 doi:10.20944/preprints202207.0156.v1

Article

A Novel Method for Lung Image Processing Using Complex Networks

Laura Broască[#1], Ana Adriana Trușculescu[#2,3] *, Versavia Maria Ancușa[1], Horia Ciocârlie[1], Cristian-Iulian Oancea[2,3], Emil-Robert Stoicescu[4,5] and Diana Luminița Manolescu[3,4]

[1] Department of Computer and Information Technology, Automation and Computers Faculty, "Politehnica" University of Timișoara, Vasile Pârvan Blvd. No. 2, 300223 Timișoara, Romania; laura.broasca@cs.upt.ro(L.B.); versavia.ancusa@cs.upt.ro(V.M.A.); horia.ciocarlie@cs.upt.ro(H.C.)
[2] Pulmonology Department, 'Victor Babes' University of Medicine and Pharmacy, Eftimie Murgu Square 2, 300041 Timișoara, Romania; ana.trusculescu@umft.ro(A.A.T.), oancea@umft.ro(C.I.O.)
[3] Center for Research and Innovation in Precision Medicine of Respiratory Diseases (CRIPMRD), 'Victor Babes', University of Medicine and Pharmacy, 300041 Timișoara, Romania
[4] Department of Radiology and Medical Imaging, 'Victor Babes' University of Medicine and Pharmacy Timisoara, Eftimie Murgu Square No. 2, 300041 Timișoara, Romania; stoicescu.emil@umft.ro(E.R.S.); dmanolescu@umft.ro(D.L.M.)
[5] Research Center for Pharmaco-Toxicological Evaluations, 'Victor Babes' University of Medicine and Pharmacy Timisoara, Eftimie Murgu Square No. 2, 300041 Timișoara, Romania;
* Correspondence: ana.trusculescu@umft.ro

Abstract: The High-Resolution Computed Tomography (HRCT) detection and diagnosis of diffuse lung disease is primarily based on the recognition of a limited number of specific abnormal findings, pattern combinations or their distributions, as well as anamnesis and clinical information. Since texture recognition has a very high accuracy percentage if a complex network approach is used, this paper aims to implement such a technique customized for diffuse interstitial lung diseases (DILD). The proposed procedure translates HRCT lung imaging into complex networks by taking samples containing a secondary lobule, converting them into complex networks and analyzing them in 3 dimensions: emphysema, ground glass opacity and consolidation. This method was evaluated on a 60 patient lot and the results show a clear quantifiable difference between healthy and affected lungs. By deconstructing the image on three pathological axes, the method offers an objective way to quantify DILD details which, so far, have only been analyzed subjectively.

Keywords: diffuse interstitial lung disease; complex networks; model; HRCT

Abbreviation List

CAD- Computer Aided Diagnosis

COPD- Chronic Obstructive Pulmonary Disease

CT-Computer Tomography

DILD- Diffuse interstitial lung Diseases

FOV – Field of View

GE- General Electrics

GGO- Ground Glass Opacity

HRCT-High Resolution Computed Tomography

HPc-Chronic Hypersensitivity Pneumonitis

HU-Hounsfield Unit

ILD-Interstitial Lung Diseases

IPF-Idiopathic Pulmonary Fibrosis

 doi:10.20944/preprints202207.0156.v1

NSIP-Non-Specific Interstitial Pneumonia

OP- Organizing Pneumonitis

PCR- Polymerase Chain Reaction

PFT- Pulmonary Function Test

SPL- Secondary Pulmonary Lobule

UIP- Usual Interstitial Pneumonia

1. Introduction

Pathological alterations that affect lung interstitium usually start with an abnormally strong inflammatory process which inhibits alveoli expansion. In time, the inflammation strain is replaced with permanent rigidity due to scar tissue, which in turn creates more inflammation, progressing towards serious clinical outcomes. This cycle of inflammation and fibrosis in the lung interstitium is the unifying characteristic for the Diffuse Interstitial Lung Diseases (DILD) group[1]. Historically, this heterogeneous group of more than 200 distinct diseases that affect the lung parenchyma have seen recurring challenges concerning the terminology, classification, and staging of the DILDs[2].

Due to having the same pathological process, the clinical and, in part, paraclinical characteristics used in DILD diagnosis tend to overlap, yet distinct pathology origins require differentiation in order to successfully issue a treatment course. There is no better example in this case than that of Idiopathic Pulmonary Fibrosis (IPF), which has a median survival rate of 2-5 years, yet its clinical diagnosis can easily be mistaken for the much more common Chronic Obstructive Pulmonary Disease (COPD) with a considerably better prognosis (mild cases have a 10-20 years survival rate)[3].

The DILD progressive aspect presents the challenge of an early and accurate diagnosis, which almost doubles the survival rate and improves life quality by employing the right treatment [4]. Clinical signs and symptoms overlap, as previously stated, therefore paraclinical methods are crucial to properly diagnose DILD. However, the more commonly used investigations, like chest X-Ray (CXR), peripheral blood tests, spirometry need to be complemented with the more specialized imagistic tools like High Resolution Computed Tomography (HRCT), lung ultrasound and, in particular cases, with bronchoscopy and surgical lung biopsy[5,6].

The HRCT has been the central noninvasive instrument in analysis since the 2011 updated imagistic diagnostic guidelines[7], offering crucial details and insights that can lead to a swift diagnosis[8]. As with any diagnostic tools, there can be intricacies which require either a very specialized technician and / or further, more invasive, investigations. Moreover, substantial inter-observer variance, even between experienced radiologists, complicates the process[9–12]. The current approach is to try and supplement human interpretation of HRCT with automated tools, like the CALIPER software [13] or various AI-based tools, like [14]–[17].

This paper starts by shortly presenting the way diagnosis is made by computers and humans respectively. Subsequently, a novel technique is presented and then assessed from both a biological and a system science perspective.

1.2. Using HRCT – humans & computers

In diagnosing DILDs, medical specialists start with the HRCT pattern recognition of a limited number of specific abnormal findings, particular combinations or patterns of these abnormalities, one or more discrete distributions of abnormal findings, and the use of basic history and clinical information.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

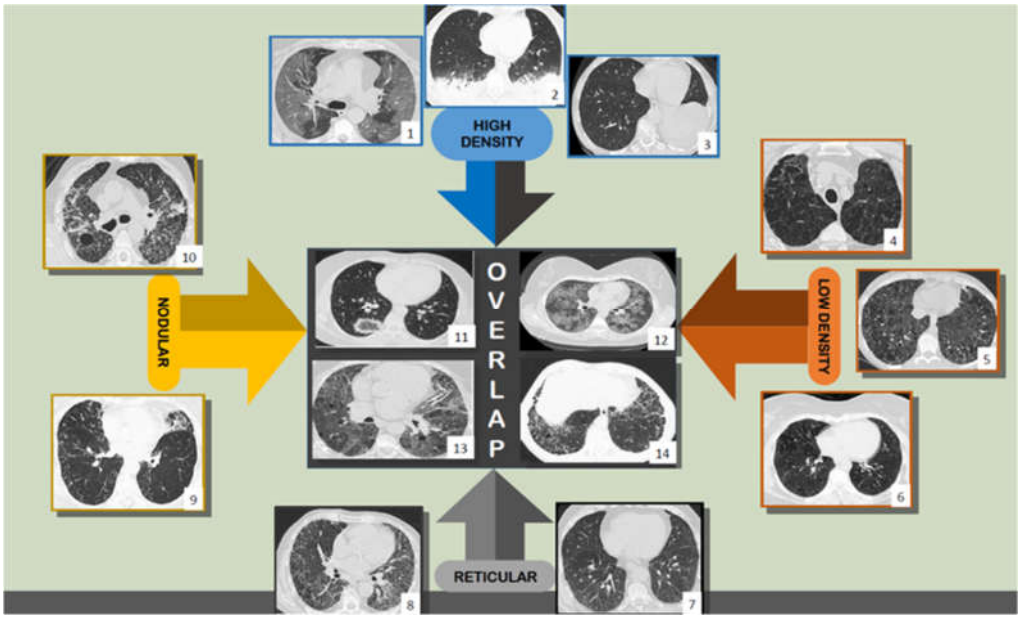


Figure 1. Axial thin-section CT scans, injury patterns: high density (1, 2, 3), low density (4, 5, 6), reticular (7, 8), nodular pattern (9, 10) and overlapping (11, 12, 13, 14). Scans belong to the 'Dr. Victor Babes' Infectious Diseases and Pneumoftiziology Clinical Hospital Timisoara database.

Radiological DILD diagnosis is pattern-based and linked to the underlying histology. It is anticipated that the future of DILD identification will involve behavior-based radiological phenotypes, in consequence determining clinical management [18]. By classifying primary lesion types into four categories – reticular pattern, nodular pattern, high attenuation and low attenuation – a diagnosis can be achieved. Their overlap and association with other lesions matter [19] (Figure 1), as well as their distribution in the lung and in the basic structural and functional unit of the lung - the secondary lobule.

Thin-section CT, inspiratory, expiratory, and prone sequences comprise the most sensitive radiologic examination to evaluate the lung parenchyma for evidence of ILD. The key anatomic components of the lung parenchyma examined in IPF are the interstitium and secondary pulmonary lobule (SPL)[20]. Consequently, histological phenotypes and lesion types (primary lesions and/or its overlapping model) as well as their lung and SPL distribution could compete and work together to indicate an accurate clinical syndrome. For example, *Usual Interstitial Pneumonia* (UIP) is the classic progressive fibrotic phenotype, yet self-sustaining progressive fibrosis is not only found in IPF patients, but also in the progressive *Non-Specific Interstitial pneumonia* (NSIP) or *chronic hypersensitivity pneumonitis* (PHc) phenotype.

