

## TIPARE ÎN BIOINFORMATICĂ

### **Teză de doctorat – Rezumat**

Pentru obținerea titlului științific de doctor la

Universitatea Politehnica Timișoara

În domeniul de doctorat Calculatoare și Tehnologia Informației

**autor ing. Laura BROASCĂ**

conducător științific Prof.Univ.Dr.Ing. Horia CIOCÂRLIE

Februarie 2023

# 1. Abstract

Teza de doctorat este structurată în 6 capitole. **Capitolul 1** prezintă o scurtă introducere în tematica rețelelor complexe și tiparelor regăsite în Bioinformatică. Mai exact, abordează două mari direcții ale studiului tiparelor în domeniul Bioinformaticii, bazate pe rețele complexe: **tipare vizuale** – putând fi observate și analizate cu ajutorul tool-urilor de vizualizare ale rețelelor complexe, dar și **tipare de numerice** – evidențiate prin modelarea, vizualizarea și analiza datelor provenite din domeniul medical al patologiilor pulmonare.

**Capitolul 2** prezintă un studiu asupra rețelelor complexe, noțiuni teoretice, dar și asupra celor mai noi algoritmi și unelte software folosite pentru vizualizarea rețelelor complexe. Acest capitol se încheie prin definirea nișei și motivația dezvoltării unei soluții care să acopere neajunsurile soluțiilor software existente actual specializate pe vizualizarea și modelarea rețelelor complexe.

**Capitolul 3** prezintă soluția tehnică propusă – aplicație software capabilă să transforme sisteme biologice în rețele complexe vizualizate într-o modalitate inovativă – vizualizări în 2 și în 3 dimensiuni în funcție de necesitățile utilizatorului astfel încât să permită evidențierea tiparelor vizuale într-o modalitate cât mai pragmatică și eficientă. Algoritmul este analizat din mai multe puncte de vedere și comparat cu alți algoritmi pentru a evidenția contribuția și îmbunătățirile aduse prin dezvoltarea acestei aplicații.

**Capitolul 4** tratează modelarea tiparelor numerice a datelor provenite din analiza unor seturi de HRCT (High Resolution Computer Tomograph) provenind de la pacienți sănătoși dar și de la pacienți care suferă de Pneumopatii interstițiale difuze (PID). În acest scop, se prezintă un nou algoritm dezvoltat pentru transformarea imaginilor HRCT în rețele complexe care modelează 3 tipuri de țesut pulmonar. Aceste tipuri de țesuturi pot prezenta cheia tratamentului cât mai rapid pentru a eficientiza recuperarea pacientului. Rețelele complexe rezultate din conversia HRCT-urilor sunt analizate din punct de vedere al distribuției pixelilor și sunt evidențiate mai multe tipuri de funcții matematice care fac diferența între pacienții sănătoși și pacienții bolnavi. Se evidențiază astfel potențialul algoritmului în vederea stabilirii unui diagnostic cât mai precis și obiectiv.

**Capitolul 5** se bazează pe algoritmul descris în capitolul 4, și pornind de la această modelare a țesutului pulmonar afectat în cazul PID, se propune o nouă metodă pentru detecția timpurie a modificărilor țesutului pulmonar, dar și o modalitate pentru măsurarea vitezei de progresie a afectării pulmonare. Se validează ipotezele prin intermediul unor T-teste (modalitate statistică de validare) și se demonstrează potențialul acestui tip de abordare în cazul a două tipuri de țesut pulmonar: Ground Glass Opacity și Consolidare.

**Capitolul 6** acoperă concluziile și contribuțiile personale pe cele două direcții propuse: tipare vizuale și tipare de date în domeniul bioinformaticii.

## 2. Obiective

Obiectivele principale ale acestei teze sunt axate pe două mari direcții.

**Tipare vizuale și tool-uri de analiză a tiparelor din Bioinformatică bazate pe rețele complexe:**

1. Analiza toolurilor existente de reprezentare a rețelelor complexe pentru a determina punctele tari și slabe în reprezentarea datelor specifice bioinformaticii.
2. Dezvoltarea unui algoritm bazat pe rețele complexe care să faciliteze detecția de tipare vizuale pentru date specifice bioinformaticii (dimensiuni mari/medii – mai mult de 200 noduri 500 muchii, conectivitate mărită). Algoritmul va putea funcționa în 3 dimensiuni pentru o vizualizare complexă. Numărul de dimensiuni vizualizate

poate fi 2 și / sau 3 în funcție de necesitățile particulare ale biologilor / medicilor / specialiștilor în bioinformatică.

3. Analiza performanței și a calității vizuale a algoritmului dezvoltat comparativ cu celelalte aplicații software existente la momentul actual.

#### **Tipare numerice și metode de identificare a tiparelor numerice în Bioinformatică:**

1. Determinarea unui domeniu din medicină în care se pretează o aplicație de analiză numerică bazată pe rețele complexe
2. Crearea unui algoritm inovator bazat pe rețele complexe pentru procesarea datelor imagistice de tip HRCT specifice plămânilor
3. Validarea acestui algoritm din punct de vedere medical dar și al științei sistemelor
4. Compararea algoritmului cu alți algoritmi existenți
5. Integrarea algoritmului într-o metodologie pluridisciplinară care să conducă la identificarea tiparelor în cadrul Penumopatiilor Interstițiale Difuze (bazat pe cel puțin unul din cele trei tipuri de densitate patologice ale țesutului pulmonar: Emfizem, Ground Glass Opacity, Consolidare)
6. Evidențierea tiparelor de date ce permit diferențierea între pacienții normali (sănătoși) și cei patologici
7. Utilizarea algoritmului propus pentru identificarea precoce a modificărilor patologice ale țesutului pulmonar
8. Propunerea unei metrici pentru a evalua viteza de progresie a afectării pulmonare

### **3. Contribuții personale**

Pentru îndeplinirea obiectivelor menționate am dezvoltat doi algoritmi bazați pe rețele complexe: un algoritm inovativ pentru vizualizarea rețelelor complexe, dar și un algoritm de Computer Aided Diagnosis (diagnostic ajutat de calculator) bazat pe rețele complexe.

Prima parte a tezei (**Capitolul 3**) descrie algoritmul hibrid de vizualizare a rețetelor complexe într-un mod combinat: 2D și 3D pentru a afișa și a sublinia **tipare vizuale** care ar putea altfel să fie slab reprezentate cu ajutorul aplicațiilor software existente.

