segmented lung binary CT scan images of COVID subjects

Table 1. Significant normal feature values of normal and COVID CT images using Invariant moments. Mean difference in the normal and COVID subjects indicate loss of shape in lungs due to incidence of COVID. Overlapping of feature values between normal and COVID subjects can be observed

Feature Name	Normal		p-value
	Normal	COVID	
	(Mean $\pm$ STD)		
Mean	0.33 $\pm$ 0.05	0.39 $\pm$ 0.11	1.51E-10
Variance	0.02 $\pm$ 0.01	0.05 $\pm$ 0.06	4.19E-08
Skewness	0.001 $\pm$ 0.002	0.01 $\pm$ 0.03	9.17E-06
Kurtosis	0.001 $\pm$ 0.002	0.01 $\pm$ 0.04	6.48E-05
M5	2.37E-06 $\pm$ 1.96E-05	0.001 $\pm$ 0.008	0.008399
M6	0.0001 $\pm$ 0.0004	0.005 $\pm$ 0.02	0.00057
M7	1.36E-06 $\pm$ 1.75E-05	0.0008 $\pm$ 0.0054	0.021871

STD-Standard Deviation

Table 3 and Table 4 represents the normal and normalised feature values of normal and COVID subjects for shape features. From the table, the mean value of feature Area for normalised inter subjects is 0.70, 0.64. Likewise from all the features it is noticeable that the normal subjects have high mean value compared to COVID subjects. It specifies the loss of region in the lungs due to incidence of COVID.

Table 2. Significant normalised feature values of normal and COVID CT images using Invariant moments. Mean difference in the normal and COVID subjects indicate loss of shape in lungs due to incidence of COVID

Feature Name	Normalised		p-value
	Normal (Mean $\pm$ STD)	COVID (Mean $\pm$ STD)	
Mean	0.18 $\pm$ 0.10	0.23 $\pm$ 0.17	0.000853
Variance	0.18 $\pm$ 0.13	0.10 $\pm$ 0.13	1.57E-07
Skewness	0.07 $\pm$ 0.14	0.05 $\pm$ 0.14	0.111226
Kurtosis	0.05 $\pm$ 0.12	0.03 $\pm$ 0.12	0.11579
M5	0.76 $\pm$ 0.07	0.02 $\pm$ 0.10	2.2E-255
M6	0.12 $\pm$ 0.21	0.03 $\pm$ 0.13	3.07E-07
M7	0.07 $\pm$ 0.06	0.05 $\pm$ 0.13	0.02027

STD-Standard Deviation

Table 3. Significant normal feature values of normal and COVID CT images using Shape features. Mean difference in inter subjects indicate loss of shape in lungs due to COVID incidence

Feature Name	Original		p-value
	Normal (Mean $\pm$ STD)	COVID (Mean $\pm$ STD)	
Area	45876.5 $\pm$ 8928.08	48386.85 $\pm$ 17859.05	0.07
Circularity	0.88 $\pm$ 0.22	0.78 $\pm$ 0.24	1.32E-05
Filled Area	45883.85 $\pm$ 8928.59	48693.92 $\pm$ 17988.48	0.04
Convex Area	70867.32 $\pm$ 13220.21	82842.16 $\pm$ 24869.72	4.13E-09
Euler Number	1.89 $\pm$ 0.34	2.39 $\pm$ 2.02	0.0006
Equiv Diameter	240.54 $\pm$ 23.54	243.69 $\pm$ 47.22	0.39
Solidity	0.64 $\pm$ 0.05	0.57 $\pm$ 0.10	5.17E-16
Extent	0.55 $\pm$ 0.05	0.49 $\pm$ 0.09	1.19E-15
Perimeter	828.23 $\pm$ 161.13	900.99 $\pm$ 232.32	0.0003
Min Feret Angle	22.77 $\pm$ 87.10	19.27 $\pm$ 87.68	0.68
Max Feret Angle	41.93 $\pm$ 160.07	34.48 $\pm$ 163.23	0.64
Min Feret Diameter	240.78 $\pm$ 34.11	254.94 $\pm$ 48.43	0.0007
Max Feret Diameter	343.66 $\pm$ 28.69	375.78 $\pm$ 50.03	3.29E-14

The extracted feature values are being classified using Neural Network classifier. Classifier performance in differentiating normal and COVID subjects using Invariant moments and shape features are evaluated which are represented in Table 5 and Table 6.

From Table 5 and Table 6, it is observed that the normalized feature values of Invariant moments (AUC-0.995, CA-0.993, F1 score-0.992) have significant performance metrics than

the normal feature values (AUC-0.698, CA-0.642, F1 score-0.641). Similarly, from the Table 4, normalized feature values (AUC-0.939, CA-0.880, F1 score-0.880) of shape features provide better performance metrics than the normal feature values (AUC-0.987, CA-0.978, F1 score-0.977). Normalized features show better metrics since the range of intensity values are brought into normal distribution. From the performance metrics of NN classifier on Invariant moments and shape features. It is identifiable that impact of NN is more on Invariant moments rather than shape features. Metric differences between normal and normalised features of both the techniques depicts impact of classification.

