CT Image Analysis Using Grayscale Statistics to Categorise Severity of Lung Abnormalities of COVID-19 Patients

Sara Ghashghaei

Shiraz University of Medical Sciences

David A. Wood

DWA Energy Limited

Erfan Sadatshojaei (✉ erfan.sadatshojaei@gmail.com)

Shiraz University

Mansoorreh Jalilpoor

Shiraz University of Medical Sciences

Research Article

Keywords: pulmonary image classification, Computed tomography grayscale, CT image attributes, confusion analysis; transparent abnormality scaling, computer-aided detection.

Posted Date: April 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-410348/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Grayscale image attributes from 456 images extracted from CT scan slices of 53 patients (49 with COVID-19 and 4 without) are used to establish a visual scale of severity of lung abnormalities (five classes: 0 to 4). The complex trends of these easy-to-derive image attributes can be used graphically to discern the visual scale of lung abnormalities in broad terms. With the aid of machine learning algorithms, the visual classes can be distinguished with close to 95% accuracy using combinations of selected grayscale attributes. Confusion matrices reveal that the best-performing machine learning models are able to distinguish more accurately between certain classes than visual inspection of CT images by experts. The adaboost, decision tree and random forest models confused on average less than 25 of the 456 CT-scan image extracts evaluated between the visual classes of lung abnormalities.

1. Introduction

The global COVID-19 virus pandemic has had devastating consequences. Accurately distinguishing and grading the severity of its pulmonary impacts for individual patients rapidly from image data remains an important objective. Computed tomography (CT) images offer a readily accessible, safe (low-radiation) and flexible modality for assessing pulmonary conditions and are now routinely used for diagnosing and assessing the severity of various lung diseases including lung cancer, bacterial and viral pneumonia and, most recently, COVID-19^{1,2}. Indeed, there are growing indications that radiological methods deliver greater sensitivity to COVID-19 diagnosis than a range of lab-based tests, including the standard diagnostic test of virus' nucleic acid by real-time reverse transcription polymerase chain reaction (rRT-PCR)^{3,4}. This makes thoracic CT scan slices the favoured image modality for COVID-19 detection and verification⁵. Much of the recent research on CT images relating to COVID-19 has been related to its diagnosis and observation during therapeutics rather than in attempting to classify different degrees of its severity⁶. The wider application of CT image information in terms of distinguishing degrees of severity of COVID-19 has been recognized and is being exploited⁷. López-Cabrera et al.⁸ considered some of the obstacles and limitations of using machine and deep learning methods for the automated classification of COVID-19 using chest CT scans. Farhat et al.⁹ also reported on some of the challenges of applying deep-learning techniques to pulmonary image data for COVID-19 diagnosis and prognosis.

Features indicative of COVID-19, typically present in thoracic CT images, are ground – glass opacity (GGO), crazy – paving and consolidation textures, air bronchograms, reverse halos, and peri-lobular patterns¹⁰. Duzgun et al.¹¹ evaluated radiological characteristics of a broad spectrum of lung pathologies, considering their resemblances with the joint signs of COVID-19 pneumonia. They identified that existing underlying lung abnormalities could interfere with the CT diagnosis of COVID-19. Deep feature classification process is proving useful in classifying the severity of lung abnormalities in patients with COVID-19 infections¹². Computer-aided detection (CAD) exploiting machine learning (ML) and deep learning (DL) algorithms has been successfully configured and applied to accurately detect and distinguish such features^{13–15}. In particular, DL methods adapting convolutional neural networks (CNN),

coupled with automatic feature extraction algorithms, have received much attention^{16,17}. Automated feature extraction is not a straightforward process as it requires effective and reliable image segmentation software^{14,18,19}.

There is now a strong research trend towards CT scan image analysis to apply deep learning coupled with automated feature extraction software⁹. Whereas such an approach is clearly effective for diagnosis and identifying severity of lung abnormalities, it is somewhat opaque to the underlying attributes that determine the features and patterns being distinguished in suites of thoracic CT scan slices now routinely collected for many COVID-19 patients. This study takes a somewhat different and original tack in an attempt to increase the transparency of the underlying image attributes that are contributing to the distinguishing features. Armed with this information much insight can be provided to the characteristics associated with different levels of severity of lung abnormalities.

This study focuses on the grayscale statistical analysis of human-selected image extracts of the pulmonary parenchyma portion of CT-scan slices. The image extracts are obtained by visual inspection and do not involve any automated feature extraction intervention. Much information can be rapidly gained from such grayscale image extracts using statistical and graphical analysis on a standalone basis. The statistical data is also suitable for further analysis with the aid of machine learning algorithms, on a supervised basis, to accurately predict degrees of severity of lung abnormalities associated with specific images.

2. Methods

2.1 CT Scan Images Compiled

In this study, images were gathered at the Namazi hospital (Shiraz, Iran; refer to Ethical Guidelines and Consent Statement) using a Philips 16-slice CT scanning machine. Image sets with slice thicknesses of 0.625mm were generated without using contrast. Unenhanced chest CT scans are widely employed as a good modality for fast diagnosis and evaluation of viral diseases such as coronavirus, complementing the laboratory rRT-PCR. Suites of CT image scans were compiled from the 49 COVID-positive patients and 4 COVID-negative patients.

CT scanning devices can have either single or multi-slice capabilities. Single-slice CT scanners generate a single image for each spin of the patient gantry. Multi-slice machines are now widely used in many hospitals because they provide more rapid and comprehensive imaging. Multi-slice CT scanners range from four to sixty-four-slice configurations²⁰. 16-slice CT scanners facilitate rapid full-organ coverage generating high-resolution images and are well-suited for detailed lung investigation.