Astfel, contribuțiile personale constau în:

1. Analiza și evaluarea soluțiilor de ultimă generație existente în domeniu pentru vizualizarea rețelelor complexe. Publicată în: [1][6][8]
2. Am propus și implementat un nou algoritm de vizualizare al rețelelor complexe. [8][10]
3. Am testat performanța algoritmului cu ajutorul mai multor seturi de date biologice de dimensiuni variabile.
4. Am implementat o nouă metrică de covarianță (similitudine) pentru nodurile rețelei și pentru poziționarea lor în layout-ul rețelei [9]
5. Am analizat comparativ algoritmul dezvoltat față de alte aplicații software pentru vizualizarea rețelelor complexe frecvent folosite [9][10]

Referitor la **Capitolele 4 și 5** ale tezei, **analiza tiparelor numerice** printr-o nouă metodă de interpretare a Computer Tomografiilor, am adus următoarele contribuții:

1. Am participat în dezvoltarea unei abordări multidisciplinare în ce privește metoda de modelare a imaginilor HRCT de plămâni, publicată în [23][25]
2. Am propus și implementat algoritmul Python de procesare a HRCT-urilor bazat pe rețele complexe, publicat în [23][25]
3. Am creat și definit un model responsabil de împărțirea pe straturi/benzi a imaginilor DICOM conform celor trei dimensiuni de interes pentru specialiștii pneumologi: Emfizem, Ground Glass Opacities, Consolidare [23][25]

4. Am procesat și pregătit toate seturile de date folosite în această abordare, în urma selecției lotului făcută de partenerii de la Spitalul de boli infecțioase și pneumoftiziologie Victor Babeș Timișoara. [23][25]
5. Am participat în definirea unei dimensiuni a secțiunii/ferestrei de observație (65x65 px) pentru seturile de date analizate, astfel încât să satisfacă două mari criterii: relevanța din punct de vedere medical (acoperirea unui întreg lobul pulmonar secundar) dar și obținerea/menținerea unei performanțe adecvate a algoritmului [23][25]
6. Am făcut o analiză în detaliu asupra distanței radiale, prin efectuarea mai multor teste cu o distanță radială variată, și am analizat impactul și relevanța modificărilor asupra modelului de rețelelor complexe a țesutului pulmonar [23][25]
7. Am definit o metrică de similaritate (delta) între pixeli bazat pe benzi Hounsfield (Emfizem, GGO, Consolidare) și am validat-o cu date medicale, împreună cu echipa de medici specialiști pneumologi [23][25]
8. Am identificat funcțiile matematice care se mapează pe distribuția gradelor nodurilor rețelelor complexe, atât pentru lotul de pacienți normali cât și pentru lotul de pacienți cu plămâni afectați [23][25]
9. Am făcut o analiză pentru a determina acuratețea acestor funcții matematice în cazul plămânilor normali și a celor afectați, dar și măsura în care aceste funcții se conformează cu Știința Sistemelor dar și cu Știința Medicală [23][25]
10. Am validat modelul propus din punct de vedere al Științei rețelelor [23][25]
11. Am validat modelul propus prin prisma a trei metrici ale rețelelor complexe: suma totală a gradelor, gradul mediu, gradul maxim [23][25]
12. Am participat în definirea unui nou tip de măsurători și o formulă matematică pentru viteza de progresie a fibrozei pulmonare. Bazat pe noțiunea clasică de *viteză* (definită ca variație pe unitate de timp), am propus o nouă formulă a vitezei relative de variație pentru fibroză. [25]

## 4. Rezumatul tezei pe capitole

### 1.1. Introducere

Scopul tezei constă în rezolvarea unor probleme întâlnite în domeniul Bioinformaticii, mai exact aducerea unor contribuții (algoritmi / software), bazate pe rețele complexe, care să ajute la descoperirea tiparelor în Bioinformatică pe două mari direcții: **tipare vizuale** și **tipare numerice**.

În acest sens, a fost prezentat contextul actual pentru cele două categorii.

Seturile de date aparținând rețelelor biologice sau genetice (ADN) au câteva caracteristici care complică și îngreunează descoperirile pe această nișă: seturile de date sunt foarte mari și dense ca dimensiuni, însă totodată sunt și incomplete. Acest lucru a condus la nevoia specialiștilor de a folosi unelte sau aplicații software prin care să poată genera vizualizări grafice cât mai fidele ale acestor sisteme, prin care să poată observa cu claritate structura lor, comunitățile, dar și legăturile sau interdependențele între elementele acestora.

În ce privește **tiparele vizuale** în domeniul geneticii și instrumentele software care există actualmente și care pot genera vizualizări ale diferitelor tipuri de rețele complexe, acestea sunt mai multe: Gephi, NetworkX (Python), Pajek, Igraph, etc.

Însă deși toate acestea pot genera reprezentări satisfăcătoare ale rețelelor complexe, ele prezintă diferite dezavantaje fie pe partea de performanță, fie fac compromisuri în ce privește aspectul vizual. Însă mai ales în domeniul biologiei sau al medicinei, necesitatea pentru redarea unor rețele complexe foarte dense constând în mii de noduri și mii sau zeci de mii de legături este una cu atât mai stringentă cu cât, de exemplu în cazul rețelelor ADN, ele ar putea ajuta la

detectarea unor legături încă nedescoperite între genele umane. Ca atare, compromisurile făcute la partea grafică sunt un dezavantaj pentru specialiștii care caută detalii vizuale. Pe de altă parte, cu cât reprezentarea vizuală este mai detaliată și mai încărcată, cu atât necesită mai multe resurse de procesare, iar acest lucru conduce la o mai slabă performanță și un timp mai lung de generare.

Astfel, în urma colaborării cu specialiști în domeniul geneticii care doreau studierea anomaliilor geneticii apărute încă din fază embrionară prin rețele complexe, a apărut motivația de a dezvolta un algoritm care să poată genera vizualizări cât mai clare ale acestor rețele într-o modalitate inovativă: într-o manieră combinată 2D și 3D.

În aspectul **tiparelor numerice** – nișa pe care o propune teza este analiza datelor provenite din HRCT-uri pulmonare prin intermediul reprezentării și modelării lor sub formă de rețele complexe și detectarea unor tipare matematice care pot caracteriza și diferenția pacienții normali (sănătoși) de pacienții care suferă de pneumopatii interstițiale difuze (PID).

Nevoia specialiștilor pneumologi este aceea de a detecta cât mai rapid modificări ale țesutului pulmonar astfel încât să poată interveni și trata cât mai eficient patologiile, dat fiind că unele dintre acestea pot avea o evoluție virulentă și pot cauza decesul într-un timp foarte scurt. Diagnosticul medical al acestor PID este predominant bazat pe imagistica medicală, care poate oferi o reprezentare clară a plămânilor, dar și pe alți parametri și teste medicale. Însă în ce privește interpretarea HRCT-urilor, acest proces poate fi unul subiectiv, dat fiind că ochiul uman poate omite detalii de dimensiuni foarte mici (câțiva pixeli).

Există actualmente câteva aplicații software capabile să identifice tipare prin analiza HRCT-urilor, însă ele sunt fie încă într-o stare incipientă (Zrimec et al), fie sunt prea complexe, ori costisitoare din punct de vedere financiar (CALIPER). În acest scop, teza propune o nouă modalitate bazată pe analiza tiparelor numerice observate în HRCT-uri, care se poate dovedi de ajutor pentru consolidarea diagnosticelor date de medici în cazul PID.

## **1.2. Noțiuni teoretice**

Capitolul 2 cuprinde o trecere în revistă a celor mai importante noțiuni teoretice în domeniul rețelelor complexe și a algoritmilor software care au ca scop identificarea tiparelor vizuale.

În aria rețelelor complexe, se prezintă câteva noțiuni fundamentale pentru înțelegerea lucrării de față. Astfel, sunt enumerate noțiunile de bază care stau la baza caracterizării unei rețele complexe: distribuția gradelor, coeficientul de clusterizare, modularitatea sau densitatea unei rețele. Totodată prezintă o clasificare a metricilor celor mai folosite, asociate cu rețelele complexe (degree centrality, betweenness centrality, closeness centrality), dar și a tipurilor de modele de rețele: modelul Albert Barabasi, modelul Erdos-Renyi, modelul Watts-Strogatz. În ce privește algoritmii de clusterizare, sunt enumerați câțiva dintre cei mai folosiți algoritmi, și anume Algoritmii Louvain, algoritmul Leiden, sau K-means clustering.