Table 4. Significant normalised feature values of normal and COVID CT images using Shape features. Mean difference in inter subjects indicate loss of shape in lungs due to COVID incidence

Feature Name	Original		p-value
	Normal (Mean $\pm$ STD)	COVID (Mean $\pm$ STD)	
Area	0.70 $\pm$ 0.15	0.64 $\pm$ 0.23	0.002
Circularity	0.63 $\pm$ 0.22	0.62 $\pm$ 0.19	0.78
Filled Area	0.70 $\pm$ 0.15	0.64 $\pm$ 0.23	0.0007
Convex Area	0.70 $\pm$ 0.15	0.65 $\pm$ 0.23	0.021
Euler Number	0.94 $\pm$ 0.17	0.39 $\pm$ 0.16	2.6E-113
Equiv Diameter	0.83 $\pm$ 0.09	0.79 $\pm$ 0.14	0.0004
Solidity	0.68 $\pm$ 0.10	0.62 $\pm$ 0.16	1.1E-05
Extent	0.61 $\pm$ 0.09	0.54 $\pm$ 0.14	6.41E-08
Perimeter	0.52 $\pm$ 0.18	0.55 $\pm$ 0.17	0.089
Min Feret Angle	0.69 $\pm$ 0.36	0.57 $\pm$ 0.45	0.002
Max Feret Angle	0.61 $\pm$ 0.45	0.59 $\pm$ 0.46	0.59
Min Feret Diameter	0.75 $\pm$ 0.12	0.73 $\pm$ 0.15	0.16
Max Feret Diameter	0.86 $\pm$ 0.08	0.83 $\pm$ 0.13	0.001

Table 5. Classification ability of Neural Network classifier using Invariant moments for normal and normalized feature values

Neural Network	Normal	Normalised
AUC	0.698	0.995
CA	0.642	0.993
F1 score	0.641	0.992
Precision	0.645	0.993
Recall	0.642	0.993

Table 6. Classification ability of Neural Network classifier using shape features for normal and normalized feature values

Neural Network	Normal	Normalised
AUC	0.939	0.987
CA	0.880	0.978
F1 score	0.880	0.977
Precision	0.880	0.978
Recall	0.880	0.978

Statistical values of extracted features are plotted in a bar graph using SPSS software. Figure 3 and Figure 4 represents invariant moments and shape features statistical analysis. Mean and standard deviation values of each feature can be observed. Mean difference between normal and COVID subjects represents variation in the lung shape.

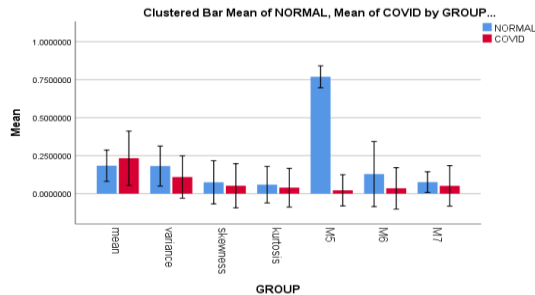
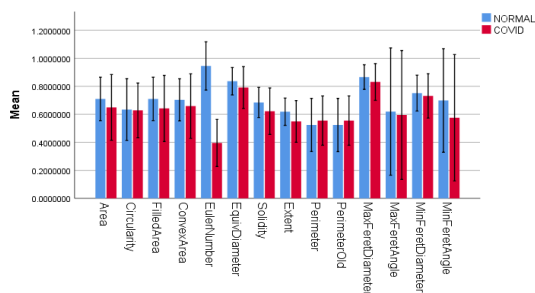
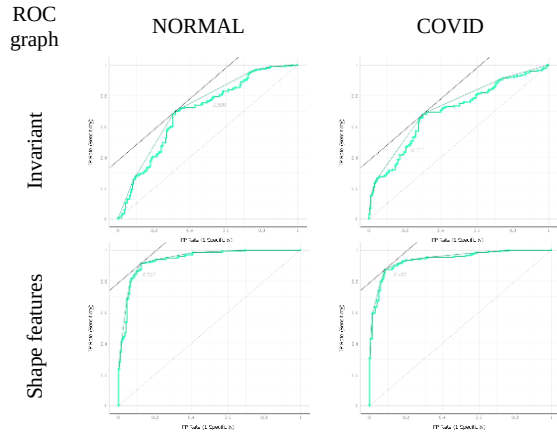
**Figure 3. Representative graphs of the statistical analysis for the extracted invariant moment features using SPSS. Mean difference with ± 1 SD between the normal and COVID subjects of each feature can be observed****Figure 4. Representative graphs of the statistical analysis for the extracted regionprops features using SPSS. Mean difference with ± 1 SD between the normal and COVID subjects of each feature can be observed**

Figure 5 represents the comparison of ROC graphs between Invariant moments and shape features specifying the classification ability of Neural Network classifier in classifying normal and

COVID subjects using the extracted feature values. From the graphs, it is noticeable that the invariant moments show better classification ability between the normal and COVID subjects than the shape features.