2.2 CT Scan Image Analysis Procedure

2.2.1 Visual CT scan image assessment

The suite of CT image slices for each individual were visually assessed by a clinician and each image was assigned to one of five classes using a score on the following 0 to 4 scale:

0 = Patient has tested negative for Covid-19 and has no obvious signs of other lung diseases

1 = Patient has tested positive for Covid-19 but has no obvious signs lung abnormalities

2 = Patient has tested positive for Covid-19 and has only minor signs of lung abnormalities

3 = Patient has tested positive for Covid-19 and has substantial signs of lung abnormalities

4 = Patient has tested positive for Covid-19 and has severe signs of lung abnormalities

These visual scores (VS) then become the prediction objective of the grayscale image analysis. Appendix 1 (see Supplementary File) shows example extract images for each of these VS classes.

2.2.2 Statistical analysis of extract grayscale images

From the suite of CT scan images for each COVID-positive patient, four to five images were identified as representative, and selected for grayscale analysis. From each of those selected CT images, two rectangular grayscale image samples were extracted, one from the left lung and one from the right lung. From the 49 COVID-positive patients this resulted in 392 individual extract images being compiled. The same approach was taken with the COVID-negative patients but more of the CT images were sampled from each of those patients. A total of 64 extract images were compiled from the four COVID-negative patients. The dataset for grayscale analysis therefore includes 456 individual rectangular extract images.

The extract images varied in size from about 2000 to 80000 pixels, averaging about 25000 pixels. That size of the extracted image was determined by the nature of the original CT scan. The objective for each extract image is that it should include only the parenchyma of the lung, avoiding the pleura, the diaphragm and the mediastinum.

A pixel in a grayscale image is associated with a single pixel value indicative of its brightness. That pixel value is typically expressed as an 8-bit integer that allows a value range from 0 to 255. By convention 0 represents black, 255 represents white, and all the intermediate values represent shades of grey. This makes it possible to conduct statistical analysis of the pixel grayscale values of all the pixels in each extract images. In this study that analysis was conducted using the open-source computer vision library²¹ for image analysis run in Python code. For each extract image the following statistical information was extracted:

1. Number of pixels
2. Average grayscale pixel value
3. Number of pixels with the average value
4. Percentage of pixels with the average grayscale value

5. Grayscale variance
6. Grayscale variance / grayscale average
7. Grayscale standard deviation
8. Grayscale standard error
9. Minimum grayscale value
10. Grayscale P10 value
11. Grayscale P50 value
12. Grayscale P90 value
13. Maximum grayscale value

Two of these statistics (#3 and #8) are substantially influenced by the number of pixels in each image, so are not suitable for comparative use between images. However, the standard error is useful for providing an indication of the uncertainty in the average grayscale values. It is encouraging that standard error is less than 0.7 of a grayscale unit for all images evaluated. Statistic #4 is independent of image size so it is a more appropriate measure to use than statistic #3 for comparative purposes.

2.2.3 Analysis with machine learning algorithms

Values for statistics #2 to #13 constitute the independent variables. The visual assessment scores (0 to 4) of lung abnormalities observed in the CT images (termed “visual score”; VS) represent the dependent variable to be predicted by machine learning (ML) using a suite of classification algorithms. Therefore, each data record in the dataset involves twelve independent variables (the grayscale statistics for a specific extract image) and one dependent variable, the visual score determined by the clinician. The models are run in python code with the objective function set to minimize the root mean squared error (RMSE) of VS actual (VS_{act}) versus VS predicted (VS_{pred}).

The 456 data records are evaluated with a suite of ten established ML classification methods, which are:

Adaboost (ADA: a boosted decision tree model)

Decision Tree (DT)

Extreme Learning Machine (ELM)

Gaussian Process Classification (GPC; based on Laplace approximation)

K Nearest Neighbour (KNN)

Multi-layer Perceptron (MLP)

Naïve Bayes Classifier (NBC)

Random Forest (RF)

Quadratic Discriminant Analysis (QDA)

Support Vector Machine (SVM)

The VS prediction accuracy achieved by each of these methods is recorded and compared using a set of established prediction accuracy statistics. Application workflows and methodologies associated with the ten ML methods evaluated are not detailed here as they are well established and discussed in detail elsewhere: ADA^{22,23}; DT^{24–26}; ELM^{27–29}; GPC^{30,31}; KNN^{13,32}; MLP³³; NBC^{13,34}; RF³⁵; QDA^{36,37}; and, SVM^{38,39}. In this study, these algorithms are run in Python code and the control parameter values applied to them are listed in Table 1.

Table 1. ML algorithm execution method and control variable values applied to predict VS using CT scan extract grayscale image dataset. All models were executed in Python.

Machine Learning Algorithm	Control Parameter Values Applied
Adaboost (ADA)	Number of estimators=2000; learning rate =2; base estimator is DT with depth =1000; splitter =best
Decision Tree (DT)	Maximum depth =10000; splitter =best; criterion =entropy
Extreme Learning Machine (ELM)	Hidden_units =5000; activation function = leaky_relu; random type = normal
Gaussian Process Classification (GPC)	Kernel = rbf; rbf hyperparameters optimized as part of training
K Nearest Neighbour (KNN)	Neighbours = 10; weighted by Euclidian distance
Multi-layer Perceptron (MLP)	3 hidden layers with 500, 250 and 250 neurons; logistic activation function; adam solver; alpha=0.0000001; maximum iterations 1000
Naïve Bayes Classifier (NBC)	Gaussian processor; variable
Random Forest (RF)	Number of estimators=1000; Maximum depth =1000
Quadratic Discriminant Analysis (QDA)	No control parameters influence predictions
Support vector machine (SVM)	Kernel = rbf; C=500000; gamma=0.00001

2.2.4 Statistical measures of prediction accuracy assessed

The VS prediction accuracy is assessed in terms of nine widely-used measures of statistical accuracy (Fig. 1).