Ca algoritmi state-of-the-art pe această nișă a reprezentării rețelelor complexe, sunt prezentate câteva dintre cele mai folosite, ca de exemplu: Force Atlas 2, Fruchterman-Reingold, OpenOrd, Yifan-Hu, Kamada-Kawai. În ce privește aplicațiile care implementează acești algoritmi, printre acestea se numără: Gephi (acesta implementează toți algoritmii enumerați anterior), NetworkX (dezvoltat în Python), iGraph (dezvoltat în limbajul R), Cytoscape, Nodetrix.

În finalul capitolului 2 se definește nișa abordată în această teză. Se face o comparație (din punct de vedere al performanței dar și dpdv vizual) între aplicațiile de vizualizare

enumerat anterior, și se identifică direcțiile de îmbunătățire dar și criteriile pe care un nou algoritm dezvoltat de mine ar trebui să le acopere: din punct de vedere vizual – algoritmul trebuie să ofere o reprezentare clară a structurii rețelei complexe, cu posibilitatea identificării componentelor, dar și a elementelor singulare, care să poată procesa rețele de dimensiuni medii și mari, care să poată funcționa într-un timp rezonabil, ideal mai bun decât al celorlalte aplicații.

### 1.3. Tipare vizuale în Bioinformatică

#### Algoritmul hibrid de vizualizare a rețelelor complexe în 2D și 3D

Capitolul 3 prezintă un nou algoritm bazat pe rețele complexe pentru evidențierea tiparelor vizuale, aplicat pe rețele biologice sau genetice. Acest algoritm a fost dezvoltat în Octave, o versiune gratuită a celebrului MatLab.

Se prezintă mai întâi criteriile care pot influența generarea grafică a rețelelor complexe și pe care dorește să le îndeplinească acest algoritm:

- Reprezentare structurală relevantă și clară,
- Grafică și coloristică atractivă ochiului uman
- Consum de resurse moderat care să permită folosirea pe scară largă
- Timpul de execuție trebuie să fie unul acceptabil chiar dacă acesta depinde de dimensiunile setului de date

Este prezentat aici conceptul algoritmului propus: generarea unor vizualizări grafice ale rețelelor complexe (din domeniul biologiei) în care comunitățile să fie distribuite într-un spațiu **3D**, însă, în funcție de necesitățile utilizatorului, componentele să poată fi afișate atât în **2D** cât și în **3D**, în mod diferențiat.

O altă abordare inovativă constă în introducerea unui **sistem de referință gradat** (axele  $xOyz$ ) care să ofere un indicator vizual în ce privește distanțele dintre elementele rețelei, dimensiunile comunităților, și care să transmită mult mai precis ideea de ierarhie între componente.

Structura vizualizării se dorește a fi ușor de înțeles, să reducă suprapunerile între muchiile grafului iar comunitățile să fie bine delimitate.

Prima versiune publicată a algoritmului se bazează pe distribuția întregului graf peste un paraboloid compus din cercuri concentrice, în care nodurile rețelei să fie dispersate de-a lungul cercurilor. Ierarhia se bazează pe gradul nodurilor, iar acesta va fi asociat cu raza cercurilor. Nodurile cu un grad mai mare de conectivitate vor fi poziționate mai sus pe axa  $Oz$ , în timp ce nodurile cu un grad mai mic vor fi plasate pe cercurile aflate mai jos. Cu toate acestea, neajunsul primei versiuni a algoritmului este acela că este dificilă gruparea nodurilor în comunități distincte vizual, și de asemenea figura rezultată este una aglomerată. Ca atare acest tip de vizualizare nu îndeplinește criteriile menționate inițial.

În consecință, am dezvoltat o versiune îmbunătățită a algoritmului bazată pe următoarele reguli:

- Fiecare comunitate va fi distribuită pe propriul paraboloid într-un spațiu 3D.
- În cadrul fiecărei comunități, nodurile vor fi plasate pe cercuri concentrice după regula: nodurile cu gradul cel mai mare sunt plasate spre vârful paraboloidului (poziția nodului pe axa  $Oz$  este dată de gradul nodului, la fel și raza cercului pe care va fi el așezat)
- Nodul va fi orientat ca poziționare înspre comunitățile externe cu care este cel mai interconectat – poziționare bazată pe centrul de greutate ponderat al celorlalte comunități cu care este conectat.

- În interiorul propriei comunități, nodul va fi plasat mai aproape de alte noduri cu care are o similaritate mare (similaritatea între două noduri este dată de proporția de vecini pe care o au în comun)

Structura algoritmului are 3 faze:

- Pregătirea setului de date: fișiere cu rețele complexe din domeniul biologiei împărțite în comunități prin algoritmul Louvain (liste de muchii)
- Calculul structurii grafului și determinarea poziționării spațiale a nodurilor și comunităților
- Generarea vizualizării bazat pe structura spațială stabilită anterior

### Împărțirea algoritmului pe pași

Mai întâi algoritmul execută parsarea fișierelor cu listele de muchii și stochează sau calculează informații despre fiecare nod: conectivitate, grad, comunitate din care face parte.

Se creează apoi în memorie o matrice de adiacență care descrie graful.

Algoritmul calculează apoi covarianța nodurilor (criteriul care dă gradul de similitudine) bazat pe matricea de adiacență și îl stochează.

Poziționarea 3D a nodurilor în cadrul unei comunități este calculată în funcție de numărul de conexiuni al nodului către comunitățile externe, centrul de greutate ponderat al comunităților cu care este conectat nodul, conexiunile și similaritatea nodului cu alte noduri din cadrul aceluiași cluster.

După calcularea celor două poziții potențiale pentru plasarea nodului pe unul din cercurile concentrice care formează paraboloidul-suport pentru comunitatea nodului curent, se ajustează poziția proporțional cu coexpresia dintre nodul curent și nodul anterior plasat – pentru a nu crea o aglomerare de noduri în aceeași arie 3D.

După calculul pozițiilor tuturor nodurilor din rețea, se generează o vizualizare cu nodurile și muchiile precalculate. Utilizatorul are posibilitatea de a specifica dacă dorește ca fiecare dintre comunități să fie afișată în 3D, sau doar una dintre ele să fie în 3D în timp ce comunitățile adiacente să fie plotate în 2D. De asemenea, există posibilitatea afișării muchiilor din interiorul clusterelor în 3D, păstrând muchiile dintre comunități în două dimensiuni pentru simplificarea imaginii.