This (human) approach is vastly different from the way Computer Aided Diagnosis (CAD) works. Most CAD use heuristics and machine learning without an analytical process, being focused on proper classification and not on the underlying causes. This approach, encountered in [14]–[17] does not allow for any type of evaluation in terms of progress or severity. Programs that try a more anatomy-based tactic [21], [22], [23], commercial and scientific alike, may require additional input data, such as Pulmonary Function Tests (PFT) (e.g.: Caliper), and their output only reports the abnormal volume. No qualification is offered with respect to the lesion severity in that volume. However, among the advantages of using such tools are: relatively fast processing times, verified results, having a good, reproducible precision and successfully assisting the medical personnel.

The HRCT slices contain a non-visual apparent information stored as Hounsfield Units (HU) that can enhance the way gradient difference between pixels relates to textural differences. The various densities interweave in their geometrical placements to create textures. Since texture recognition has a very high accuracy percentage when a complex

 doi:10.20944/preprints202207.0156.v1

network approach is used[24], [25], [26], this paper aims to implement such a technique customized for DILD.

2. Materials and Methods

2.1. Lot selection

To choose the eligible patients we used 'Dr. Victor Babes' Infectious Diseases and Pneumoftiziology Clinical Hospital Timisoara database, stored in their private cloud archive. From more than 30000 imaging exams, stored in Digital Imaging and Communications in Medicine (DICOM) format, a total of 60 scans were selected, based on the following inclusion criteria:

- 30 patients with CT exams and exploratory function tests with the diagnosis of DILD (diffuse interstitial lung disease);
- 30 patients with normal CT imaging that were considered the control group;

All the participants provided a written consent for the usage of their HRCT scans. In addition, the Ethics Committee of Clinical Hospital for Infectious Diseases and Pneumoftiziology Dr. Victor Babeș Timișoara (protocol code 11835/26.11.2021) also approved this study.

2.2. Imaging parameters

All examinations were performed with a General Electrics (GE) Healthcare Optima 520 16-slices, with 32-slices reconstruction. The scanner is a 0.5 mm x 16 detector-row, allowing for an 8 mm total z-axis length. Every patient was examined with a constant setting portocol, variation occuring only on radiation dose, due to variable tissue penetration.

The HRCT parameters used are the following:

- slice thickness: 1.25 mm
- scan time: 1 second
- kV: 120
- mAs: 130
- collimation: 2.5 mm
- matrix size: 768 x 768
- Field of View (FOV): 35 cm
- reconstruction algorithm: high spatial frequency
- window: lung window
- patient position: supine (usually) or prone position (if DILD is suspected)

The slice is narrower than the recommended 1.5mm from the Radiology Working Group of the Pulmonary Fibrosis Foundation, to allow for a better and smoother lesion detection as well as a higher accuracy – both very crucial aspects of DILD diagnosis. Spatial resolution (pixel spacing) for these settings is 0.74 mm.

The HRCTs were stored in a DICOM format, as it is the universal form for encrypted medical imaging with a high transmissibility property. The algorithm behind the DICOM encodes the personal information of the patients, CT information, technical parameters, and medical images, making it difficult to read without a specific application.

The main criteria for analyzing image data were the tissue densities/opacities and these were determined by applying Hounsfield scale's principles. The Hounsfield Units (HU) are commonly used for a quantitative analysis of radio density and tissue tightness, being useful for the interpretation of CT scans. Image reconstruction relies on the tissue properties regarding X-ray beam penetration and attenuation in order to define a grayscale image system. These grayscale intervals vary between approximately -1000 HU (air) to 3000 HU (metals like silver, steel), according to the attenuation range of tissue absorption. This transformation is represented by a gray tones scale and has as a landmark the density of distilled water, which is defined as zero HU.

Table 1. HU intervals from to Lin Li et al. and Maria Paola Belfiore et al. reports (9,10,11). These values are specific for the General Electrics Healthcare Optima 520

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

Pulmonary zones	HU intervals
Emphysema	[-1024, -977)
Normal pulmonary parenchyma	[-977, -703)
Ground-glass opacities	[-703, -368)
Others (crazy-paving, pleural fat)	[-368, -100)
Consolidations	[-100, 5)
Others (interstitial vessels)	>5 HU

According to the HU intervals illustrated in Table 1, each element of this lesional picture will have an equivalent. For example, the honeycombing-pattern is a mixt of cysts (emphysema) and reticulations (consolidations); the reticular fibers' network is a consolidation equivalent, since ground-glass opacities are already represented in the table.

For the studied pathologies, the selected intervals are those representing emphysema, normal pulmonary parenchyma, ground-glass opacities, and consolidations.

2.3. Image lot selection

A 65x65 pixel area was manually selected out of one of the HRCT slices, for each one of the HRCTs lot. The argument behind choosing such areas manually instead of processing the entire image at once is based on the idea of analyzing the most relevant samples for the chosen pathologies, taken in isolation. Only after specific patterns have been discovered would it be sensible to apply the findings on a larger scale.

In order to remover intra- and inter- observer variability the most relevant area for diagnosis was a majority intersection of selections made by four independent observers, two radiologists (10+ and 5 years thoracic experience) and two pneumologists (15+ and 5+ years DILD experience). For the DILD affected lot, these selections represent an extraneous diagnosis confirmation, since the images were already annotated by at least 3 specialists from the National Fibrosis Center of the 'Dr. Victor Babes' Infectious Diseases and Pneumoftiziology Clinical Hospital Timisoara.

The dimension for this sample area has been chosen based on multiple factors:

1. The more pixels a sample contains, the more processing power it requires for transforming it into a matrix and furthermore into a complex network. This also influences the processing time which could span from seconds to minutes.
2. This area should be both wide enough to capture relevant lung tissue for the diagnosis yet small enough to eliminate any extra types of tissue which might "contaminate" or add unnecessary complexity to the selected sample.
3. The selected square area should capture at least one functional component of the lung (secondary pulmonary lobule) in its entirety and with it, any type of illness it might suffer from. Given that one secondary lobule has an area ranging from 1 cm^2 to 2.5 cm^2 and that the pixel spacing within the selected HRCTs varies between 0.70 and 0.80 (this setting is machine dependent and is encoded into the HRCT metadata) then a sample rectangle of 65 x 65 px should normally include at least one secondary lobule. E.g.: Actual pixel spacing value for the lot is PS = 0.74 mm, retrieved as DICOM parameter. Given that the area of a secondary lobule is a 2.5 cm^2 x 2.5 cm^2, then the smallest valid DICOM sample of a secondary lobule should be 25 / 0.74 = 33.7837 mm. However, having in mind the idea of capturing at least one full secondary lobule, the sample area size is set to almost double that value. Alternative studies have also tried similar experiments with a cropped DICOM sample of only 11 x 11 px, yet it is not clear why this value was chosen. [22], [23].

2.4. Image processing algorithm

Each of the selected samples is then processed with the help of a Python written program developed specifically for this purpose. Using a specialized CT library, pydicom, the DICOM slices are cropped to the pre-established size (65 x 65 px) around the interest area.

The program consists of an algorithm meant to carry out the following steps:

1. Iterate over a set of HRCT slices (DICOM files)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

2. For each one, crop out a 65 x 65 pixel area
3. Analyze selected area from 3 perspectives:
   a. Convert pixel gradient into Hounsfield Unit value according to the formula:
   HUv = rescaleSlope * PxGradient + rescaleIntercept,
   where rescaleSlope and rescaleIntercept are constant values dependent on the CT equipment and embedded into the DICOM metadata, and PxGradient is the color code of a pixel.
   b. Isolate all emphysema-like tissue, GGO (Ground Glass Opacity), consolidation densities in the cropped image and leave out any other types of tissue (Figure 2)
   c. Separate each HU strip in the sample into a separate layer (Figure 2)
4. Generate complex networks out of each layer
5. Analyze connectivity, closeness, distribution of nodes (pixels)
6. Determine patterns of normal lungs, affected lungs

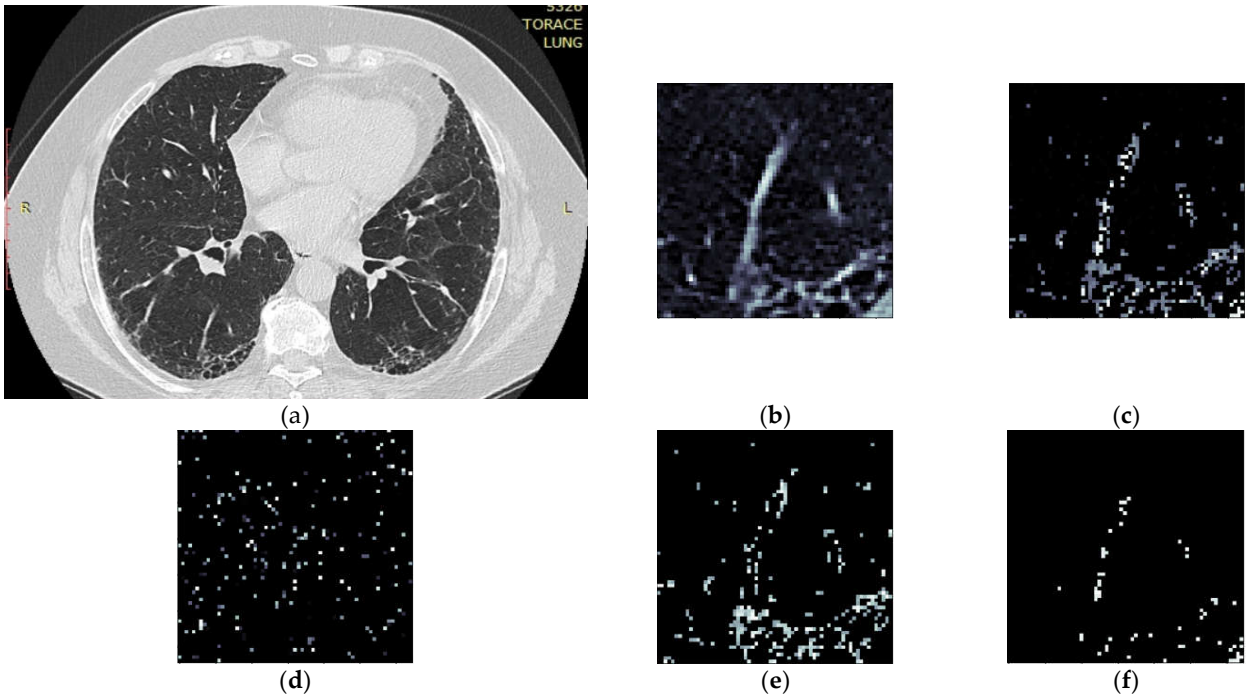


Figure 2. Splitting CT sample into layers
(a) original CT, (b) sample crop, (c) combined Emphysema, GGO, Consolidation layers, (d) Emphysema Layer, (e) GGO layer, (f) Consolidation layer.

