

Covid-19 related Lung Consolidation's Digital signature in CT scans

La signature digitale de la consolidation pulmonaire relative au Covid-19 dans les images tomodynamométriques

Adnane Laghrissi¹, Mohammed Douimi¹, El Moukhtar Zemmouri¹, Wajih Rhalim², Najwa Benslima³, Najib Alidrissi⁴, Khaoula Echabbi¹, Hassan Ghazal⁵, and Salsabil Hamdi⁶

1. Moulay Ismail University, Meknes, Morocco,
2. Mohamed 5 University, Rabat, Morocco
3. Hôpital Universitaire International Mohammed VI, Bouskoura, Morocco
4. Université Mohammed VI des Sciences de la Santé, Casablanca, Morocco
5. Centre National pour la Recherche Scientifique et Technique, Rabat, Morocco
6. Institut Pasteur du Maroc, Casablanca, Morocco

ABSTRACT

Introduction: Research analyzing COVID-19 CT imagery categorizes visual findings into three specific groups, the third of which is lung consolidation (LC).

Objective: This study evaluates CT scans from RT-PCR-confirmed COVID-19 cases to define the digital signature of lung consolidation based on grayscale pixel intensity.

Methods: By leveraging statistical analysis and modelization techniques, we implemented a dual-threshold filter in Python to achieve efficient CT image segmentation.

Results:

1- The pixel's grayscale values of Covid-19 related lung consolidation are normally distributed

2- The proposed filter successfully isolates the lung consolidation spots

Conclusion: This study presents a novel statistical framework for identifying COVID-19 lung consolidation patterns in CT imagery. Given the observed variability across different imaging hardware and geographic cohorts, further investigation is required to improve the model's generalizability.

Keywords: Covid-19, Computed tomography, Lung consolidation, Statistics, Image segmentation

RÉSUMÉ

Introduction: Les recherches analysant l'imagerie scanner de la COVID-19 classent les signes visuels en trois groupes spécifiques, le troisième étant la condensation pulmonaire (LC).

Objectif: Cette étude évalue des scanners CT de cas de COVID-19 confirmés par RT-PCR afin de définir la signature numérique de la condensation pulmonaire en se basant sur l'intensité des pixels en niveaux de gris.

Méthodes: En tirant parti de l'analyse statistique et des techniques de modélisation, nous avons implémenté un filtre à double seuillage en Python afin de réaliser une segmentation efficace des images CT.

Résultats:

1- Les niveaux de gris des pixels correspondant aux consolidations pulmonaires liées à la COVID-19 suivent une distribution normale.

2- Le filtre proposé isole avec succès les zones de consolidation pulmonaire.

Conclusion: Ce travail introduit une approche statistique novatrice pour la détection des profils de condensation pulmonaire liés à la COVID-19 dans les scanners CT. Toutefois, les résultats présentent des variations significatives dépendant des appareils et de la géographie, ce qui nécessite des recherches approfondies pour améliorer la généralisabilité des résultats.

Mots clés: COVID-19, Tomodynamométrie, Consolidation pulmonaire, Statistiques, Segmentation d'images

Correspondance

Adnane Laghrissi

Moulay Ismail University, Meknes, Morocco

Email: laghrissiadnane@gmail.com

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious illness caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). First identified in Wuhan, China, in December 2019, the virus rapidly spread worldwide, prompting the World Health Organization (WHO) to declare it a pandemic on March 11, 2020 (1). The global response to COVID-19, including widespread lockdowns, led to significant social, economic, and health disruptions (2). However, it also had unintended benefits, such as a 26% reduction in global CO₂ emissions (3). By late October 2021, confirmed cases surpassed 246.5 million, with nearly 5 million deaths reported (4).

Detection Methods

Combating the pandemic relies on accurate detection, which involves two complementary scientific fields: molecular biology and radiology (5).