### Rezultate

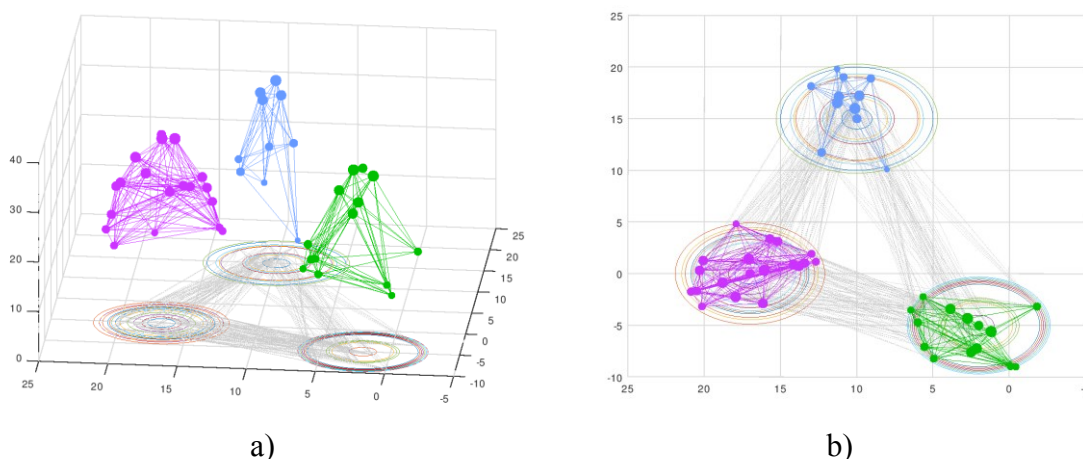


Fig 1. Vizualizare 3D – comunități afișate în 3D, muchii inter-cluster 2D  
a) Privire în 3D b) Privire de sus asupra aceleiași rețele

## Discuție

Algoritmul propus se dorește a fi o alternativă îmbunătățită la aplicațiile existente în domeniu. Comparativ cu Gephi sau cu NetworkX (Python) – avantajele constau în claritatea vizuală a structurii rețelei, datorită distribuției separate a comunităților și eliminarea unor elemente vizuale care pot aglomera imaginea finală (muchii inter-cluster sunt păstrate în 2D).

Există avantajul unui **sistem de referință vizual** (gradat), care nu apare la niciunul din celelalte aplicații menționate (Gephi, NetworkX, iGraph, R, etc) și care oferă informații prețioase utilizatorului despre raportul dintre elementele grafului, dar și despre ierarhia sau dimensiunea lor. Acesta dă rețelei un adevărat aspect 3D, spre deosebire de celelalte tool-uri software. Graficul poate fi rotit de-a lungul celor 3 axe și mărit pentru o mai bună vizualizare.

Comparativ cu ceilalți algoritmi, din punct de vedere al performanței, rezultatele algoritmului hibrid se observă că sunt mai bune în unele aspecte. Timpul de rulare este proporțional cu dimensiunea rețelei, însă acolo unde **densitatea** rețelei este foarte **mare**, **algoritmul** propus oferă **timpi de execuție mai buni**.

Am comparat algoritmi din mai multe puncte de vedere: aspect vizual, viteză, relevanță, distribuție a nodurilor, existența unui sistem de referință clar, opțiunea integrării unor metrici definite de utilizator, adaptabilitate, capacitate de integrare a algoritmilor de machine learning, user friendliness, interactivitate – iar rezultatul cumulat al punctajului pe aceste criterii a demonstrat că algoritmul hibrid este într-adevăr o alternativă viabilă, dacă nu chiar mai bună la aplicațiile software existente, depinzând desigur de nevoile utilizatorului.

### 1.4. Modelarea tiparelor numerice

#### O nouă metodă pentru interpretarea imaginilor CT

Capitolul 4 prezintă o scurtă introducere despre Patologiile interstițiale difuze (PID) și provocările medicilor în a diagnostica și a trata rapid aceste tipuri de boli. Diagnosticul medical în aceste cazuri se bazează foarte mult pe imagistica medicală ca radiografiile sau HRCT-uri. Însă acest proces este unul inerent subiectiv întrucât identificarea zonelor afectate din plămân se bazează pe observația vizuală a radiologilor, iar detaliile pot face o diferență majoră.

Pentru a ajuta la îmbunătățirea diagnosticului, exista la momentul actual o serie de aplicații software care pot analiza și interpreta HRCT-uri medicale pulmonare. Unele dintre ele sunt încă într-o fază incipientă (Zrimec et al.), pe când altele necesită o suită de alte teste pulmonare pentru a putea genera analiza unui CT (CALIPER) iar utilizarea lor necesită o licență plătită. Un alt dezavantaj al acestor soluții este că nu pot cuantifica gradul sau extinderea afectării pulmonare, ci fac doar o clasificare a diagnosticului, fără a oferi nuanțe.

Date fiind provocările și nevoia specialiștilor de soluții mai cuprinzătoare pentru suportul diagnosticului medical în cazul PID, am propus și implementat un algoritm care să transpună porțiuni din imaginea plămânului (preluată dintr-un CT) într-o rețea complexă și să o analizeze pentru a evidenția tipare în tipurile de țesut pulmonar. Imaginea este mai întâi împărțită în 3 straturi, în funcție de tipul de țesut: Emfizem, Ground Glass Opacity, Consolidare.

Rezultatul obținut sunt 3 imagini separate care conțin doar pixelii aparținând unuia dintre cele 3 tipuri de densități pulmonare. Aceste imagini sunt apoi transformate în rețele complexe în care pixelii sunt echivalentul nodurilor, iar conexiunile dintre ei sunt date de gradul de asemănare. În urma obținerii celor 3 rețele complexe, acestea sunt analizate din punct de vedere al distribuției gradelor nodurilor și se observă tipurile de funcții matematice care se mapează cel mai bine pe cele două categorii de pacienți: pacienți cu plămâni normali (sănătoși) și pacienți cu patologii aparținând PID.



## **Privire de ansamblu**

Sunt prezentate aici pe scurt tiparele medicale pe care se bazează diagnosticul medical, în termeni de distribuție spațială în plămân, dar și din punct de vedere al poziționării în lobulul pulmonar secundar. Distribuția diverselor tipuri de țesut pulmonar cu densități diferite, întinderea lor și modul în care se combină oferă indicii legate de tipul patologiei.

HRCT-ul este modalitatea prin care specialiștii pot analiza vizual aceste tipare. Acesta e format din multiple “felii foarte subțiri”/radiografii ale corpului, suprapuse la o distanță de cel mult 1.25 mm. Aceste imagini sunt apoi procesate de către algoritmi de reconstrucție pentru a produce o prezentare detaliată și precisă a unei părți a corpului.

## **Formatul DICOM pentru stocarea digitală a HRCT-urilor**

Standardul DICOM este un standard internațional folosit în mod uzual pentru reprezentarea și stocarea digitală a imaginilor HRCT. Aparatura medicală imagistică (RMN, radiografii, HRCT, ultrasunete) preia informațiile despre pacient și le convertește și stochează în acest format. Fișierele de acest fel au o extensie *.dcm*. Acesta este format dintr-un două secțiuni: un antet și conținutul efectiv al fișierului. Antetul stochează metadate despre parametrii aparatului cu care a fost preluată proba, date despre pacient, date despre rezoluția imaginii, sau date despre poziționarea pacientului în timpul procedurii. Aceste date sunt necesare unui software care citește imaginile pentru a putea interpreta corect restul conținutului.

Conținutul efectiv al fișierului constă într-o înșisuire binară care descrie pixelii imaginii și intensitatea lor.