In order to transform each of the crop layers (emphysema tissue, GGO tissue, consolidation) into complex networks (Step 4), the following are assumed:

1. Each pixel represents a network node and the pixel color (gradient) constitutes its value.
2. Two pixels are presumed to be connected if the following conditions are met:
   - The radial distance (Rd) between them (within the crop) is Rd ≤ 4 pixels. Assuming each pixel (Px) is the origin O of a circle with radius r = 4, every other pixel (Py) within the circle area can be considered connected. In other words: $\{\exists E(Px, Py) | d(Px, Py) \leq 4\}$.
   - The gradient difference between Px and Py is less than or equal to 50

Given the above conditions, the algorithm generates sets of nodes and connecting edges, exporting them into separate files for each individual layer. Thus, each lung HU

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

layer is converted into a complex network and analyzed from a degree distribution point of view.

Paragraphs 2.4.1 and 2.4.2 offer further insight into the threshold values selection processes.

2.4.1. Radial distance selection

In order to determine the radial distance at which lesions are singular or coupled, several trials have been carried out, using values in the range $1 \leq Rd \leq 8$ pixels.

Values less than 3 pixels resulted in a sparse network and very few connections, meaning that a small number of similar pixels were found in the immediate vicinity of each other. This leads to a relatively large amount of scattered clusters with less than 3 nodes in total. When compared to other Rd values, it does not convey much relevant information about the lung profile.

On the other hand, with Rd values above 5, while being more integrative, the algorithm becomes too permissive due to the specific complex network process of nodes attachment, linking similar nodes without an anatomical just cause. Defining a circle with radius r between 5 and 8 ($5 \leq r \leq 8$) allows for a more interconnected network, less clusters and a different degree distribution (Figure 3).

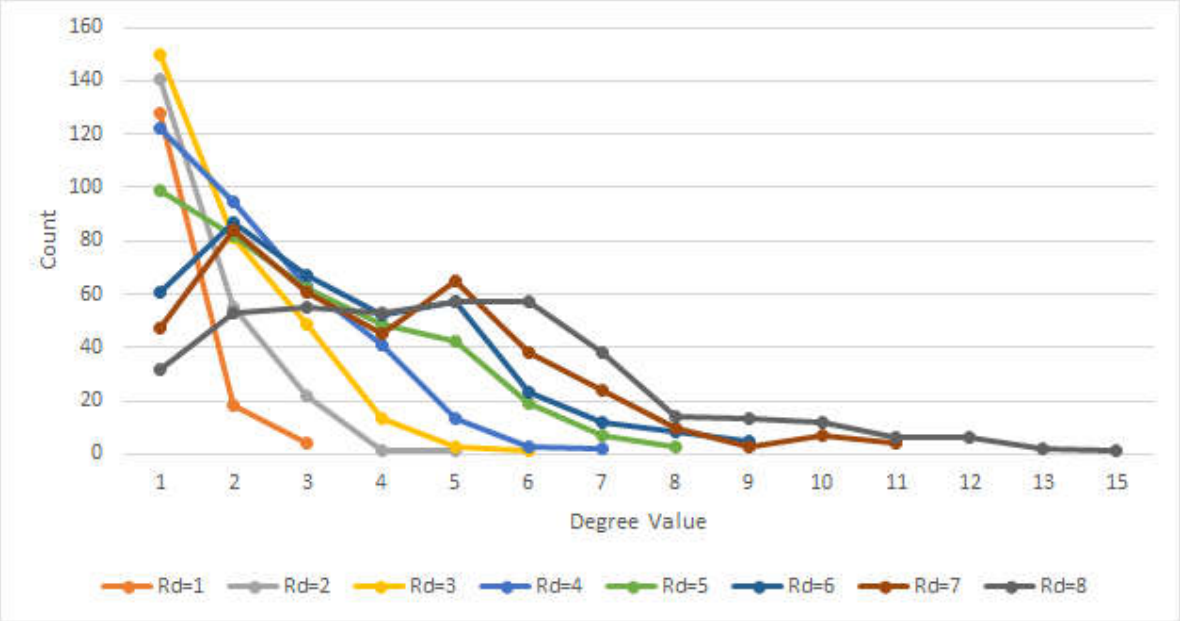


Figure 3. Degree distributions for various Rd.

Therefore, given the above experiment, it has been determined based on multiple trials that the most suitable Rd value is $Rd \leq 4$ pixels, which is big enough to generate dense clusters, yet small enough to make a difference in terms of degree distribution especially when comparing normal lungs with affected lungs.

This is confirmed by [27], which, in a clinical setting, uses an initial size for detectable lesions between 3–17 mm. Since a Rd = 4 pixels corresponds to a metric value of 4*0.74 = 2.96 mm, the obtained empirical result matches their results.

Further discussion regarding the distribution fit with a logarithmic or power function is presented in section 4.1, as it refers to model fit in the network science context.

2.4.2. Gradient difference threshold

In terms of gradient difference, the chosen delta determines whether two pixels are close enough in terms of graytones, or if they are too far apart on the grayscale to be considered adjacent. While a delta D = 50 covers the entire Emphysema strip, for the GGO and Condensation strips, it helps break the network into clusters. This rule can be summarized as:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

$$|G(Px) - G(Py)| \leq D,$$

where G(Px) and G(Py) are the respective gradient values of two pixels Px and Py, and D = 50 is the delta max threshold above which two pixels are not considered related.

In the end, each network layer can be defined as:

$N(P, E)\ where\ E = \{E(Px, Py) | d(Px, Py) \leq Rd\ and\ |G(Px) - G(Py)| \leq D\}$,

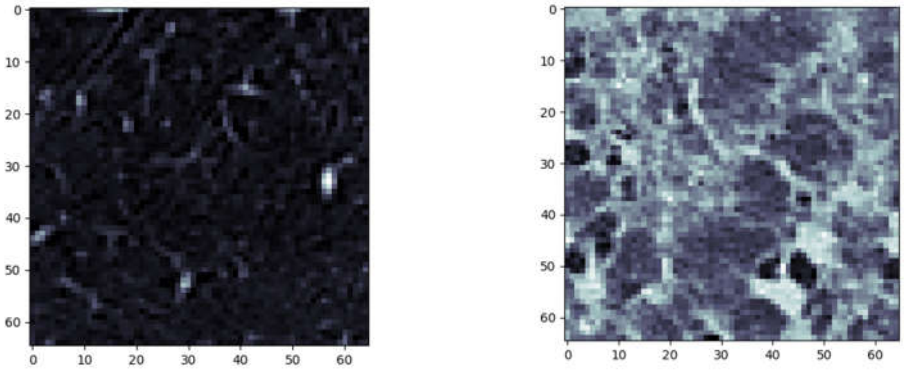
where P is the set of vertices or pixels, E is the set of edges, Rd = 4 and D = 50.

3. Results

Following the previously described method, all HRCTs (of both normal and affected lungs) were processed. The individual steps for a single normal and DILD affected patient (Figure 4) are showcased in subsection 3.1, with a further lot analysis presented in subsection 3.2

3.1 Normal and DILD case sample results

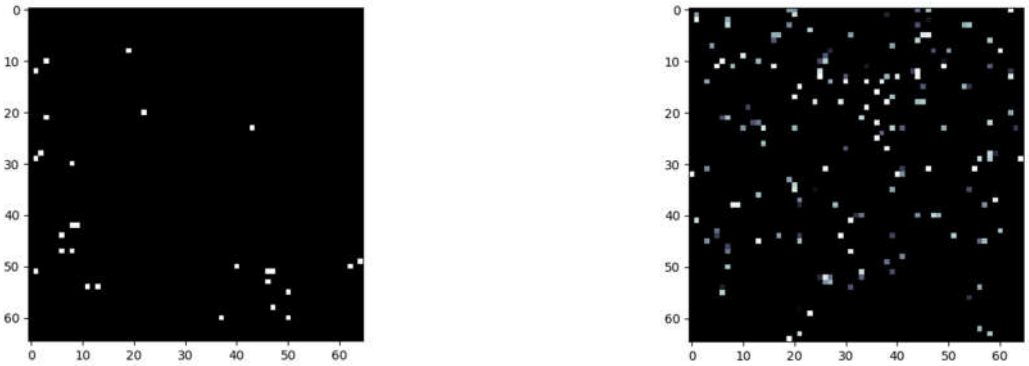
The first step is sample cropping into 65 x 65 pixels.



(**a**) (**b**)

Figure 4. Algorithm step 1 – sample selection
(**a**) Normal sample (**b**) DILD (IFP) sample.

Next steps imply splitting everything into layers and converting those layers into complex networks. First the emphysema layer is examined.