Molecular Biology

This field focuses on cellular molecules and their interactions. A key method for viral detection is Real-Time Polymerase Chain Reaction (RT-PCR), which amplifies viral DNA to measure viral load in samples (e.g., nasopharyngeal swabs). While RT-PCR is widely used for COVID-19 diagnosis, it carries a 25% false-positive risk (6), necessitating confirmation through radiological analysis.

Radiology

Radiographic imaging plays a critical role in diagnosing and monitoring COVID-19 (7). The two primary techniques are:

- Computed Tomography (CT Scans)
- X-rays

Radiologists play a critical role in disease diagnosis by interpreting radiological images (up to 300 per scan) alongside clinical data. This process relies heavily on visual analysis and professional expertise, which can sometimes lead to interpretation errors or diagnostic discrepancies among practitioners (8).

Current research trends in radiology focus on automating diagnostic processes or at least reducing error rates through technological assistance (9). Given the widespread digitization of medical imaging and clinical data, computer-aided interpretation has emerged as the most promising approach to achieve these objectives.

For our analysis, we selected chest CT scans due to their superior image quality and lower noise interference compared to conventional chest X-rays. Recent studies on COVID-19 progression through chest CT imaging reveal a distinct visual evolution pattern (10):

- Initial Stage: Ground Glass Opacity (GGO)
- Intermediate Stage: Crazy Paving Pattern
- Advanced Stage: Lung Consolidation (LC)



Figure 1. Examples of visual manifestation of The three stages

Approach

Implementing and fine-tuning neural networks presents significant challenges, particularly when processing large datasets where computational demands escalate exponentially (11). To address these limitations, we propose an alternative computer-aided detection method that operates independently of neural network architectures.

Our approach marks a pioneering step in automating the identification of Lung Consolidation (LC)- a key radiological manifestation of COVID-19 in chest CT scans. This methodology offers both simplified implementation and enhanced processing efficiency compared to

conventional deep learning solutions.

METHODS

Research Objectives and Methodology

This study seeks to establish the statistical distribution governing grayscale intensity values of Lung Consolidation (LC) pixels in chest CT scans from RT-PCR-confirmed COVID-19 patients. Upon identifying a consistent statistical pattern, we propose a two-stage filtering framework to isolate lung consolidation regions through image segmentation.

Statistical Model Design

The filter is derived from a statistical model designed to characterize COVID-19 lesion grayscale levels. A primary challenge in this characterization stems from significant grayscale variability, attributable to three key factors:

- **Infection Duration:** Grayscale levels evolve dynamically across disease stages.
- **Scanner Heterogeneity:** Grayscale values vary by scanner manufacturer and model.
- **Noise Interference:** Stochastic noise introduces grayscale fluctuations.

Mitigation Strategies

To minimize these variabilities, we adopt a relative grayscale characterization approach:

- **Temporal Variability (Infection Duration)**
- **Align analysis with the three distinct pathological stages outlined in this paper's introduction.**
- **Propose stage-specific filter adaptations to account for grayscale evolution.**
- **Scanner Dependency**
 - Restrict analysis to CT images acquired from a single scanner model (selected for superior image quality and consistency).
- **Noise Reduction**
 - Apply a cascaded preprocessing pipeline:
 - Grayscale normalization
 - Gaussian blurring to suppress residual noise while preserving lesion morphology.

Data set specifications

Our study utilizes a collection of 10,200 CT images stored in 12-bit DICOM format. The dataset comprises scans from 35 RT-PCR-confirmed COVID-19 positive patients, with an average of approximately 300 images per patient. All images were acquired using an "OPTIMA CT660" scanner (General Electric Healthcare) and were provided by Cheikh Khalifa Hospital in Casablanca, Morocco.