These measures of prediction accuracy complement each other in the information they provide about the VS predictions versus the actual values. RMSE is of particular interest as this is the metric used as the objective function for the ML algorithms.

3. Results

3.1 Statistical analysis of grayscale image extracts

A summary of the grayscale image statistics is provided in Table 2 considering the dataset as a whole and distinguishing the images for patients testing positive for COVID-19 and those testing negative. It is clear that there is a substantial variation in all of the grayscale statistics, which is encouraging for the possibility of exploiting them collectively to distinguish the severity of lung abnormalities in the patients. Even more encouraging are the substantial differences between the value distributions of several of these statistics for COVID positive and negative patients. Of particular note, are the higher average values of

statistics #2 and #5 to #12 and the lower average values of statistics #3 and #4 for the COVID-19 positive patients compared with the COVID-19 negative patients (Table 2).

As well as the differences observed between the patient groups, there are also some notable correlations between the grayscale statistics and the VS indicator (Table 3). Two correlation coefficients are calculated: the Pearson correlation coefficient (R) that assumes the variables are normally distributed (i.e., parametric) and the Spearman correlation coefficient (ρ) which is non-parametric and is calculated using ranked data. The ninety percentile of the grayscale distribution (P90) shows the highest R and ρ values with VS of 0.86 and 0.87, respectively. The average grayscale values also show high positive correlations with VS ($R = 0.79$ and $\rho = 0.81$), as does the fifty percentile (P50; $R = 0.73$ and $\rho = 0.78$) and variance ($R = 0.67$ and $\rho = 0.72$). On the other hand, the percentage of pixels with grayscale average value shows a strong negative correlation with VS ($R = -0.73$ and $\rho = -0.77$). In broad terms, these relationships show greater grayscale variability in the COVID-19 sufferers.

Table 2. Statistical summary of grayscale statistical value distributions for 456 image extracts from 53 patients (49 testing positive for COVID-19 and 4 testing negative for COVID-19. VS refers to the visual score (scale 0 to 4) assigned to each image by a clinician in terms of the severity of lung abnormalities based on inspection of the suite of CT scans available.

	Pixels	Average Grayscale	Average Grayscale #Pixels	Average Grayscale %Pixels	Variance Grayscale	Variance / Average Grayscale	Standard Deviation Grayscale	Standard Error Grayscale	Minimum Grayscale	P10 Grayscale	P50 Grayscale	P90 Grayscale	Maximum Grayscale	VS Objective Function
Statistic Number	1	2	3	4	5	6	7	8	9	10	11	12	13	
Full Set : 456 Data Records														
Minimum	1830	64.6	7	0.073%	126.9	1.93	11.3	0.057	0	31	54	75	188	0
Maximum	80472	214.4	2171	5.572%	3671.0	32.28	60.6	0.860	78	186	222	235	255	4
Range	78642	149.8	2164	5.499%	3544.1	30.34	49.3	0.803	78	155	168	160	67	4
Average	24547	113.5	330.8	1.257%	1413.5	12.45	36.4	0.291	44.4	78.1	104.8	163.8	250.1	2.2
Standard Deviation	16540	30.8	350.8	0.835%	687.7	5.49	9.3	0.152	16.5	22.5	34.1	43.7	10.2	1.2
Covid-19 Negative : 64 DataRecords														
Minimum	1830	65.7	19	1.038%	126.9	1.93	11.3	0.057	5	44	61	75	188	0
Maximum	62592	111.9	2171	5.572%	1676.1	17.27	40.9	0.673	55	74	101	168	255	0
Range	60762	46.2	2152	4.534%	1549.2	15.34	29.7	0.616	50	30	40	93	67	0
Average	27912	79.9	715.7	2.387%	735.2	8.99	26.4	0.207	35.6	58.3	73.7	106.2	237.9	0
Standard Deviation	16972	12.0	537.3	0.923%	336.0	3.43	6.3	0.132	13.7	7.5	11.8	20.9	16.9	0
Covid-19 Positive : 392 DataRecords														
Minimum	2600	64.6	7	0.073%	189.4	2.35	13.8	0.071	0	31	54	82	214	1
Maximum	80472	214.4	1638	4.340%	3671.0	32.28	60.6	0.860	78	186	222	235	255	4
Range	77872	149.8	1631	4.267%	3481.6	29.93	46.8	0.789	78	155	168	153	41	3
Average	23998	119.0	267.9	1.073%	1524.3	13.02	38.1	0.304	45.8	81.3	109.9	173.3	252.1	2.5
Standard Deviation	16425	29.5	261.6	0.657%	666.7	5.56	8.7	0.151	16.5	22.5	33.8	38.9	6.9	0.9

Table 3. Pearson correlation coefficients for the grayscale statistics of the 456 extract images. The last column displays the visual score Spearman rank correlation coefficients for comparison.