(**a**) (**b**)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

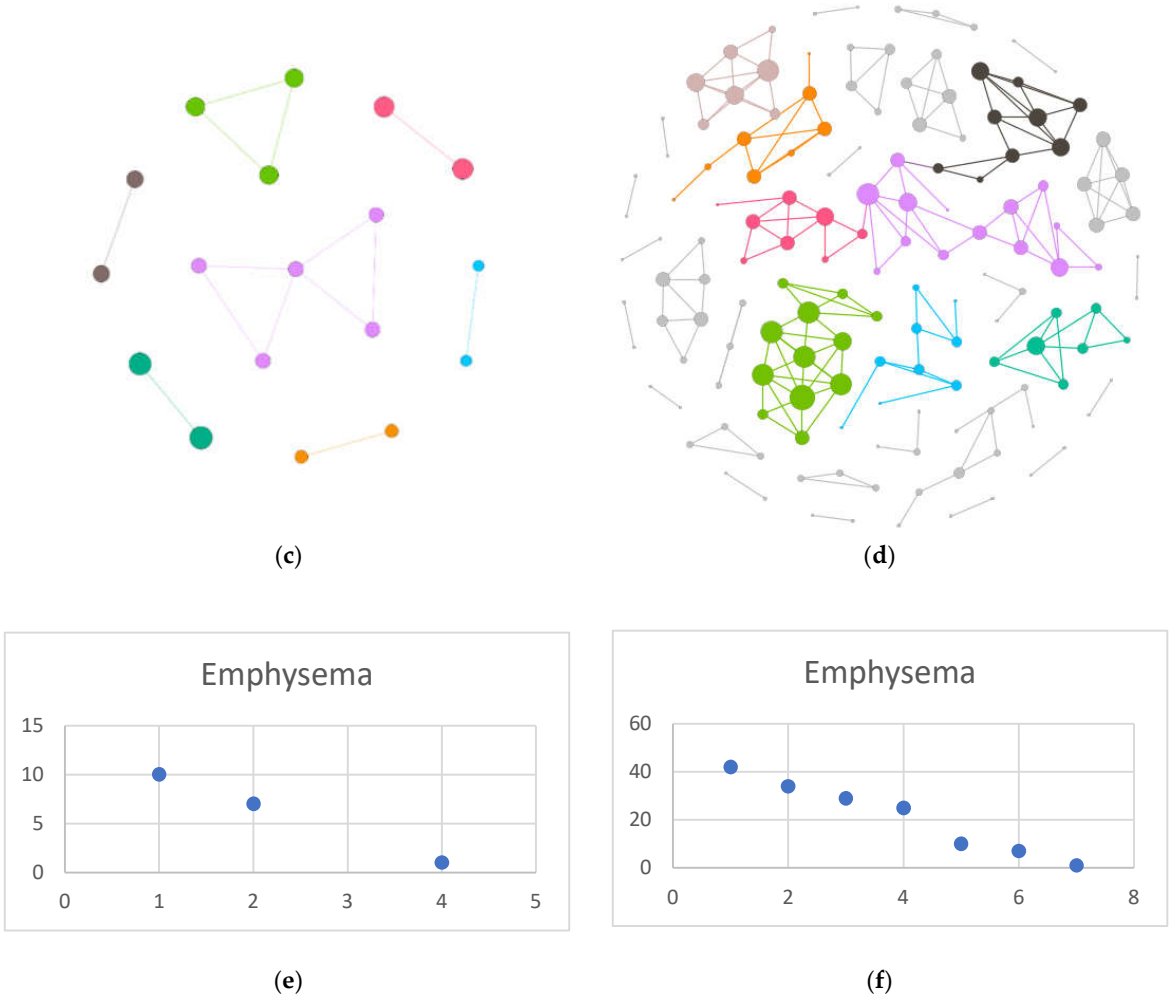


Figure 5. Emphysema processing
(**a**) HU filtered layer for the normal sample; (**b**) HU filtered layer for the DILD sample
(**c**) Complex network built according to the proposed algorithm corresponding to the normal sample, Fruchterman Reingold render layout, node sizes proportional to node degrees, edge width invariant (1.5 pixels). (**d**) Complex network built according to the proposed corresponding to the DILD sample, Fruchterman Reingold render layout, node sizes proportional to node degrees, edge width invariant (1.5 pixels). (**e**) Degree distribution the normal sample network (**f**) Degree distribution the DILD sample network.

Next, is the ground glass layer, and this is where major differences occur. Even though a visual inspection might evaluate distributions in Figure 6a and b as random, the network degree distribution shows a completely different story: a logarithmic distribution for the normal process and a polynomial one for the IFP.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

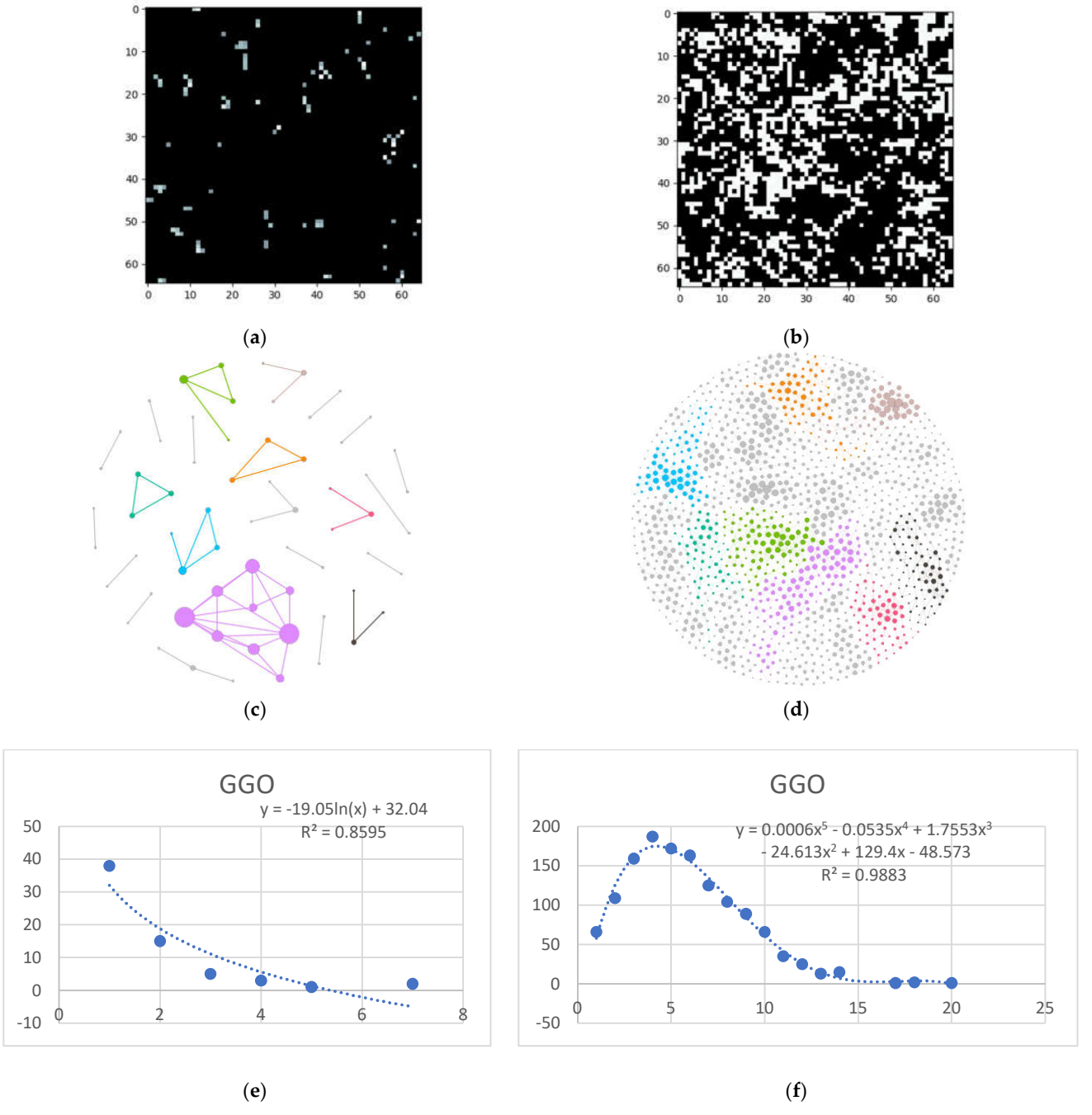


Figure 6. GGO processing
(**a**) HU filtered layer for the normal sample; (**b**) HU filtered layer for the DILD sample
(**c**) Complex network built according to the proposed algorithm corresponding to the normal sample, Fruchterman Reingold render layout, node sizes proportional to node degrees, edge width invariant (1.5 pixels). (**d**) Complex network built according to the proposed corresponding to the DILD sample, Fruchterman Reingold render layout, node sizes proportional to node degrees, edge width invariant (1.5 pixels). (**e**) Degree distribution the normal sample network (**f**) Degree distribution the DILD sample network. Equations for curve fit and R^2 are also presented on the relevant distributions.

Last, but not least, the consolidation layer.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