## **Seturile de date**

Seturile de date au fost selectate de către echipa de specialiști a spitalului de Boli infecțioase și Pneumoftiziologie Victor Babeș din Timișoara. Numărul total al pacienților a fost de 60. Aceste CT-uri au fost împărțite în două categorii: 30 de cazuri de plămâni sănătoși (grupul de control) și 30 de cazuri de pacienți cu plămâni suferind de PID.

Criteriile de includere ale pacienților în acest studiu includ condiția ca un pacient să fi fost diagnosticat de cel puțin 3 specialiști pneumologi cu o experiență de minim 5 ani în PID, fiecare CT trebuie să aibă aceleași caracteristici și calitate stabilite pentru tot lotul, toți pacienții trebuie să aibă investigații imagistice cu o frecvență anuală, toți trebuie să aibă în portofoliu o suită de suită de investigații precum DLco, FEV sau detalii ale evaluării clinice și rezultatul lor, toate CT-urile trebuie să fi fost adnotate de către experți medicali cu indicații și descriere a zonelor de plămân afectate.

Printre criteriile de excludere se numără lipsa consimțământului scris pentru folosirea în cadrul studiilor științifice, lipsa investigațiilor recurente anuale, exemple de imagini CT cu o calitate sub cea stabilită pentru acest studiu, prezența unor alte patologii asociate.

## **Privire de ansamblu asupra algoritmului**

Pașii algoritmului implementat pentru procesarea CT-urilor constau în:

1. Pregătirea și selecția seturilor de date
2. Preprocesarea CT-urilor prin selecția câte unul sample de plămân afectat (65 x 65 px) pentru fiecare pacient
3. Algoritmul convertește pixelii fiecărei sample în echivalentul pe scara Hounsfield:
  - a. Fiecare sample este tratată ca o matrice de pixeli

- b. Se convertește gradientul fiecărui pixel în echivalentul său Hounsfield (HU)
  - c. Se copiază valorile Hounsfield în matrici nou create pentru fiecare tip de țesut afectat: Emfizem, Ground Glass Opacity, Consolidare.
4. Fiecare dintre cele 3 matrici noi sunt convertite într-o matrice de adiacență corespunzătoare unei rețele complexe echivalente după regulile:
    - a. Fiecare pixel este un nod din rețea
    - b. Muchia dintre noduri este creată doar dacă:
      - i. Distanța radială dintre cei doi pixeli este  $rd \leq 4$
      - ii. Diferența de gradient Hounsfield între cei doi pixeli să fie  $\Delta \leq 50$
      - iii. Valorile HU dintre doi pixeli aparțin aceleiași benzi Hounsfield (aceluiși tip de țesut)
  5. Cele 3 matrici împreună cu matricea care le înglobează pe toate sunt analizate din perspectivă matematică, definind funcții care pot aproxima distribuția gradelor nodurilor din rețelele rezultate.
  6. Procedul este executat pentru toate CT-urile din cele două seturi: pacienți normali și pacienți cu PID
  7. Funcțiile matematice obținute pentru cele două tipuri de plămâni sunt comparate pentru a determina similarități și deosebiri.

### Definirea sample-ului

Dimensiunea aleasă pentru sample-urile analizate a fost aceea de 65x65 pixeli. În urma analizei studiilor existente în domeniu, și a discuțiilor avute cu specialiști pneumologi, s-a decis această dimensiune care corespunde dezideratului lor din mai multe puncte de vedere: această arie ar trebui să fie destul de mare încât să cuprindă unitatea funcțională de bază a plămânului (LPS - lobulul pulmonar secundar) care are o dimensiune între [1 cm; 2.5 cm]. Conversia acestei dimensiuni în termeni de pixeli se face printr-o formulă care depinde de spațierea pixelilor (PS). Pentru acest experiment,  $PS = 0.74$ , astfel că printr-un calcul scurt, dimensiunea minimă a sample-ului este de 33.7837 px. Însă acest lucru ne-garantând încadrarea perfectă a unui LPS, am decis o valoare aproape dublă pentru dimensiunile sample-ului: 65 x 65 px.

### Distanța radială

Definim distanța radială (rd) drept distanța maximă (distanța liniară euclidiană) la care se pot afla două noduri pentru a putea fi considerate conectate. În urma unui experiment în care am evaluat structura rețelelor complexe rezultate bazate pe distanțe radiale de la 1 la 8, am observat că o  $rd = 4$  este distanța ideală cu ajutorul căreia algoritmul generează rețele complexe care pot reda cel mai fidel imaginea inițială. Valoarea aleasă de  $rd = 4$  corespunde de asemenea și altor studii în domeniu care se referă la leziunea minimă detectabilă ca având dimensiuni de 3-17 mm. Într-adevăr, convertind din pixeli în mărime fizică  $4 \times 0.74 = 2.96$  mm – este dimensiunea minimă a leziunii pe care o poate reda algoritmul propus, iar aceasta se conformează cu studiile existente.

### Benzile Hounsfield

Hounsfield Units sunt unitățile de bază pentru cuantificarea absorbției și atenuării razelor X de către țesutul uman. Specialiștii fac uz de aceste unități atunci când vor să confirme anumite suspiciuni de diagnostic care ar fi altfel dificil de evaluat doar vizual. Spectrul de valori HU variază între -1000 HU în cazul aerului și 3000 HU în cazul metalelor dure ca oțelul. Pentru ochiul uman, aceste valori sunt asociate cu o anumită nuanță de gri (pe o scară de la alb la

negru). Pentru acest studiu au fost alese trei benzi HU (trei intervale) care corespund la trei tipuri de leziuni pulmonare: Emfizem [-1024, -977), Ground Glass Opacity [-703, -368) și Consolidare [-100, 5). Un plămân afectat poate avea o combinație a acestor leziuni, în proporții diferite.

### Similaritatea nodurilor bazată pe diferența de gradient

Una dintre condițiile ca doi pixeli să fie considerate conectați este ca diferența de gradient între ei să fie de maxim 50 de unități. Această valoare a fost aleasă bazându-se pe două criterii: plaja de valori a unor benzi HU fiind foarte largă, aceasta necesită împărțirea pe sub-intervale pentru a ajuta la o diferențiere mai bună a pixelilor chiar și în cadrul aceleiași benzi. În al doilea rând, din punct de vedere biologic și medical, două puncte o diferență de gradient mai mare decât 50 nu reprezintă același tip de țesut, sau același grad de afectare pulmonară, astfel că acest tip de diferențiere este necesară.