Pearson Correlation Coefficient (R)	Number of Pixels	Average Grayscale	Average Grayscale # Pixels	Average Grayscale % Pixels	Variance Grayscale	Variance / Average	Standard Deviation Grayscale	Standard Error Grayscale	Minimum Grayscale	P10 Grayscale	P50 Grayscale	P90 Grayscale	Maximum Grayscale	Visual Score (VS)	VS Spearman Correlation Coefficient
Number of Pixels	1	-0.2095	0.6564	0.1532	-0.2134	-0.1258	-0.2079	-0.7488	0.0202	-0.1116	-0.1981	-0.2398	-0.1759	-0.1804	-0.1992
Average Grayscale		1	-0.4868	-0.5611	0.4818	-0.0022	0.5030	0.3944	0.4892	0.8838	0.9792	0.8983	0.4460	0.7886	0.8129
Average Grayscale # Pixels			1	0.7710	-0.6109	-0.4654	-0.6690	-0.7062	-0.0177	-0.2505	-0.4103	-0.6569	-0.5170	-0.5765	-0.5739
Average Grayscale %Pixels				1	-0.8083	-0.6463	-0.8750	-0.5283	-0.0349	-0.2637	-0.4545	-0.8008	-0.5594	-0.7306	-0.7669
Variance Grayscale					1	0.8399	0.9876	0.6500	-0.1270	0.0852	0.3805	0.7660	0.4330	0.6681	0.7152
Variance / Average						1	0.8358	0.5224	-0.4949	-0.3716	-0.1065	0.3774	0.3065	0.3540	0.3984
St. Dev. Grayscale							1	0.6472	-0.1134	0.1137	0.3972	0.7911	0.4926	0.6936	0.7152
Standard Error Grayscale								1	-0.0897	0.1256	0.3414	0.5496	0.3832	0.4432	0.5078
Minimum Grayscale									1	0.7108	0.4520	0.3213	0.1895	0.1807	0.1896
P10 Grayscale										1	0.8697	0.6697	0.3414	0.5525	0.5577
P50 Grayscale											1	0.8112	0.3807	0.7339	0.7676
P90 Grayscale												1	0.5179	0.8594	0.8700
Maximum Grayscale													1	0.4660	0.3792
Visual Score (VS)														1	

3.2 Graphical analysis of grayscale statistics

Graphical displays better illustrate the continuity of the grayscale distributions than the tabular statistical analysis and highlight some key relationships between these variables. Figure 2 shows the relative spread of values relating to grayscale average, grayscale variance and percentage of pixels at the grayscale average. The scaling factors are applied to optimize the spread of the data points in this triangular display.

The three statistics displayed in Fig. 2 can successfully distinguish the severity of COVID-19-related lung abnormalities with a continuous gradation from the lower left of the triangle. That position indicates a high percentage of pixels at the grayscale average which is of dark grey shades with low variance. As the data points move towards the top right the increasing variance and average grayscale indicate more lighter grey shades are present as the lung abnormalities become more severe (higher VS values indicated by darker symbols). Finally, for the grayscale extract images with the most severe lung abnormalities the data points in Fig. 2 move towards the lower right of the triangle as variance decreases and light grey shades dominate the images.

A three-dimensional graphic of grayscale P90 versus percentage of pixels at the average grayscale value versus grayscale variance provides another graphic that meaningfully distinguishes the severity of lung abnormalities related to COVID-19 from CT scan images (Fig. 3).

The graphical and statistical relationships established indicate that the grayscale statistics can be used effectively to identify the severity of lung abnormalities in relation to COVID-19 infection. Indeed, the contrast between these statistical variables suggest that relatively simply formulaic relationships between some of them could be used to provide a useful automated scale of severity of lung abnormalities with which to compare the VS score visually assigned by a clinician. This goes beyond the objective of the current study. However, the approach of exploiting grayscale images extracted from CT scans clearly has promise for categorizing lung abnormalities associated with diseases such as pneumonia and lung cancer as well as COVID-19 from lung CT scans. Future research is planned to further explore these possibilities.

3.3 Feature selection for machine learning analysis

The grayscale statistical data set (456 data records with 13 input variables; Table 2) can be used with the commonly used ML algorithms to predict the VS value for each image extract to a high degree of accuracy. To demonstrate this, two cases of ML models are executed with VS as the objective function: A) using eleven of the statistical variables (all excluding number of pixels and Variance/Average) available; and B) using just nine variables (excluding number of pixels, standard error, minimum grayscale and number of pixels at the average grayscale value). The feature selection for the variables to use is based upon the information provided in Table 3 and Figs. 2 and 3. The most influential variables, in terms of their correlation coefficients with VS, and each other, are included as input variables. On the other hand, the variables with the lowest R and p values with VS, and those that are influenced by the number of pixels in the extract images, are excluded from the 9-variable case.

3.4 VS predictions from machine learning methods

Tables 4 and 5 present the prediction accuracy statistics for the ten ML algorithms in terms of their ability to correctly assign the VS value to each of the 456 data records, for the 11-variable and 9-variable models, respectively. Each algorithm's prediction performance is ranked in terms of RMSE and the number of prediction errors (right-hand columns of Tables 4 and 5). Most algorithms yield better VS prediction accuracy in terms of RMSE for the 11-variable models. However, the adaboost model results in lower RMSE (and number of errors) for the 9-variable model making it the best performing algorithm for that data variable combination. Random forest generates the best performance for the 11-variable models in terms of RMSE, although the decision tree model results in the lowest number of errors (23 out of 456; 95% of its predictions being correct). The KNN model achieves similar performance for the 11-variable and 9-variable models with lower RMSE in the former but lower errors in the latter.

Table 4. Prediction accuracy results for 11-variable grayscale statistics machine learning models (Case A) applied to 456 grayscale extract images from CT scans of COVID-19 positive (VS = 1 to 4) and COVID-19-negative (VS = 0) patients.