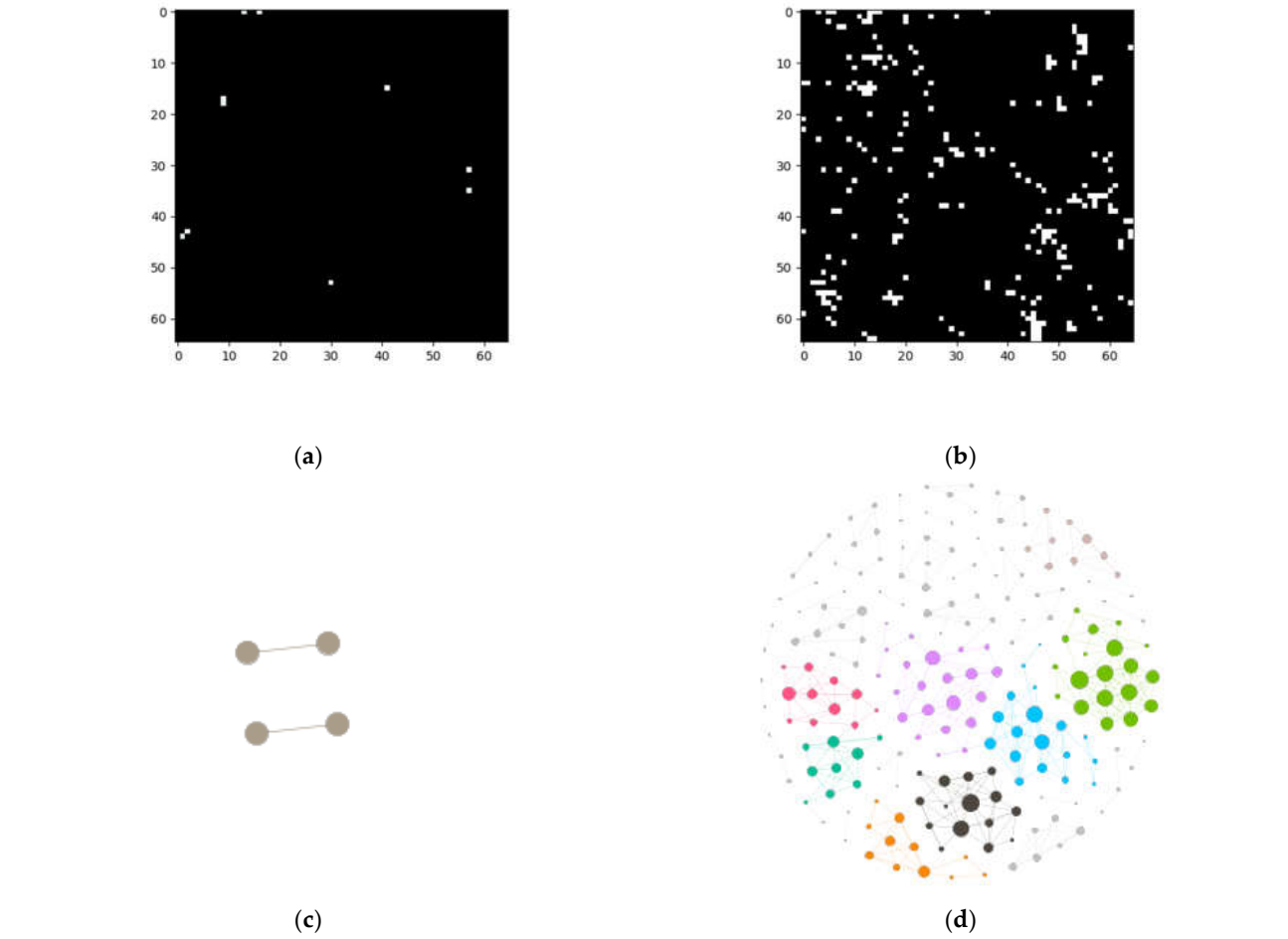


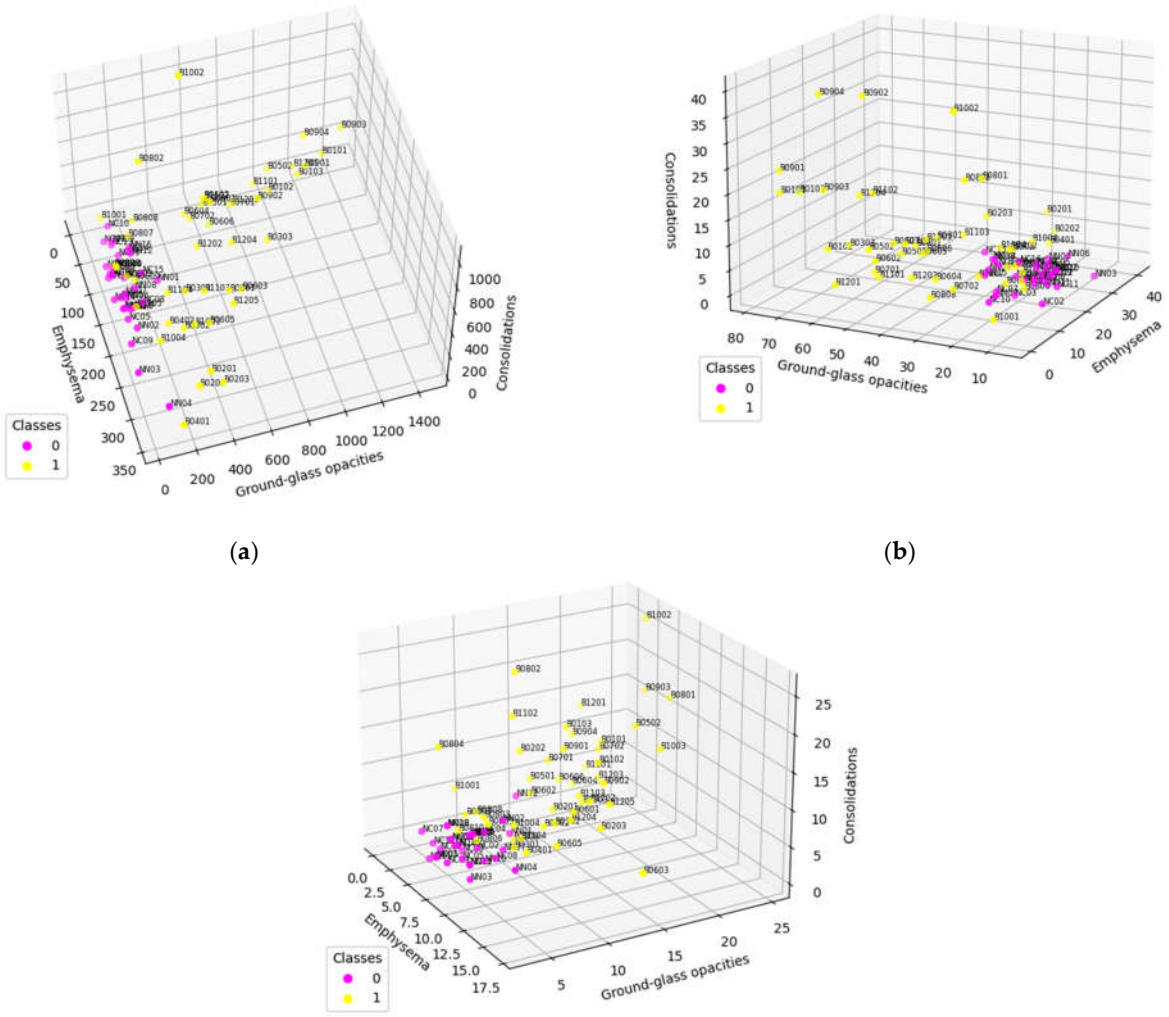
Figure 7. Consolidation processing
(**a**) HU filtered layer for the normal sample; (**b**) HU filtered layer for the DILD sample (**c**) Complex network built according to the proposed algorithm corresponding to the normal sample, Fruchterman Reingold render layout, node sizes proportional to node degrees, edge width invariant (1.5 pixels). (**d**) Complex network built according to the proposed corresponding to the DILD sample, Fruchterman Reingold render layout, node sizes proportional to node degrees, edge width invariant (1.5 pixels). (**e**) Degree distribution the normal sample network (**f**) Degree distribution the DILD sample network. Equations for curve fit and R^2 are also presented on the relevant distributions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

3.2. Results

At an individual level the differences can be fairly distinctive and the entire image lot analysis presented the challenge of determining network metric relevance, in a broader context.

In order to measure the network invariants entropy, a metric based on degree sequences is usually preferred. However, the differences shown in the previous subsection present the challenge of adding a measurement for the network size. Figure 8 shows three metrics, selected due to their overall balance between metrics that measure network complexity and size: total count (the degree sum), average count (average degree) and maximum degree.



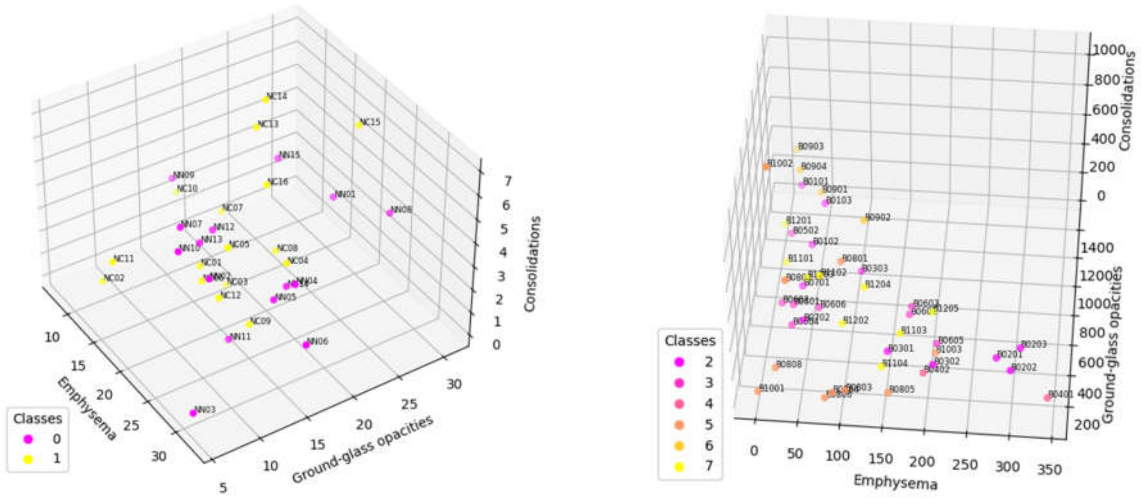
(**a**) (**b**)

(**c**)

Figure 8. Population distributions comparison according to specific complex network parameters: (**a**) Total count (**b**) Average count (**c**) Maximum degree
Class 0 (fucsia) represents normal lungs, while class 1 (yellow) is formed of DILD affected lungs.

To further study these results, normal and DILD patients distributions were plotted separately, adding another data layer (Figure 9). With normal patients, a distinction was made between patients diagnosed prior covid era and during covid. As for DILD patients, individual disease types were highlighted.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1



(**a**) (**b**)

Figure 9. (**a**) Normal population plotted based on average degree. Class 0 is normal population investigated prior covid, class 1 are cases diagnosed as normal in the pandemic era (**b**) DILD population plotted based on average degree. Class 2 is UIP, 3 probable UIP, 4 UIP and emphysema, 5 organizing pneumonitis (OP), 6 hypersensitivity pneumonitis (HP), 7 sarcoidosis.

As seen, there are a couple of outliers in what is otherwise a very tight distribution, and they will be further assessed in the discussion section.