Formula după care gradientul unui pixel este convertit în valoare Hounsfield este una dependentă de valorile aparatului medical:

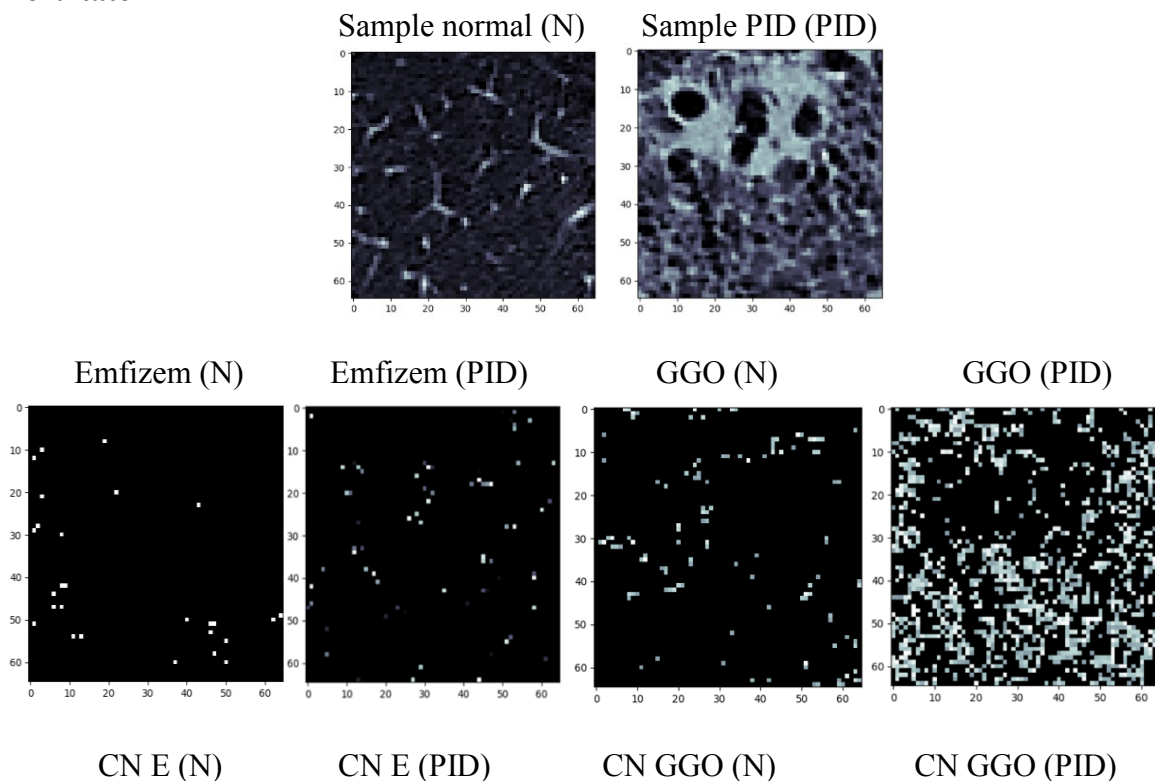
$$\text{pixel\_hu\_value} = \text{pixel\_value} * \text{RescaleSlope} + \text{RescaleIntercept}$$

unde RescaleSlope și RescaleIntercept sunt parametri stocați în antetul fișierului DICOM.

### Conversia imaginilor cu benzi HU în rețele complexe

Algoritmul creează câte o matrice de adiacență pentru cele 3 benzi HU alese (Emfizem, GGO, Consolidare) pe baza a două condiții: ca două noduri să fie conectate, ele trebuie să se afle la o distanță de maxim 4 pixeli, iar diferența de gradient între ele trebuie să fie de maxim 50 de unități.

### Rezultate



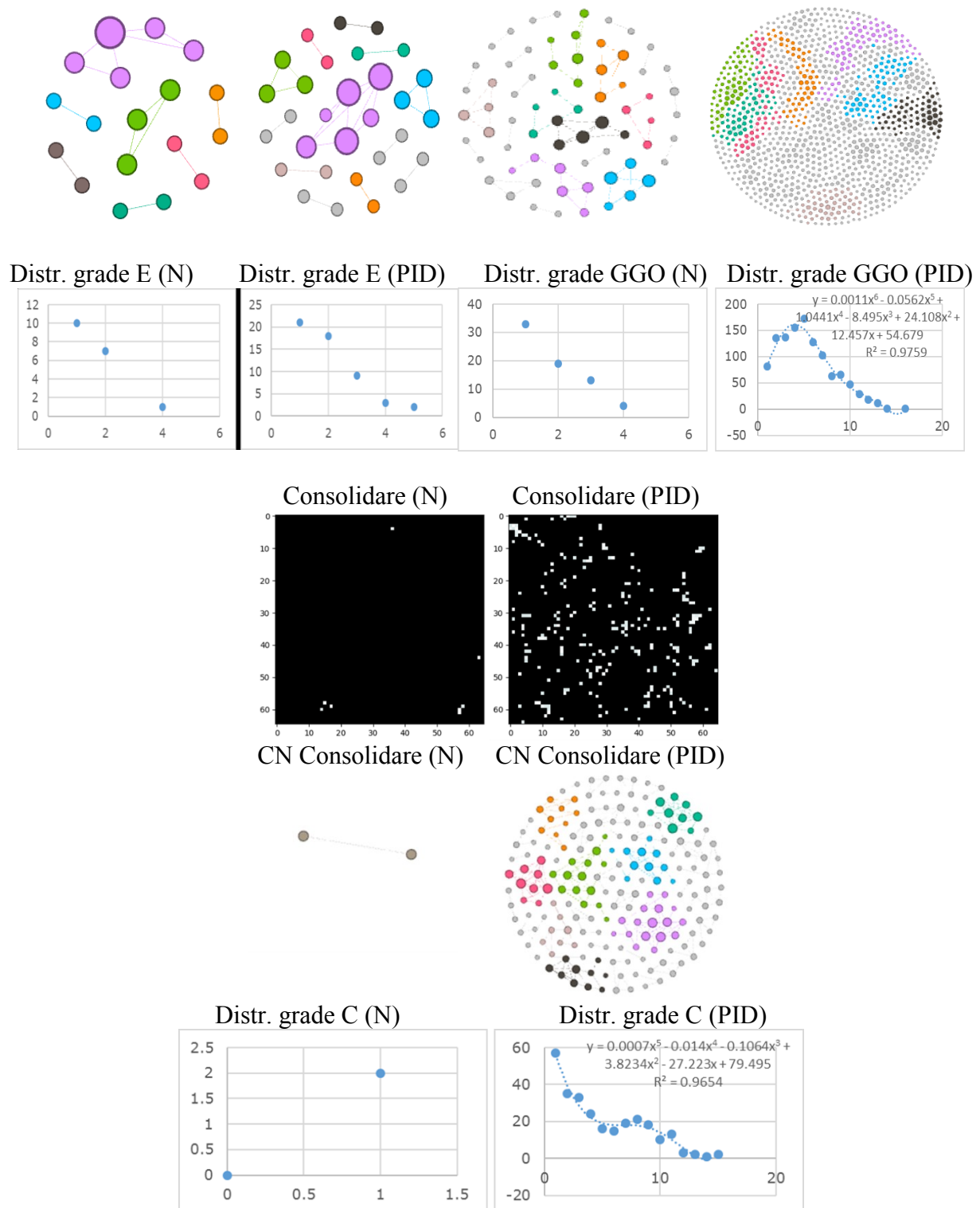


Fig. 2 Conversia unui DICOM sample în rețele complexe pe cele trei dimensiuni E, GGO, C

N – normal, P – Pneumopatii Interstițiale Difuze, E – Emfizem, GGO – Ground Glass Opacity, C – Consolidare.

## Discuții

În urma rezultatelor, **funcția care aproximează plămânii normali** s-a dovedit a fi cea **logaritmică**. Cercetări anterioare în domeniu arată că într-adevăr aceste tipuri de ecosisteme se

încadrează în categoria funcțiilor logaritmice sau funcțiilor de power law. În acest caz însă, fiind vorba de un sistem biologic fără feedback loop, ele se potrivesc într-adevăr cu o funcție logaritmică.