Statistical Measures of Prediction Accuracy for 11-Variable Machine Learning Models										
Case A 12 Input Variables	RMSE (VS)	MAE (VS)	APD (%)	AAPD (%)	Sdev (VS)	R ²	Number of Errors	% Correct	Rank by RMSE	Rank by Errors
ADA	0.2864	0.062	-0.757	2.510	0.287	0.9404	26	94.30%	5	4
DT	0.2757	0.056	-1.913	2.887	0.275	0.9448	23	94.96%	4	1
ELM	0.2608	0.056	-0.730	2.270	0.261	0.9508	25	94.52%	2	3
GPC	0.4290	0.168	-1.260	5.807	0.429	0.8670	80	82.46%	8	8
KNN	0.2683	0.060	-0.507	2.393	0.269	0.9477	27	94.08%	3	5
MLP	0.3715	0.118	-1.483	4.650	0.372	0.9007	54	88.16%	7	7
NBC	0.6465	0.318	-3.823	14.310	0.647	0.7225	134	70.61%	10	10
QDA	0.4960	0.190	-3.220	8.247	0.496	0.8230	84	81.58%	9	9
RF	0.2569	0.054	-0.290	2.277	0.257	0.9520	24	94.74%	1	2
SVM	0.3286	0.08	-0.137	3.230	0.328	0.9227	34	92.54%	6	6

Table 5. Prediction accuracy results for 9-variable grayscale statistics machine learning models (Case B) applied to 456 grayscale extract images from CT scans of COVID-19 positive (VS = 1 to 4) and COVID-19-negative (VS = 0) patients.

Statistical Measures of Prediction Accuracy for 9-Variable Machine Learning Models										
Case B 9 Input Variables	RMSE (VS)	MAE (VS)	APD (%)	AAPD (%)	Sdev (VS)	R ²	Number of Errors	% Correct	Rank by RMSE	Rank by Errors
ADA	0.2324	0.05	0.053	1.653	0.232	0.9606	24	94.74%	1	4
DT	0.3376	0.066	-1.467	3.267	0.338	0.9174	24	94.74%	6	4
ELM	0.3376	0.062	-1.52	3.313	0.338	0.9184	24	94.74%	6	4
GPC	0.5329	0.212	-4.470	10.497	0.533	0.8006	91	80.04%	7	7
KNN	0.2720	0.058	-0.633	2.600	0.272	0.9462	25	94.52%	2	5
MLP	0.5727	0.240	-4.610	11.190	0.572	0.7808	100	78.07%	8	8
NBC	0.6588	0.330	-3.710	15.190	0.659	0.7123	139	69.52%	10	10
QDA	0.5848	0.254	-0.717	9.883	0.582	0.7686	109	76.10%	9	9
RF	0.2793	0.058	-0.507	2.660	0.279	0.9435	24	94.74%	3	4
SVM	0.3225	0.08	-0.217	3.250	0.322	0.9255	35	92.32%	4	6

The ADA, DT, ELM, KNN and RF models substantially outperform the other models for cases A and B. The GPC, NBC and QDA models substantially underperform the other models for cases A and B with the NBC resulting in the lowest prediction accuracy for both cases evaluated. The MLP and SVM models show intermediate prediction accuracy with the SVM model achieving similar results for both cases (34 and 35 prediction errors, respectively). On the other hand, the MLP models performance deteriorates significantly from Case A (54 prediction errors) to case B (100 prediction errors). These results suggest that of the ML models evaluated ADA, DT, ELM, KNN and RF are best suited to the VS prediction task for the data set evaluated. Future studies are planned to evaluate the potential of deep learning methods to improve upon the results achieve by these deep learning methods.

4. Discussion

The results presented are considered to be very encouraging in terms of the use of statistical analysis of grayscale image extracts from suites of CT scans to identify the severity of lung abnormalities in COVID-19 suffers and potentially those with other lung diseases. The broad continuous nature of the distributions for several grayscale statistics (Tables 2 and 3) can be usefully exploited to grade lung abnormalities on a continuous basis. This can be achieved graphically (Figs. 2 and 3) without recourse to machine learning. Alternatively, by exploiting ML methods it can accurately (up to about 95% accuracy) predict visual scales of lung damage assessed by a clinician. These results justify further research and application of the method.

How error is assessed in the ML methods is worthy of more detailed consideration. The ML methods are driven to minimize RMSE as their objective function. This is considered an appropriate approach as RMSE provides a continuous and highly sensitive error scale for this purpose. However, the overall objective of the prediction models is to minimise the number of prediction errors. It is therefore important to compare the performance of these two error measures, and the other measures shown in Tables 4 and 5 to more fully understand the accuracy of the measures. It is possible for the RMSE value to go up slightly but the number of prediction errors to go down slightly (e.g., compare the performance of the KNN algorithm for Cases A and B). It is worth exploring why this is possible and its implications.

Confusion matrices are useful ways to explore in more detail how the prediction errors are distributed across the VS classes (0 to 4). Figure 4 plots a confusion matrix for the two best performing models (RF for Case A and ADA for Case B).

The comparison of these two high-performing ML models is interesting because they both achieve the same number of prediction errors overall, that is 24, but the ADA model for Case B results in a lower RMSE value (0.2324) compared with that of the RF Case A model (RMSE = 0.2569). Close inspection of the confusion matrices shows that both models achieve similar performance for VS class 0 (COVID-negative patients) with two errors. However, the RF model confuses its two errors for class 0 with VS class 2, whereas the ADA model confuses its two errors with class 1. In RMSE terms, the RF model's errors for this class will contribute more to the RMSE than the ADA errors. The ADA model performs better in its predictions for VS classes 1 and 2 than the RF model. On the other hand, the RF model performs better in its predictions for VS classes 3 and 4 than the ADA model. What is relevant though, is that only one of the 24 errors generated by the ADA model is more than one class away from the correct VS class (i.e., for VS class 4 one erroneous prediction falls into class 2). On the other hand, three of the errors generated by the RF model are more than one class away from the correct VS class (i.e., two for VS class 0 and one for VS class 2). The RMSE magnitude is therefore affected by not only the number of prediction errors but also their magnitude. From an RMSE perspective, it is better for errors to fall into neighbouring classes rather than into more distant cases.