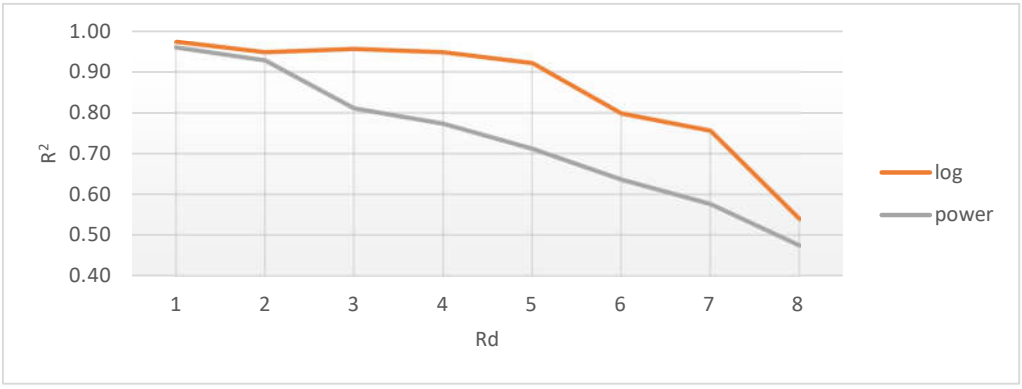
4. Discussion

As stated in the beginning of the paper, the goal was to create a complex-networks based model based on HRCT lung imaging. Having done so, an assessment needs to be made as to how well that model fits known frameworks from Network System Science and Medical Science.

4.1. Network System Science

One way to describe network systems based on real-world data is through their degree distributions, more specifically by the function type best fitting those distributions. Novel research, like [28], shows that the power and logarithmic functions define these systems. Empirical results, like those presented in Figure 3, Figure 5e, Figure 6e, Figure 7e showcase a logarithmic distribution at the proper biological resolution (Rd = 4) for normal patients. However, the power function fit on all the normal patients, even varying the radius, to safeguard from biological variations, is a very poor fit, especially when compared with the logarithmic function.

In Figure 10 the different fit of these distributions were tested against different relative distances between lung entities. Values less than 3 show a relatively similar fit, which



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

is mathematically correct, yet biologically incorrect because 1 and 2 pixels separation translate to a 0.74 mm to 1.48 mm gap, too small to be relevant.

Figure 10. Average coefficient of determination (R^2) for logarithmic and power distributions, relative to radial distance (Rd).

A possible rationalization for this result comes from the perspective that biological systems with feed-back have a power distribution, yet those without feed-back are characterized by a logarithmic distribution, as postulated in [29]. Since the lung is a system without tightly coupled feed-back loops, its distribution should follow the logarithmic model, as confirmed by our model.

Pathological lungs have an entirely different distribution, as shown in Figure 5f, Figure 6f, Figure 7f, best fitted with a polynomial function, not a logarithmic one. Literature results show that proliferative processes have polynomial distributions[30], [31] and since the studied DILDs have proliferative inflammation and fibrosis, they can be assimilated to the literature processes. Indeed, the proliferation cause is not necessarily a virus, however the histopathological propagation still follows the same principles.

Depending on the type of pulmonary damage, the degree of the function may vary in the range of [2,8] for the studied lot. This demands further exploration with enough data to tie the degree of the polynomial function to the type or complexity of disease that a patient suffers from. To be more specific, since lung diseases manifest themselves as a composition of the three mentioned axes (Emphysema, GGO, Consolidation), these three factors may vary differently in each case. So far, it can only be ascertained that there is a disease, but not its specific type. In order to be able to associate the illness complexity to a certain degree of a polynomial function, a more in-depth study, comprising separate large datasets, should be carried out.

The differences between the DILD affected networks and normal networks are not only distinctive, as presented in Figure 8, but can be further quantified if a simple standard deviation for all patient data series is computed. The results, presented in Figure 11 on all the three measurements considered for the networks (maximum degree, total count and average degree), for each HU band and for the combined pathological HU bands, prove a clear separation between the pathological and normal networks.

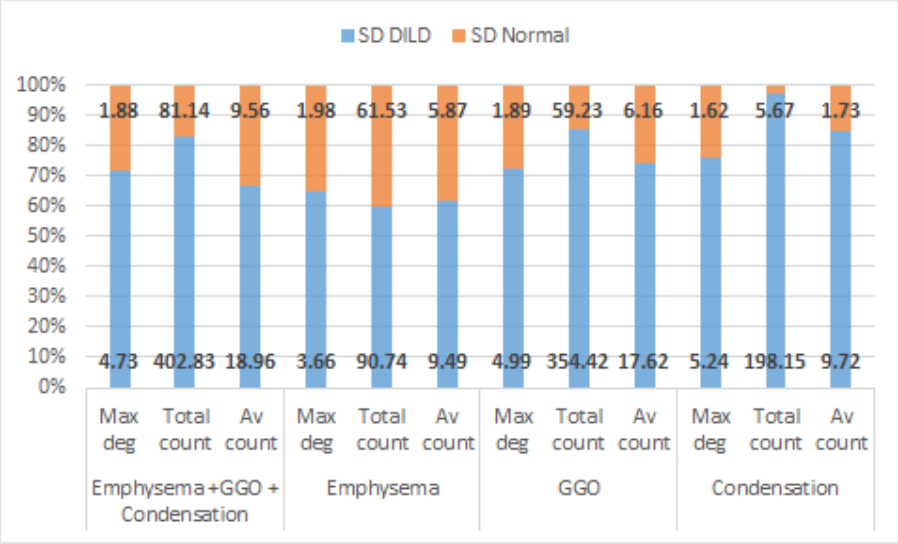


Figure 11. Relative percentage of standard deviation for DILD vs normal lungs on all the pathological HU bands taking into account maximum degree, total count and average degree.
Absolute values are also given on each data point.

In conclusion, these results show that the defined model is valid from a system science perspective, accurately reflecting the underlying process that defines them.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

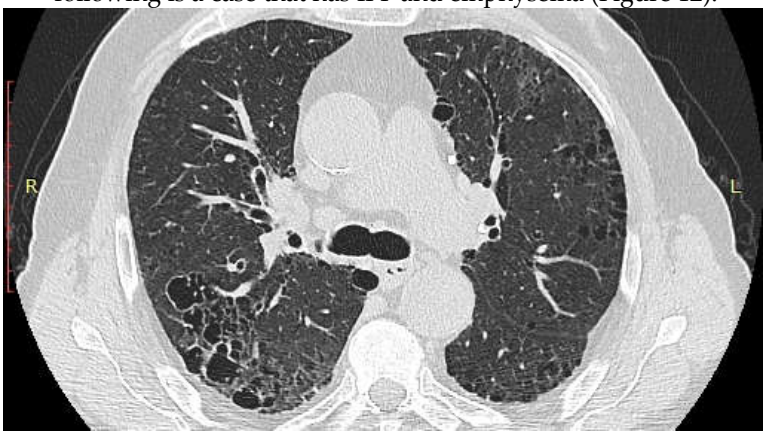
4.2. Medical Science

To properly model the biological system, the presented method should reflect different anatomical and, more importantly, pathological aspects of the lung.

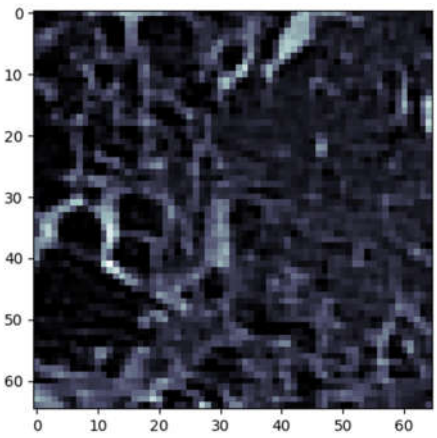
In Figure 9a, only the normal patients are presented, classified according to the epoch in which the exams were taken, as in pre or during the pandemic. There are three post-covid cases NC13, NC14 and NC15 that present higher GGO and consolidation values. Studying their clinical data, NC13 and NC14 are recuperating after severe covid, which would explain their artifacts. NC15, however, has a more special story, i.e.: this investigation was taken before the clinical onset of COVID, when the PCR test result was negative. The patient went on to develop severe COVID, confirmed by a positive PCR test just two days later. In this case, the algorithm did detect an outlier despite the doctor's initial diagnosis. This indicates that such an algorithm might be able to detect early changes in a patient's lung texture and therefore offer the possibility of fast treatment, if the situation warrants it. The clinical data did not show any other outlier in the NC group, as confirmed by our model.

In the pre-covid (NN) group, the outliers may occur due to patient particularities such as: smokers, asthma sufferers or convalescing post-infectious patients. For example, NN06 and NN03 (Figure 9a) are the only two heavy smokers in the normal group, whom the radiology team classified as normal. The model, however, shows them very close to the hypothetical boundary of the normal zone, reinforcing the remark that pathological and non-pathological processes are not a discrete but rather a continuous process. Therefore, the granularity offered by the proposed approach enhances classical CT interpretation and offers details which could easily escape the human eye.

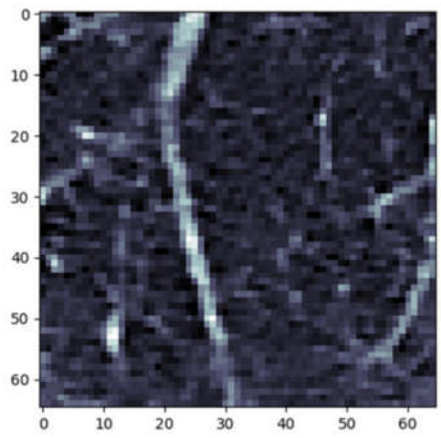
To showcase the fitting of this model onto the pathological process, presented in the following is a case that has IPF and emphysema (Figure 12).