Sampling Methodology

From this comprehensive dataset, we systematically selected three representative sample groups. The selected images meet the following criteria:

Table 1. Image grouping conditions and number of images by group

Group ID	Conditions on images	Nbr of images
1	a. Contains at least one LC spot b. The LC spot is significantly large c. The contour of the LC spot is clearly defined	50
2	a. Contains at least one LC spot b. The LC spot is not significantly large c. The contour of the LC spot is not clearly defined	40
3	a. Doesn't contain any LC spot	40

Image Preprocessing Pipeline

To enhance image quality and reduce noise, we implemented a two-stage preprocessing approach:

Grayscale Conversion:

Applied uniformly to all image groups¹³

Transforms RGB values to grayscale using channel mean equalization:

For each pixel: $gsv = (R + G + B)/3$

Gaussian Smoothing:

Implemented using a discrete 3×3 Gaussian kernel

Kernel parameters: $\sigma = [X]$ (standard deviation)

Performs weighted neighborhood averaging to reduce high-frequency noise (12).

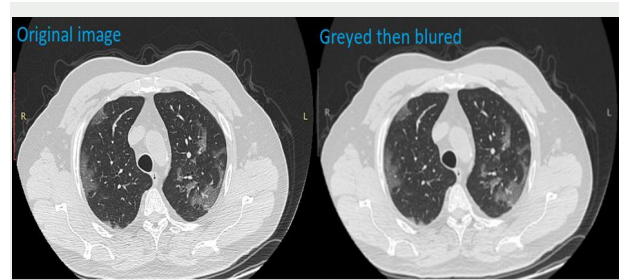


Figure 2. A chest CT scan before and after image processing example

Data extraction

For each image in group 1, and under radiologists control, we choose a 10 pixel² square in the significant center of every LC spot contained in the image. Then for each square we calculate the 100 pixels² values mean, which are listed in the following dataset:

Table 2. 10x10 pixel mean by LC spot in chest CT scans for images in group 1

3534	3282	3522	3335	3629	3468	3281	3570	3548	3612
3572	3287	3607	3610	3573	3522	3408	3510	3432	3372
3452	3452	3683	3630	3292	3503	3342	3602	3310	3508
3326	3422	3227	3530	3400	3518	3361	3547	3581	3571
3423	3372	3510	3548	3685	3570	3341	3313	3227	3681

Statistical analysis

We are interested in the characteristic X , which is the mean of each 10x10 pixel square grayscale level of lung consolidation spot contained in Table 2.

We model X as a random variable that follows a distribution within the population P . This distribution has a mean E and variance V :

$$E(X) = \mu \text{ and } V(X) = \sigma^2$$

The images in the data set are 12 bits in depth, which means that the number of possible grayscale values per pixel is $2^{12} = 4096$

Since it is impossible to consider P in its entirety, we propose using a sample of $n=40$ selections from the images in P . A selection is a matrix A .

$$A = (a_{ij})_{1 \leq i, j \leq p} \quad \text{with } a_{ij} \in \{0, \dots, 4095\}$$

Representing the grayscale level of Lung Consolidation lesions in the pixels, and $p=100$ is the dimension of A. The selection is made by choosing A in the significant center of the lung consolidation spot. We strongly assume that the sample selections are drawn randomly. Each selection has an equal probability of being drawn.

We consider the random experiment E, which consists of randomly drawing a selection A from P. We associate with E the random variable X, which follows a certain distribution. We conduct the same experiment n times under independent conditions. The n experiments produce a sample, which is the set $\{X_1, X_2, \dots, X_n\}$ of n independent and identically distributed random variables with the same distribution as X, represented in Table.2.

Our objective is to extrapolate (statistical inference) the characteristics of the mean μ and the variance σ^2 of the grayscale levels of the lung consolidation spots. The estimates of these unknown characteristics μ and σ^2 for the population are derived from the sample data.

We propose the following point estimator for the mean:

$$\underline{X} = \frac{1}{n} \sum_{k=1}^n X_k$$

And the following estimator for the Variance:

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \underline{X})^2$$

Realization of $\underline{X}_k$ is :

$$\underline{x}_k = \frac{1}{p^2} \sum_{i=1}^p \sum_{j=1}^p a_{ij}^k$$

And realization of $\underline{X}$:

$$\underline{x} = \frac{1}{n} \sum_{k=1}^n \underline{x}_k$$

These estimators are convergent and unbiased:

$$\underline{X} = \mu \quad \text{and} \quad E(\underline{X}) = \mu$$

Because it is the mean of means. Therefore, the empirical mean calculated on a sample is a good estimate of the mean in the population, and the population is normally distributed:

$$\mu = \underline{X}$$

Using Kolmogorov-Smirnov test, we consider the following hypothesis :