As well as, providing insight to the location of the prediction errors, the confusion matrices highlight the classes of the dependent variable for which the model works well and less well. Consequently, Fig. 4 suggests that the RF could potentially do better with its predictions at the middle / lower end of the VS scale (VS 0 to 2), whereas ADA improved could potentially do better with its predictions at the middle/ upper end of the VS scale (VS 2 to 4). Work is underway to consider how such improvements could be achieved.

What should be taken into account is that the VS classes are made by human visual observation (a clinician's considered view). The boundaries between the classes are therefore, to an extent, associated with a degree of subjectivity. As is clear from Figs. 2 and 3 the severity of lung abnormalities likely falls into a continuous spectrum from none to extreme and any class boundaries placed visually in that spectrum by human interpretation will to an extent be arbitrary. In that context, only 24 VS misclassifications out of 456 images (almost 95% correct) by the better ML models is surprisingly good. It is unlikely, considering the broad nature of the VS class definitions applied in this study that any expert system would be able to get close to zero VS prediction errors.

Another observation worthy of discussion is that the better performing ML models confuse very few of the images between classes VS = 0 (COVID-19-negative) and VS = 1 (COVID-19-positive, but with no visible signs on the CT images). By definition the clinician is unable to distinguish classes 0 and 1 by visual inspection. Yet with > 95% accuracy (Fig. 4B) the ADA 9-variable model is able to distinguish these two classes with only three misclassifications between the groups from 114 images analysed. Clearly, the

grayscale statistics are able to distinguish distinct attributes of these classes that are not visible to even to the trained observer. This is considered to be further justification for applying and further evaluating this the application of grayscale statistical analysis of CT scan images.

5. Conclusions

Consideration of the grayscale image attributes of the pulmonary parenchyma portion of CT-scan slices reveals valuable information with which to categorize the severity of lung abnormalities associated with COVID-19. Such information usefully complements their visual analysis by medical experts and supports a classification system (0 = COVID negative; 1 = COVID positive with no lung abnormalities; 2 to 4 represent varying grades of severity of lung abnormalities) that can accurately be discerned with the aid of machine learning. The analysis compiles 456 extract images from CT scans from 49 patients with COVID-19 (confirmed by rRT-PCR testing) and 4 individuals testing negative for the disease. Thirteen grayscale statistical features were measured for each extract image. Several of these features follow systematic but complex continuous trends that can be graphically related to the 0 to 4 visual assessment trend of lung abnormalities. Ten machine learning methods were evaluated to establish their ability to accurately identify the visual classification of lung abnormalities applying supervised learning to selected grayscale features. The best of these machine learning models was able to achieve almost 95% prediction accuracy using nine to eleven grayscale variables. Adaboost, Decision tree and random forest models averaged 23 to 25 prediction errors from the 456 image dataset with respect to the lung abnormality severity classes. Confusion matrices reveal that the best performing models make very few errors in distinguishing between the visual classes. Of particular interest is their ability to distinguish between visual class zero (COVID-19 free class) and class one (COVID-19 positive, with no visually discernible abnormalities), a distinction that is not possible by visual inspection alone.

Declarations

The authors declare that they have no conflicts of interest and have received no external funding associated with this research.

Conflict of Interest Statement

The authors declare that they have no conflict of interests regarding the information presented in this study.

Ethical Guidelines and Consent Statement

All analysis of CT scans of human patients used in this study were performed in accordance with relevant guidelines and regulations of the Namazi Hospital (Shiraz, Iran). In particular, the anonymity of individuals and confidentiality of biographical information has been maintained at all times. Furthermore, the informed consent for CT scan analysis was obtained from all CT scan subjects from which images were derived for this study. All analytical protocols were approved by the Namazi Hospital.

Funding Statement

This study has received no funding from third parties.

References

- [1] Bai, H. Hsieh, B. Xiong, Z. Halsey, C. Choi, J.W. et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology* **296** (2), E46–E54 (2020). DOI: 10.1148/radiol.2020200823.
- [2] Hare, S.S. Tavaré, A.N. Dattani, V. Musaddaq, B. Beal, I. et al.. Validation of the British Society of Thoracic Imaging guidelines for COVID-19 chest radiograph reporting. *Clinical Radiology* **75** (9), 710.e9-710.e14 (2020). <https://doi.org/10.1016/j.crad.2020.06.005>
- [3] Ai, T. Yang, Z. Hou, H. Zhan, C. Chen, C. et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* **296** (2), E32-E40 (2020). <https://doi.org/10.1148/radiol.2020200642>
- [4] Fang, Y. Zhang, H. Xie, J. Lin, M. Ying, L. et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* **296** (2), E115–E117 (2020). <https://doi.org/10.1148/radiol.2020200432>
- [5] Xie, X. Zhong Z. Zhao, W. Zheng, C. Wang, F. Liu, J. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology* **296**, 200343 (2020). <https://doi.org/10.1148/radiol.2020200343>
- [6] Ng, M-Y. Lee, E.Y. Yang, J. Yang, F. Li, X. et al. Imaging profile of the COVID-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging* **2** (1), e200034 (2020). <https://doi.org/10.1148/ryct.2020200034>
- [7] Li, K. Wu, J. Wu, F. Guo, D. Chen, L. et al. The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. *Invest Radiology* **55** (6), 327–31 (2020). DOI: 10.1097/RLI.0000000000000672
- [8] López-Cabrera, J.D. Orozco-Morales, R. Portal-Díaz, J.A. Lovelle-Enríquez, O. & Pérez-Díaz, M. Current limitations to identify COVID-19 using artificial intelligence with chest X-ray imaging, *Health and Technology* **11**(2), 411-424 (2021).
- [9] Farhat, H. Sakr, G.E. & Kilany, R. Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19. *Machine Vision and Applications* 31-53 (2020). <https://doi.org/10.1007/s00138-020-01101-5>
- [10] Chung, M. Bernheim, A. Mei, X. Zhang, N. Huang, M. Zeng, X. et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology* **295**, 202–7 (2020). <https://doi.org/10.1148/radiol.2020200230>