(**a**)



(**b**)



(**c**)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

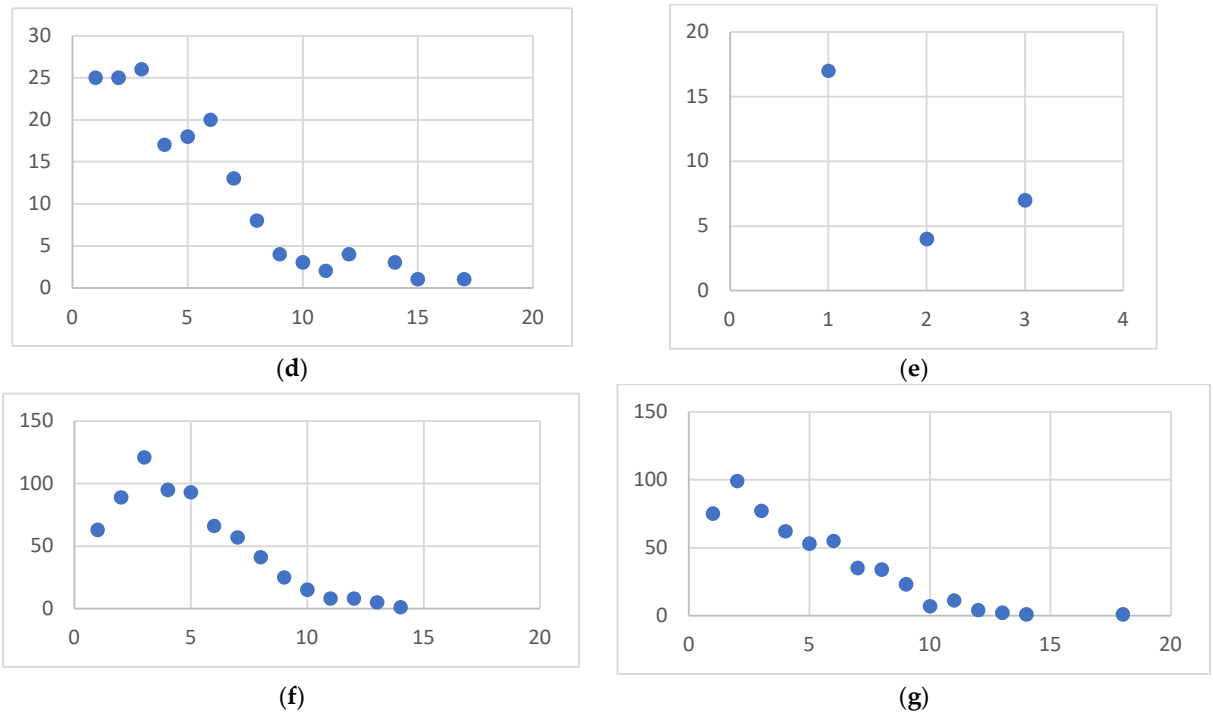


Figure 12. (**a**)HRCT slice under analysis (**b**) Sample 1 (**c**) Sample 2 (**d**) Degree distribution for sample 1 on the emphysema layer (**e**) Degree distribution for sample 2 on the emphysema layer (**f**) Degree distribution for sample 1 on the GGO layer (**g**) Degree distribution for sample 2 on the GGO layer

The emphysema bubble found in sample 1 is clearly reflected in the degree distribution of the same sample. However, both samples present similar inflammation (the GGO layer distributions) showcasing the underlying disease – IPF. The proposed model has successfully dealt with overlapping patterns, in this case .

Regarding the pathological and the normal case distribution presented in Figure 8, there are some cases in which the pathological points are very close to the normal ones. Zooming in, as shown in Figure 9 b, those cases belong to OP (organizing pneumonitis). The OP is the usual manner of reaction to lung lesions, during the healing process, consecutively most likely to a lung infection, but also found after radiation therapy, inhaling injury, neoplasm, drug toxicity. In HRTC evaluation it may have a multitude of appearances, including nodular images, irregular GGO pattern, but most often peripheral bilateral consolidation (atoll sign).[32], [33] Patients that are overlapping the normal lot are (close to being) healed, therefore the model rightfully clustered them with the normal patients.

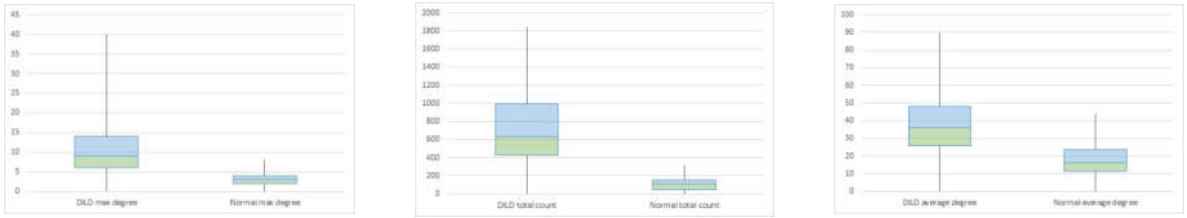


Figure 13. Box plot for DILD (left) VS Normal (right) for complex network parameters of (**a**) maximum degree (**b**) total count (**c**) average degree

Comparing normal lungs with diseased lungs from a statistical perspective, is challenging due to different DILD phenotypes and the relative small lot size/ disease class. To prove the method and model work in an overall manner, a t-test: two-sample assuming unequal variances was conducted comparing normal to DILD samples. The results, summarized in Table 2 and Figure 13, show that measured p is less than 0.05 (3.97×10^{-17},

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2022 doi:10.20944/preprints202207.0156.v1

$8.52x10^{-23}$ and $5.31x10^{-9}$) and observed t (10.49, 14.91 and 6.29) is larger than critical t (1.98, 1.99 and 1.98), therefore rejecting the null hypothesis; i.e. being 95% confident that the differences between groups are not due to chance.

Table 2 Statistical comparison

	Maximum degree		Total count		Average Count	
	DILD	Normal	DILD	Normal	DILD	Normal
Mean	15.96875	7.032258	846.5692	7.1	51.65253	32.53397
Variance	39.45933	3.365591	206084.5	3.334483	362.9068	113.4483
Observations	30	30	30	30	30	30
Hypothesized Mean Difference	0		0		0	
Df	82		64		92	
t Stat	10.49451		14.9084		6.288591	
P(T<=t) one-tail	3.97E-17		8.52E-23		5.31E-09	
t Critical one-tail	1.663649		1.669013		1.661585	
P(T<=t) two-tail	7.93E-17		1.7E-22		1.06E-08	
t Critical two-tail	1.989319		1.99773		1.986086	

A proper comparison between disease phenotypes would require a much larger study in order to be relevant. However, the purpose of this paper was to test if the complex network model can accurately reflect the biological process and the quantitative data agrees. From a qualitative Medical Science perspective, the matter needs further study, yet the results seem encouraging.

4.3. Comparison with other HRCT analysis methods

In this section, this method is compared with existing ones. Assessed against the normal, established way of analyzing the HRCT by human radiologists and doctors, the proposed method is almost too simple. The full medical analysis is not limited to the HRCT, it will almost always require clinical data and more often than not other paraclinical investigations. Regarding the modality, the human analysis uses a difficult to reproduce mixture of analytical and empirical processes ("clinical sense") and its disease progress measurement is mostly subjective [9]–[12]

CAD methods vary from commercial to research ones. The most well known commercial approach, Caliper [21], does not use just HRCT, it also needs a way to calculate lung expansion, like PFT. However, it is a very objective, stand-alone way to measure lung diseases and works remarkably fast. The proposed method is not nearly as fast, as it needs an estimated median 2 minutes/sample for the whole three layers, therefore requiring 242 minutes/full slice and 3872 minutes/patient. The time values are measured on an average PC, running a single threaded program. Amdahl's law indicates that there is room for improvement, with some limitations. This is a downfall and needs work in order to reach full analysis capacity, although the information offered is multifaceted compared with Caliper, due to the complex network methodology.

Research-stemmed approaches, like the one from [22], [34] and any of the ones based on machine learning like [14]–[17] use just the HRCT, but the way they offer a measurement for the disease is inexistent in most cases and volumetric, in others. Most machine learning approaches are oriented towards proper classification and pattern identification and not as a way to quantify it. Also, the time aspect is mostly unspecified for all these approaches, so no assessment can be made.

A summary for all these comparisons is offered in Table 3.

Table 3. Methodology comparison

	Just HRCT	Analytical	Empirical	Works alone	Measurement

 doi:10.20944/preprints202207.0156.v1

Doctor	N	Y	Y ("clinical sense")	Y	Subjective
Caliper [21],	N, PFT	Y	N	Y	Yes, 1 dimensional, size
Zrimec [22], [34]	Y	Y	N	Mostly	Maybe
Machine learning	Y	N	Y	Maybe	Maybe
Proposed model	**Y**	**Y**	**N**	**N**	**Yes, 3 dimensional**

None of the aforementioned approaches offer a way to mathematically characterize affected areas of the lung, unlike the present method. Using network characteristics, it can quantify and qualify a pathological process on three axes, however, it is still unable to work alone yet and needs many more cases to allow for proper classification methods.

5. Conclusions

In this paper, a novel method of using complex networks to transform lung HRCT has been presented. The methodology section delves deeper into the algorithm steps and the justification of each chosen parameter. The sample size is justified by anatomical bounds of the secondary pulmonary lobule; the radius influencing network connectivity is correlated with injury granularity and the Hounsfield units intervals are dependent upon the device and resolution. The results section presents in parallel the processing steps for two sample patients (a normal and a pathological one), as well as a whole-lot perspective. In the discussion section, the correctness of this model is justified from a System Science perspective, by using the degree distributions as the main system characterisation tool. Furthermore, the network measurement clusterization is described, showing that it results in clear disparities between the normal and pathological lots. From a Medical Science perspective it is showcased how the chosen model reflects clinical data and how its low granularity presents an advantage in the diagnosis process. In the end a comparison between this method and other existing ones highlights the advantage that it has: to offer a complex qualitative and quantitative measurement. Pitfalls of the proposed model include its inability to work alone yet and the relatively small lot on which it was tested, which will all need to be addressed in further research.