În ce privește **plămânii bolnavi**, distribuția gradelor pentru aceștia se potrivește cel mai bine cu o **funcție polinomială**. Studii anterioare din literatură confirmă acest tip de funcție dar ea este validată și de caracteristica patologiilor PID de a fi procese proliferative, în pofida faptului că nu sunt cauzate de un virus. De menționat aici că rezultatele nu arată un singur tip de funcție polinomială care poate caracteriza pe toate cazurile de PID, ca atare nu se poate trage o concluzie legată de gradul maxim al funcției. Acest lucru poate fi explicat prin faptul că patologiile PID sunt o combinație ale celor 3 benzi HU, în proporții diferite, iar acest lucru necesită un studiu mai detaliat.

Este prezentată în cadrul tezei și validarea din punct de vedere medical al acestor rezultate, dat fiind că întreg studiul a fost dezvoltat în colaborare cu echipa de medici specialiști din cadrul Spitalului de boli infecțioase și pneumoftiziologie Victor Babeș din Timișoara.

În continuare teza mai cuprinde și o validare din punct de vedere statistic a rezultatelor algoritmului și analizei prin calcularea unui T-test aplicat pe cele două serii de pacienți (normali și PID). Acest T-test confirmă că diferențele observate dintre cele două seturi de date nu sunt întâmplătoare.

Comparativ cu alte metode de analiză și evaluare a CT-urilor (Zrimec et al, CALIPER, machine learning), metoda bazată pe rețele complexe propusă aici se dovedește a avea avantaje din mai multe puncte de vedere, ca de exemplu este o metodă care acoperă 3 dimensiuni (E, GGO, C), este o metodă analitică, și oferă o evaluare cantitativă a leziunilor pulmonare. Printre dezavantaje, de menționat necesitatea unei interfețe grafice prietenoasă cu utilizatorul sau dependența de un set de date pre-procesat cu ajutorul radiologilor.

## 1.5. Analiza tiparelor numerice

### Îmbunătățirea diagnosticului imagistic al Patologiilor Interstițiale Difuze

Algoritmul propus în capitolul 4 este folosit în continuare pentru a furniza și mai multe indicii necesare unui diagnostic medical cu o acuratețe sporită. Detectia timpurie este unul dintre nevoile stringente în patologiile pulmonare PID. Actualmente însă, specialiștii evaluează vizual CT-uri succesive ale aceluiași pacient pentru a detecta progresia acestor patologii, însă și acest proces este unul subiectiv, și constă în observarea unor detalii de o granularitate care poate depăși capacitatea ochiului uman. Pentru un diagnostic cât mai corect, medicii se bazează în plus și pe alte teste pulmonare.

Diagnosticul ajutat de calculator (computer-aided diagnosis) este unul în curs de dezvoltare în care se încearcă integrarea diverselor tehnici și algoritmi computaționali pentru obținerea unor diagnostice cât mai precise. Dintre aplicațiile existente actual, CALIPER este capabil să analizeze câte un singur HRCT însă nu poate corobora rezultatele mai multora astfel încât să ofere o măsură a progresiei bolii.

Acest neajuns deschide o nișă neexploatăată pentru abordări precum pattern matching sau rețele complexe.

Cele două mari puncte pe care algoritmul descris aici dorește să le rezolve sunt:

1. Evaluarea cantitativă a evoluției PID cu ajutorul unui model bazat pe rețele complexe
2. Aducerea unei contribuții pentru ajutorul detecției timpurii al patologiilor de acest tip

Lotul de subiecți analizați constă din 96 de HRCT-uri împărțite în: 31 pacienți normali, 65 pacienți cu o serie de 2, 3 sau chiar 4 CT-uri luate la momente diferite (ani diferiți). Toți pacienții au fost anterior evaluați de medici specialiști și prezentau semne ale mai multor PID:

sarcoidoză (S), Fibroză pulmonară idiopatică (IPF), Pneumonie interstițială non-specifică (NSIP). Aceste patologii sunt o combinație între cele 3 tipuri de țesut pulmonar analizat de către algoritmul propus: Emfizem (E), Ground Glass Opacities (GGO), Consolidare (C)

În continuare în teză sunt descrise specificitățile din punct de vedere medical al celor trei tipuri de țesuturi și cum se prezintă acestea vizual într-un HRCT.

Seturile de date au fost procesate prin algoritmul descris în **Capitolul 4**, urmând aceiași pași:

1. Se preiau secțiuni de 65x65 px din zonele de plămâni de interes pentru patologiele alese
2. Pixelii sunt convertiți în echivalentul Hounsfield Units
3. Se separă imaginea în cele 3 straturi corespunzătoare tipurilor de leziuni E, GGO, C
4. Imaginile rezultate sunt convertite în matrici de adiacență și rețele complexe
5. Sunt analizate specificitățile conform cu metricile propuse

### Metrici relevante

Pentru analiza progresiei patologiilor selectate, rețelele complexe rezultate trebuie să respecte specificitățile din punct de vedere biologic ale plămânilor, astfel încât să descrie cât mai fidel evoluția, forma și densitatea țesutului afectat. În consecință am ales următoarele metrici ale rețelelor complexe pentru evaluare: gradul maxim, gradul total, gradul mediu.

Aceste metrici reprezintă din punct de vedere medical următoarele trăsături: gradul total – afectarea cumulată asupra întregii secțiuni de plămân, gradul mediu sugerează localizarea leziunilor, în timp ce gradul maxim arată intensitatea maximă a unei leziuni.

### Progresia leziunii

Pentru evaluarea progresiei leziunilor pulmonare, am propus o funcție inspirată din formula binecunoscută vitezei în fizică. Însă aceasta a fost definită ca o viteză relativă, deoarece un pacient poate fi realist comparat doar cu el însuși în ce privește evoluția bolii, dar nu cu alții.

$$v = \begin{cases} \frac{(s - s_0)}{s_0 \times t}, & \text{for } s_0 \neq 0 \\ \frac{s}{t}, & \text{otherwise} \end{cases}$$

Unde  $s$  este metrica evaluată (gr total, gr mediu, gr max) iar  $s_0$  este punctul inițial folosit pentru normalizare.  $t$  este o măsură de timp exprimată în ani, dat fiind că pacienții cu PID efectuează controale de acest tip anual.

### Rezultate

Ca exemple relevante, au fost selectate în continuare două cazuri de pacienți cu PID în cazul cărora a fost evaluată și comparată viteza de progresie pe cele 3 metrici menționate bazat pe evaluări imagistice din ani consecutivi.

Rezultatele au arătat modul de progresie al celor trei benzi HU (E, GGO, C) bazat pe cele 3 metrici, comparativ cu testele pulmonare efectuate în cadrul analizelor medicale de către specialiști.

În acest scop am calculat și un T-test (metodă statistică) pentru validarea rezultatelor obținute. Acest test arată că viteza de progresie relativă detectată cu ajutorul metodei propuse, față de evaluările clinice (DLco) este valabilă în cazul benzilor GGO și C, dar nu și în cazul benzii E.

În continuare, în scopul validării ipotezei de detecție timpurie a acestor tipuri de

patologii, pacienții au fost împărțiți în două categorii: pacienți normali și pacienți cu cazuri de PID în stadii incipiente având parametri funcționali acceptabili. Am creat astfel un context în care metoda propusă să poată detecta semne timpurii ale PID.