- [11] Duzgun, S.A. Durhan, G. Demirkazik, F.B. Akpınar, M.G. & Ariyurek, O.M. COVID-19 pneumonia: the great radiological mimicker. *Insights Imaging* **11**, 118 (2020). <https://doi.org/10.1186/s13244-020-00933-z>
- [12] Yu, Z. Li, X. Sun, H. Wang, J. Zhao, T. et al. Rapid identification of COVID-19 severity in CT scans through classification of deep features. *BioMed Eng OnLine* **19**, 63 (2020). <https://doi.org/10.1186/s12938-020-00807-x>
- [13] Hussain, L. Nguyen, T. Li, H. Abbasi, A.A. Lone, K.J. Zhao, Z. et al. Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection. *BioMed Eng OnLine* **19**, 88 (2020). <https://doi.org/10.1186/s12938-020-00831-x>
- [14] Shi, F. Wang, J. Shi, J. Wu, Z. Wang, Q. et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering* **14**, 4-15 (2021). DOI: 10.1109/RBME.2020.2987975.
- [15] Dong, D. Tang, Z. Wang, S. Hui, H. Gong, L. et al. The role of imaging in the detection and management of COVID-19: A Review. *IEEE Rev Biomed Eng.* **14**, 16-29 (2021). DOI: 10.1109/RBME.2020.2990959.
- [16] Li, L. Qin, L. Xu, Z. Yin, Y. Wang, X. et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology* **296** (2), E65–E71 (2020). <https://doi.org/10.1148/radiol.2020200905>
- [17] Song, Y. Zheng, S. Li, L. Zhang, X. Zhang, X. et al. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Trans Comput Biol Bioinform* (2021) DOI: 10.1109/TCBB.2021.3065361.
- [18] Haque, I.R.I. & Neubert, J. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked* **18**, 100297 (2020). <https://doi.org/10.1016/j.imu.2020.100297>
- [19] Zhou, T. Canu, S. & Ruan, S. 2021. An automatic COVID-19 CT segmentation based on U-Net with integrated spatial and channel attention mechanism. *Int J Imaging Syst Technol.* **31**, 16–27 (2021). DOI: 10.1002/ima.22527
- [20] Amber Diagnostics. Different types of CT machines [Accessed 30 March, 2021] <https://www.amberusa.com/blog/types-of-ct-machines/>.
- [21] OpenCV. Open source computer vision library. [Accessed 30 March 2021] <https://docs.opencv.org/master/d1/dfb/intro.html>
- [22] Barstugan, M. Rahime Ceylan, R. Sivri, M. & Erdogan, H. Automatic liver segmentation in abdomen CT images using SLIC and adaboost algorithms. *Proceedings of the 2018 8th International Conference*

on *Bioscience, Biochemistry and Bioinformatics* 129–133 (2018).

<https://doi.org/10.1145/3180382.3180383>

[23] Kuwahara, M. Kido, S. & Shouno, H. Classification of patterns for diffuse lung diseases in thoracic CT images by adaboost algorithm. *Proceedings Volume 7260, Medical Imaging 2009: Computer-Aided Diagnosis*; 726037 (2009). <https://doi.org/10.1117/12.811497>

[24] Agarwal, C. & Sharma, A. Image understanding using decision tree based machine learning. *Proceedings of the 5th international Conference on Information Technology & Multimedia* (2011) DOI: 10.1109/ICIMU.2011.6122757

[25] Rajendran, P. & Madheswaran, M. Hybrid medical image classification using association rule mining with decision tree algorithm. *Journal of Computing*, **2** (1), 127-136 (2010).

[26] Yoo, S.H. Geng, H. Chiu, T.L. Yu, S.K. Cho, D.C. et al. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Front. Med.* **7**, 427 (2020). DOI: 10.3389/fmed.2020.00427

[27] Albadr, M.A.A. Tiun, S. Ayob, M. AL-Dhief, F.T. Omar, K. & Hamzah, F.A. Optimised genetic algorithm-extreme learning machine approach for automatic COVID-19 detection. *PLoS ONE* **15** (12), e0242899 (2020). <https://doi.org/10.1371/journal.pone.0242899>

[28] Dogantakin, A. Ozyurt, F. Avci, E. & Koc, M. A novel approach for liver image classification: PH-C-ELM. *Measurement* **137**, 332–338 (2019). <https://doi.org/10.1016/j.measurement.2019.01.060>

[29] Huang, G-B. Zhou, H. Ding X. & Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **42**, 513–529 (2011). <https://doi.org/10.1109/TSMCB.2011.2168604> PMID: 21984515

[30] Zhu, F. Carpenter, T. Gonzalez, D.R. Atkinson, M. Wardlaw, J. Computed tomography perfusion imaging denoising using gaussian process regression. *Phys Med Biol.* **57** (12), N183-98 (2012). DOI: 10.1088/0031-9155/57/12/N183.

[31] Hussein, S. Gillies, R. Cao, K. Song, Q. & Bagci, U. TumorNet: Lung nodule characterization using multi-view convolutional neural network with gaussian process. *IEEE 14th International Symposium on Biomedical Imaging* (2017). DOI: 10.1109/ISBI.2017.7950686

[32] Ramteke, R.J. & Khachane Monali, Y. Automatic medical image classification and abnormality detection using K-nearest neighbour. *International Journal of Advanced Computer Research.* **2** (4,6), 190-196 (2012).