H_0 : the X characteristic of population P is normally distributed

In this case, P-value = 0.2827 which is higher than $\alpha = 0.1365$, so H_0 is accepted, hence the population P is normally distributed

According to the Central Limit Theorem:

$$\underline{X} \xrightarrow{L} n \rightarrow +\infty N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

And due to the normality of the distribution of this population we can confidently say that 95% of the means are contained in the interval [A, B] where:

$$A = \mu - 2\delta \quad \text{and} \quad B = \mu + 2\delta$$

Hence we can define the two level filter "F" as having A as lower level, and B as higher level, this filter changes the gray scale value (gsv) of every image pixel in the following way:

If $A < \text{gsv} < B$

Change Gsv to 0

Else

Change Gsv to 4000

The application of this filter to group 2 and 3 will be shown in the result and discussion section

RESULTS AND DISCUSSION

In order to verify the efficiency of the filter "F" we apply it to group 2 and group 3 images, examples of results are shown in Fig. 3, and Fig. 4 respectively:

The visual results obtained by application of filter F to group 2 images are conclusive, and as expected, the filter is successful in highlighting the lung consolidation spots only, for a calculated accuracy of 97% for this groupe

The visual results obtained by application of filter F to group 3 images, are conclusive, and as expected, the filter ignores all spots that are not lung consolidation, for a calculated accuracy of 100% for this group.

CONCLUSION

While this paper presents a novel statistical approach for identifying digital signatures of COVID-19 related lung consolidation patterns, an important limitation must be acknowledged. Our exclusive use of imaging data from a single scanner model means the identified signature characteristics may be scanner-dependent. Consequently, these findings should be validated through comparative studies incorporating data from additional scanner manufacturers before broader clinical application can be considered. This necessary extension of our work represents an important direction for future research.