In conclusion, the stated goal is considered to have been achieved, by showing how a complex network model can be used to transmute lung HRCT into a quantifiable and qualifiable structure that can enhance the DILD diagnosis.

Supplementary Materials: The datasets generated or analyzed within the current research study are available from the corresponding author on reasonable request.

Author Contributions: Conceptualization, L.B., V.M.A. and A.A.T.; methodology, V.M.A.; software, L.B.; validation, D.L.M.; formal analysis, L.B. and V.M.A.; investigation, L.B., A.A.T and V.M.A; resources, E.R.S., D.L.M and C.I.O. ; data curation, E.R.S.; image sample selection: A.A.T, C.I.O, D.L.M, E.R.S; writing—original draft preparation, L.B., A.A.T and V.M.A; writing—review and editing, L.B., A.A.T and V.M.A; visualization, L.B., V.M.A and E.R.S.; supervision, H.C., C.I.O, D.L.M; project administration, C.I.O.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Clinical Hospital for Infectious Diseases and Pneumoftiziology Dr. Victor Babeș Timișoara (protocol code 11835/26.11.2021) for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] J. H. Ryu, C. E. Daniels, T. E. Hartman, and E. S. Yi, "Diagnosis of Interstitial Lung Diseases," *Mayo Clin. Proc.*, vol. 82, no. 8, Art. no. 8, Aug. 2007, doi: 10.4065/82.8.976.

[2] S. A. Guler and T. J. Corte, "Interstitial Lung Disease in 2020: A History of Progress," *Clin. Chest Med.*, vol. 42, no. 2, Art. no. 2, Jun. 2021, doi: 10.1016/j.ccm.2021.03.001.

[3] M. Molina-Molina *et al.*, "Importance of early diagnosis and treatment in idiopathic pulmonary fibrosis," *Expert Rev. Respir. Med.*, vol. 12, no. 7, Art. no. 7, Jul. 2018, doi: 10.1080/17476348.2018.1472580.

[4] M. Kolb *et al.*, "Nintedanib in patients with idiopathic pulmonary fibrosis and preserved lung volume," *Thorax*, vol. 72, no. 4, Art. no. 4, Apr. 2017, doi: 10.1136/thoraxjnl-2016-208710.

[5] K. C. Meyer, "Diagnosis and management of interstitial lung disease," *Transl. Respir. Med.*, vol. 2, p. 4, Feb. 2014, doi: 10.1186/2213-0802-2-4.

[6] D. Manolescu, L. Davidescu, D. Traila, C. Oancea, and V. Tudorache, "The reliability of lung ultrasound in assessment of idiopathic pulmonary fibrosis," *Clin. Interv. Aging*, vol. 13, pp. 437–449, 2018, doi: 10.2147/CIA.S156615.

[7] N. Sverzellati, "Highlights of HRCT imaging in IPF," *Respir. Res.*, vol. 14 Suppl 1, p. S3, 2013, doi: 10.1186/1465-9921-14-S1-S3.

[8] R. M. du Bois, "An earlier and more confident diagnosis of idiopathic pulmonary fibrosis," *Eur. Respir. Rev. Off. J. Eur. Respir. Soc.*, vol. 21, no. 124, Art. no. 124, Jun. 2012, doi: 10.1183/09059180.00000812.

[9] M. Inomata *et al.*, "Clinical impact of the radiological indeterminate for usual interstitial pneumonia pattern on the diagnosis of idiopathic pulmonary fibrosis," *Respir. Investig.*, vol. 59, no. 1, Art. no. 1, Jan. 2021, doi: 10.1016/j.resinv.2020.07.001.

[10] S. L. F. Walsh, L. Calandriello, N. Sverzellati, A. U. Wells, D. M. Hansell, and UIP Observer Consort, "Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT," *Thorax*, vol. 71, no. 1, Art. no. 1, Jan. 2016, doi: 10.1136/thoraxjnl-2015-207252.

[11] S. L. F. Walsh *et al.*, "Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study," *Lancet Respir. Med.*, vol. 4, no. 7, Art. no. 7, Jul. 2016, doi: 10.1016/S2213-2600(16)30033-9.

[12] A. A. Trusculescu, D. Manolescu, E. Tudorache, and C. Oancea, "Deep learning in interstitial lung disease-how long until daily practice," *Eur. Radiol.*, vol. 30, no. 11, Art. no. 11, Nov. 2020, doi: 10.1007/s00330-020-06986-4.

[13] M. S. Crews *et al.*, "Automated CT Analysis of Major Forms of Interstitial Lung Disease," *J. Clin. Med.*, vol. 9, no. 11, Art. no. 11, Nov. 2020, doi: 10.3390/jcm9113776.

[14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ArXiv14091556 Cs*, Apr. 2015, Accessed: Feb. 06, 2022. [Online]. Available: http://arxiv.org/abs/1409.1556

[15] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, Dec. 2014, pp. 844–848. doi: 10.1109/ICARCV.2014.7064414.

[16] S. L. F. Walsh, L. Calandriello, M. Silva, and N. Sverzellati, "Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study," *Lancet Respir. Med.*, vol. 6, no. 11, Art. no. 11, Nov. 2018, doi: 10.1016/S2213-2600(18)30286-8.

[17] Q. Li, W. Cai, and D. D. Feng, "Lung image patch classification with automatic feature learning," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2013, pp. 6079–6082, 2013, doi: 10.1109/EMBC.2013.6610939.

[18] S. L. F. Walsh and M. Kolb, "Radiological diagnosis of interstitial lung disease: is it all about pattern recognition?," *Eur. Respir. J.*, vol. 52, no. 2, Art. no. 2, Aug. 2018, doi: 10.1183/13993003.01321-2018.

[19] "Signs and Patterns of Lung Disease - Chest Radiology: The Essentials, 2nd Edition," Feb. 06, 2022. https://doctorlib.info/medical/chest/2.html (accessed Feb. 06, 2022).

[20] S. Hobbs, J. H. Chung, J. Leb, K. Kaproth-Joslin, and D. A. Lynch, "Practical Imaging Interpretation in Patients Suspected of Having Idiopathic Pulmonary Fibrosis: Official Recommendations from the Radiology Working Group of the Pulmonary Fibrosis Foundation," *Radiol. Cardiothorac. Imaging*, vol. 3, no. 1, p. e200279, Feb. 2021, doi: 10.1148/ryct.2021200279.

[21] A. Chen, R. A. Karwoski, D. S. Gierada, B. J. Bartholmai, and C. W. Koo, "Quantitative CT Analysis of Diffuse Lung Disease," *RadioGraphics*, vol. 40, no. 1, pp. 28–43, Jan. 2020, doi: 10.1148/rg.2020190099.

[22] T. Zrimec and S. Busayarat, "Computer-aided Analysis and Interpretation of HRCT Images of the Lung," 2011. doi: 10.5772/14507.

[23] A. Depeursinge, T. Zrimec, S. Busayarat, and H. Müller, "3D Lung Image Retrieval Using Localized Features," *SPIE Med. Imaging*, Oct. 2011, doi: 10.1117/12.877943.

[24] G. V. L. de Lima, T. R. Castilho, P. H. Bugatti, P. T. M. Saito, and F. M. Lopes, "A Complex Network-Based Approach to the Analysis and Classification of Images," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Cham, 2015, pp. 322–330. doi: 10.1007/978-3-319-25751-8_39.

[25] G. Veloso, T. Castilho, P. Bugatti, P. Saito, and F. Lopes, *A Complex Network-Based Approach to the Analysis and Classification of Images*. 2015. doi: 10.1007/978-3-319-25751-8_39.

[26] Y. Mourchid, M. E. Hassouni, and H. Cherifi, "A General Framework for Complex Network-Based Image Segmentation," *Multimed. Tools Appl.*, vol. 78, no. 14, pp. 20191–20216, Jul. 2019, doi: 10.1007/s11042-019-7304-2.

[27] M. Hiramatsu *et al.*, "Pulmonary ground-glass opacity (GGO) lesions-large size and a history of lung cancer are risk factors for growth," *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer*, vol. 3, no. 11, pp. 1245–1250, Nov. 2008, doi: 10.1097/JTO.0b013e318189f526.

[28] K. M. Smith, "Explaining the emergence of complex networks through log-normal fitness in a Euclidean node similarity space," *Sci. Rep.*, vol. 11, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41598-021-81547-3.

[29] M. Adler, A. Mayo, and U. Alon, "Logarithmic and Power Law Input-Output Relations in Sensory Systems with Fold-Change Detection," *PLOS Comput. Biol.*, vol. 10, no. 8, p. e1003781, Aug. 2014, doi: 10.1371/journal.pcbi.1003781.

[30] Y. Shang, "Degree distribution dynamics for disease spreading with individual awareness," *J. Syst. Sci. Complex.*, vol. 28, Feb. 2015, doi: 10.1007/s11424-014-2186-x.

[31] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Mod. Phys.*, vol. 87, no. 3, pp. 925–979, Aug. 2015, doi: 10.1103/RevModPhys.87.925.

[32] M. Hovinga, R. Sprengers, H.-U. Kauczor, and C. Schaefer-Prokop, "CT Imaging of Interstitial Lung Diseases," *Multidetector-Row CT Thorax*, pp. 105–130, Feb. 2016, doi: 10.1007/978-3-319-30355-0_7.

[33] S. R. Desai, H. Prosch, and J. R. Galvin, "Plain Film and HRCT Diagnosis of Interstitial Lung Disease," in *Diseases of the Chest, Breast, Heart and Vessels 2019-2022: Diagnostic and Interventional Imaging*, J. Hodler, R. A. Kubik-Huch, and G. K. von Schulthess, Eds. Cham (CH): Springer, 2019. Accessed: Feb. 06, 2022. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK553872/

[34] A. Depeursinge, T. Zrimec, S. Busayarat, and H. Müller, "3D Lung Image Retrieval Using Localized Features," *SPIE Med. Imaging*, Oct. 2011, doi: 10.1117/12.877943.