[33] Kuruvilla. J. & Gunavathi, K. Lung cancer classification using neural networks for CT images. *Computer Methods and Programs in Biomedicine* **113** (1), 202-209 (2014). <https://doi.org/10.1016/j.cmpb.2013.10.011>

- [34] Park, D-C. Image classification using naïve Bayes classifier. *International Journal of Computer Science and Electronics Engineering* **4** (3),135-139 (2016).
- [35] Nedjar, I. El Habib Daho, M. & Settouti, N. Random forest based classification of medical x-ray images using a genetic algorithm for feature selection. *Journal of Mechanics in Medicine and Biology* **15** (2):1540025 (2015).DOI: 10.1142/S0219519415400254
- [36] Suzuki, K. Machine learning in computer-aided diagnosis of the thorax and colon in CT: a survey. *IEICE Trans Inf Syst.* **E96** (D4),:772-783 (2013).DOI: 10.1587/transinf.e96.d.772
- [37] Huidrom, R., Chanu, Y.J., Singh, K.M. Pulmonary nodule detection on computed tomography using neuro-evolutionary scheme. *SIVIP* **13**, 53–60 (2019). <https://doi.org/10.1007/s11760-018-1327-4>
- [38] Cyran, K.A. Kawulok, J. Kawulok, M. Stawarz, M. Michalak, M. et al. Support vector machines in biomedical and biometrical applications. In: S. Ramanna, L.C. Jain, & R.J. Howlett (Eds.): Emerging Paradigms in ML. Smart Innovation, Systems and Technologies, vol 13. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28699-5_15
- [39] Zhang, Z. & Sejdic, E. Radiological images and machine learning: Trends, perspectives, and prospects. *Computers in Biology and Medicine* **108**, 354-370 (2019).
<https://doi.org/10.1016/j.compbiomed.2019.02.017>

Figures

Statistical Measures of Prediction Accuracy Used to Assess TOB Performance

Mean Squared Error (MSE):	$MSE = \frac{1}{n} \sum_{i=1}^n ((X_i) - (Y_i))^2$
Root Mean Squared Error (RMSE):	$RMSE = \sqrt{MSE}$
Mean Absolute Error (MAE):	$MAE = \frac{1}{n} \sum_{i=1}^n X_i - Y_i $
Percentage Deviation (PD):	$PD_i = \frac{X_i - Y_i}{X_i} \times 100$
Average Percentage Deviation (APD):	$APD = \frac{\sum_{i=1}^n PD_i}{n}$
Absolute Average Percentage Deviation (AAPD):	$AAPD = \frac{\sum_{i=1}^n PD_i }{n}$
Standard Deviation (SD):	$SD = \sqrt{\frac{\sum_{i=1}^n (D_i - Dmean)^2}{n-1}}$
Correlation Coefficient (R):	$R = \frac{\sum_{i=1}^n (X_i - Xmean)(Y_i - Ymean)}{\sqrt{\sum_{i=1}^n (X_i - Xmean)^2 \sum_{i=1}^n (Y_i - Ymean)^2}}$
Coefficient of Determination (R ²):	R is expressed on a scale of -1 to +1 R ² is expressed on a scale of 0 to 1

Notes:

X_i is the measured value and Y_i is the predicted value for data record i

n is the number of data records in the set or subset being evaluated

D_i is $(X_i - Y_i)$ for data record i

$$Dmean = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)$$

Figure 1

Nine Statistical measures of prediction accuracy, their calculation formulas and reference abbreviations.

Figure 2

Triangular display of key grayscale statistics for distinguishing severity of lung abnormalities. Grayscale average and variance are low in normal lungs or those with minor abnormalities present but increase as lung abnormalities increase. For the grayscale image extracts with the most severe lung abnormalities average increases and variance decreases, as the lighter grey shades become more dominant.

Lung CT Scan Grayscale Extract Image Exploitable Statistical Relationships

Figure 3

3D graphic display of key grayscale statistics for distinguishing severity of lung abnormalities. Grayscale P90, variance and percentage of pixels at the grayscale average, considered together are very effective at identifying the degree of severity of lung abnormalities from grayscale extract images of lung CT scans of healthy and COVID-19-positive patients.

Confusion Matrix Comparison for Best Performing ML Models

A) Random Forest 11-Variable Model (456 Data records)

Visual Scores (VS)	Class	Visual Scores (VS)				
		0	1	2	3	4
0	0	62	3	1	0	0
1	1	0	49	3	0	0
2	2	2	1	138	4	0
3	3	0	0	5	124	4
4	4	0	0	0	1	59
Errors:		2	4	9	5	4
Accuracy /class (%):		96.88	92.45	93.88	96.12	93.65
Total Errors: 24 out of 456		Overall Accuracy: 94.74%		RMSE: 0.2569		

B) Adaboost 9-Variable Model (456 Data records)

Visual Scores (VS)	Class	Visual Scores (VS)				
		0	1	2	3	4
0	0	62	1	0	0	0
1	1	2	52	5	0	0
2	2	0	0	139	2	1
3	3	0	0	3	123	6
4	4	0	0	0	4	56
Errors:		2	1	8	6	7
Accuracy /class (%):		96.88	98.11	94.56	95.35	88.89
Total Errors: 24 out of 456		Overall Accuracy: 94.74%		RMSE: 0.2324		

Figure 4

Confusion matrices for the two best performing ML models, illustrating how errors are distributed differently among the VS classes although both models achieve the same accuracy in terms of the total number of prediction errors they generate.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile.docx](#)