**Figure 5. ROC graphs represent Invariant moments signifying the classifying ability of the classifier in differentiating normal and COVID subjects using obtained feature values**

Dogantekin classified 16 different kinds of human parasite eggs using microscopic images. For extraction of features (Dogantekin *et al.*, 2008). Hu invariant moments was employed to differentiate normal and COVID subjects. Adaptive Network based Fuzzy Inference System (ANFIS) classifier is used for classification and achieved a performance rate of 93.49% and concluded as incorrect identification or classification is due to similar shape in the parasite egg. Ramteke segmented and analysed five handwritten devanagari vowels using fuzzy gaussian membership function for the classification (Ramteke, 2010). He obtained a performance rate of 94.56% for handwritten devanagari vowels. George used shape features to detect brain tumour and to extract features for MRI scans (George *et al.*, 2015). Six features were considered which include Minor axis length, Major axis length, Solidity, Euler Number, Circularity and Area for feature extraction. Classification is made using tree diagrams and MLP algorithms. Achieved 95% performance rate by considering 176 MRI scan samples.

In this analysis CT images of lungs classification for early identification of COVID-19 is carried out. The shape analysis for classification is made by using invariant moment and shape features. By using the Neural Network classifier it is confined that the invariant moments are better than the shape features. Machine learning classifiers like

SVM, K-NN, logistic regression provide better performances. By using Neural Network classifier, excellent performance was observed since they are used for both regression and classification problem. NN classifier requires large amount of input data where it overwhelmed as a drawback in application. The classifier accuracy for invariant moments and shape features are 99% and 98%. The limitation for this study is the usage of limited dataset. By providing the large lung CT scan datasets of COVID patients better analysis will be performed.

Conclusions

Shape features are used as shape descriptors in comparison with invariant moments. Seven features are extracted using invariant moments. From shape features thirteen features of the function regionprops are considered. Extracted features are applied to Multi Level Perceptron classifiers. By comparing the performance metrics of invariant and shape features, it is concluded that the Invariant moments have better classification ability than the shape features in differentiating normal and COVID CT scan images. Performance of classifier using Invariant moments and shape features can be made together when same number of features are considered for the classification.

Acknowledgement

The authors would like to thank the following authors for providing the open access data of COVID-19 CT Lung and Infection Segmentation dataset and thanks for the Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS) for providing the resources to complete the study.

References

Al-Quteimat, Osama, M., and Amer Mustafa Amer. (2020). The Impact of the COVID-19 Pandemic on Cancer Patients. *American J. of Clinical Oncology*, 43(6):452-455

Asodekar, Bhagyashri H., Sonal, A. Gore, and A.D. Thakare. (2019). Brain Tumor Analysis Based on Shape Features of MRI Using Machine Learning. 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), 19-21 doi.org/10.1109/iccubea.47591.2019.9129512.

Dmitriev, E.A., A.A. Borodinov, A.I. Maksimov and S.A. Rychazhkov. (2019). Automatic Detection of Constructions Using Binary Image Segmentation Algorithms Information Technology and Nanotechnology., <https://doi.org/10.18287/1613-0073-2019-2391-264-268>.

Dogantekin, Esin., Mustafa Yilmaz., Akif Dogantekin., Engin Avci., and Abdulkadir Sengur. (2008). A Robust Technique Based on Invariant Moments-ANFIS for Recognition of Human Parasite Eggs in Microscopic Images. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2007.07.020>.

Faul, Franz., Edgar Erdfelder., Axel Buchner., and Albert-Georg Lang. (2009). Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behavior Research Methods*, 41(4):1,149-60.

George, D.N., Jehlol, H.B., and Oleiwi, A.S.A. (2015). Brain tumor detection using shape features and machine learning algorithms. *Int. J. of Advanced Research in Computer Science and Software Eng.*, 5(10):454-459.

Huang, Zhihu and Jinsong Leng. (2010). Analysis of Hu's Moment Invariants on Image Scaling and Rotation. 2010 2nd International Conference on Computer Engineering and Technology., <https://doi.org/10.1109/iccet.2010.5485542>.

Li, Meng. (2020). Chest CT Features and Their Role in COVID-19. *Radiology of Infectious Diseases*, 7(2):51-54

Ma Jun., Ge Cheng., Wang Yixin., An Xingle., Gao Jiantao., Yu Ziqi., and He Jian. (2020). COVID-19 CT Lung and Infection Segmentation Dataset (Version Version 1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3757476>.

Mingqiang, Yang., Kpalma Kidiyo., and Ronsin Joseph. (2008). A Survey of Shape Feature Extraction Techniques. *Pattern Recognition Techniques, Technology and Applications*, <https://doi.org/10.5772/6237>.

Öztürk, Şaban., Umut Özkaya., and Mucahid Barstuğan. (2021). "Classification of Coronavirus (COVID-19) from X-ray and CT Images Using Shrunk Features. *Int. J. of Imaging Systems and Technol.*, 31(1):5-15

Ramteke, R.J. (2010). Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition. *Int. J. of Computer Applications*, 1(18):975-8,887