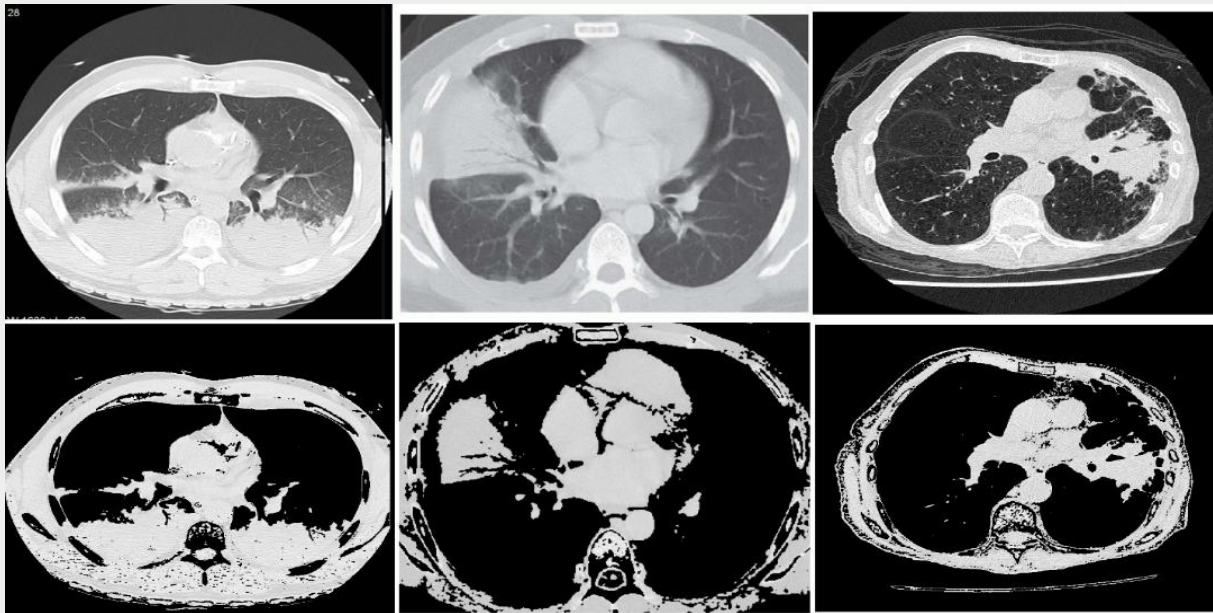


Figure 3. Filter F applied to group 2 images



Figure 4. Filter F applied to group 3 images

REFERENCES

1. Domenico Cucinotta MV. WHO Declares COVID-19 a Pandemic. National Center of Biotechnology Information [Internet]. 2020. Available from: <https://pubmed.ncbi.nlm.nih.gov/32191675/>
2. Tison GH, M M. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. *Annals of internal medicine* [Internet]. 2020; Available from: <https://www.acpjournals.org/doi/full/10.7326/M20-2665>
3. Le Quéré C, Jackson RB, Jones MW, Smith AJP, Abernethy S, Andrew RM, et al. Temporary reduction in daily global CO₂ emissions during the COVID-19 forced confinement. *Nat Clim Chang*. 2020 Jul;10(7):647–53.
4. Johns Hopkins Coronavirus Resource Center [Internet]. [cited 2025 Apr 26]. COVID-19 Map. Available from: <https://coronavirus.jhu.edu/map.html>
5. Buddhisha Udugama PK. Diagnosing COVID-19: The Disease and Tools for Detection [Internet]. American Chemical Society; 2020. Available from: <https://pubs.acs.org/doi/10.1021/acsnano.0c02624>
6. Surkova E, Nikolayevskyy V, Drobniewski F. False-positive COVID-19 results: hidden problems and costs. *The Lancet Respiratory Medicine*. 2020 Dec 1;8(12):1167–8.
7. The Role of Emergency Radiology in COVID-19: From Preparedness to Diagnosis - Muhammad Umer Nasir, James Roberts, Nestor L. Muller, Francesco Macri, Mohammed F. Mohammed, Shahram Akhlaghpour, William Parker, Arash Eftekhari, Susan Rezaei, John Mayo, Savvas Nicolaou, 2020 [Internet]. [cited 2025 Apr 26]. Available from: <https://journals.sagepub.com/doi/full/10.1177/0846537120916419>
8. Delrue L, Gosselin R, Ilse B, Van Landeghem A, de Mey J, Duyck P. Difficulties in the Interpretation of Chest Radiography. In: Coche EE, Ghaye B, de Mey J, Duyck P, editors. *Comparative Interpretation of CT and Standard Radiography of the Chest* [Internet]. Berlin, Heidelberg: Springer; 2011 [cited 2025 Apr 26]. p. 27–49. Available from: https://doi.org/10.1007/978-3-540-79942-9_2
9. Gadde SS. Artificial Intelligence - The Future of Radiology. *International Journal for Research in Applied Science & Engineering Technology* [Internet]. 2020 Jan 1 [cited 2025 Apr 26]; Available from: https://www.academia.edu/43389861/Artificial_Intelligence_The_Future_of_Radiology
10. Radiologic diagnosis of patients with COVID-19 [Internet]. [cited 2025 Apr 26]. Available from: <http://www.elsevier.es/en-revista-radiologia-english-edition--419-avance-resumen-radiologic-diagnosis-patients-with-covid-19-S2173510721000033>
11. Chang KW. Large Linear Classification When Data Cannot Fit in Memory. *dblp computer science bibliography* [Internet]. 2012. Available from: https://www.researchgate.net/publication/220345144_Large_Linear_Classification_When_Data_Cannot_Fit_in_Memory
12. Mirzaei Shahnam ES. Noise reduction and image smoothing using gaussian blur. California State University, Northridge. Department of Electrical and Computer Engineering [Internet]. 2015. Available from: <https://scholarworks.csun.edu/handle/10211.3/142402?show=full>