Metricile evaluate sunt tot cele trei menționate anterior (grad maxim, grad mediu, grad total). Bazat pe aceste metrici, un alt T-test a fost efectuat iar rezultatele arată că în cazul benzilor GGO și C rezultatele sunt valide, însă nu și în cazul benzii de Emfizem.

## **Discuție**

Date fiind rezultatele T-testelor, se poate spune că această metodă de analiză a CT-urilor poate evalua viteza de progresie a leziunilor pulmonare în cazul a două benzi HU (GGO și Consolidare) în mod precis și cantitativ. În cazul benzii de Emfizem însă, nu se poate spune că detectează variații semnificative, sau nu unele care pot fi corelate cu testele funcționale (DLco) ale pacienților.

În cazul celei de-a doua ipoteze propuse, cea a detecției timpurii, metoda propusă demonstrează că pe benzile GGO (toate cele 3 metricile) și C (două dintre metrici) poate observa diferențe notabile din punct de vedere statistic între pacienți normali și cei la limită, însă nu și în cazul benzii E.

În concluzie, referitor la cea de-a doua ipoteză propusă, modelul bazat pe rețele complexe s-a dovedit eficient în detectarea modificărilor timpurii pe banda GGO, parțial valid în cazul benzii Consolidare, și fals în cazul benzii de Emfizem.

## **5. Concluzii**

### **1. Concluzii – Tipare vizuale (Algoritm de vizualizare hibrid bazat pe rețele complexe)**

Algoritmul hibrid propus în **Capitolul 3** se dorește a fi o alternativă îmbunătățită la soluțiile software existente în domeniul vizualizării de rețele complexe și al căror scop este evidențierea tiparelor vizuale. Acest tip de algoritm s-a dovedit eficient în a umple nișa lăsată nevalorificată de aplicațiile software existente: un tip de vizualizare hibridă/ combinată între 2D și 3D, cu un sistem de referință integrat și care generează un tip de vizualizare interactivă, lăsând utilizatorul să manipuleze sau să reorienteze imaginea de-a lungul celor 3 axe xOyz. Imaginea poate fi mărită sau micșorată fără a pierde din acuratețe. Nodurile sunt bine distribuite în spațiul 3D, iar comunitățile sunt bine delimitate, oferind modalități clare de observare a ierarhiei între elemente, a dimensiunii, și a distanțelor. Comparativ cu celelalte soluții software din domeniu, algoritmul prezintă avantaje atât vizuale, prin sistemul de referință, coloristică atractivă, o agregare mai clară a comunităților, dar și prin performanță, acolo unde seturile de date biomedicale au o dimensiune considerabilă de sute de noduri și o densitate mare a muchiilor.

Viitoare direcții de îmbunătățire vizează eficientizarea timpului de rulare în special în afișarea muchiilor, reducerea suprapunerii muchiilor sau adăugarea unei interfețe grafice prietenoasă cu utilizatorul.

### **2. Concluzii – Modelarea tiparelor numerice (O nouă metodă pentru interpretarea imaginilor CT)**

Am dezvoltat un nou algoritm (**Capitolul 4**) bazat pe rețele complexe capabil să evidențieze tipare numerice în cazul HRCT-uri medicale de plămâni. Această abordare analizează date medicale pe trei dimensiuni relevante în analiza Pneumopatiilor Interstițiale

Difuze (Emfizem, GGO, Consolidare) și implică aproximarea distribuției gradelor prin funcții matematice.

Am adus un argument solid în ce privește dimensiunea secțiunii analizate (65 x 65px) iar comparativ cu metodele existente deja în domeniu, este chiar mai cuprinzătoare. Aceste dimensiuni se justifică și din punct de vedere medical deoarece pot acoperi o întreagă unitate funcțională a plămânului (lobulul pulmonar secundar).

Criteriile de interconectare ale nodurilor și distanța radială selectate sunt demonstrate și susținute prin experimente extensive coroborate cu principiile de atașament ale acestor tipuri de ecosisteme.

Metoda propusă folosește benzile Hounsfield (HU) pentru a împărți imaginea în mai multe straturi dar și drept criteriu de similitudine între pixeli potențial înrudiți, permițând o granularitate mai mare în observarea leziunilor pulmonare. Intervalele benzilor HU sunt dependente și de rezoluția aparatului folosit pentru generarea HRCT-urilor.

Secțiunea de discuții justifică din punct de vedere al științei sistemelor coerența și corectitudinea algoritmului bazat pe rețele complexe, prin folosirea metricii de distribuție a gradelor. Se arată de asemenea diferențele în termeni de clusterizare între loturile de pacienți normali și cei bolnavi. De asemenea, metoda este validată din punct de vedere medical prin capacitatea sa de a reprezenta fidel și cu o granularitate foarte mare datele clinice, iar acest lucru se poate dovedi crucial în procesul de diagnosticare.

În final, această metodă este comparată cu alte metode existente în domeniu pentru a sublinia avantajul folosirii ei: acela că oferă un instrument de analiză calitativă și cantitativă a imaginilor medicale.

Între dezavantaje și viitoare direcții de îmbunătățire, de menționat faptul că în stadiul curent, algoritmul nu funcționează ca o aplicație software de sine stătătoare (necesită cunoștințe tehnice de programare), și de asemenea lotul modest care a fost folosit pentru a testa algoritmul. Tot la capitolul viitoare îmbunătățiri menționez implementarea unei interfețe grafice prietenoasă cu utilizatorul.

În concluzie, noul algoritm bazat pe rețele complexe s-a dovedit a fi extrem de util în cazul procesului de diagnostic al PID, transformând HRCT-uri pulmonare în structuri calificabile și cuantificabile din punct de vedere matematic și statistic.

### **3. Concluzii Analiza tiparelor numerice (Îmbunătățirea diagnosticului imagistic al Patologiilor Interstițiale Difuze)**

Această metodă bazată pe rețele complexe (prezentată în Capitolul 4) a fost dezvoltată ca mod de suport în cadrul procesului de diagnostic al PID. În acest scop, au fost evaluate două ipoteze: viteza lor de progresie a patologiilor din categoria PID dar și detecția timpurie (**Capitolul 5**). Motivația a constat în faptul că acestea sunt două dintre cele mai complicate aspecte ale diagnosticului PID cu care specialiștii se confruntă.

Pentru prima dintre ipoteze, aceea a vitezei de progresie, algoritmul și metoda propusă s-au dovedit eficiente. Dat fiind gradul de precizie de 3mm pentru leziunile pulmonare, această abordare a demonstrat o bună corelare cu simptomele clinice la un asemenea nivel care nu poate fi obținut prin testele clinice obișnuite. De notat aici că stratul de Emfizem reprezintă o excepție în acest sens.

În ce privește cea de-a doua ipoteză, cea de detecție timpurie, cele mai bune rezultate au fost obținute pentru benzile de GGO și Consolidare. Aceste două benzi sunt foarte importante din punct de vedere medical în diagnosticarea PID și este o abilitate esențială demonstrată de algoritmul propus, spre deosebire de alte soluții software precum CALIPER.

În ce privește direcțiile de îmbunătățire, de menționat eficientizarea timpului de execuție, integrarea într-o soluție software mai mare care să includă și capabilități de machine



learning și segmentare, dar și competențe de pattern recognition bazate pe analiza rețelelor complexe.

### **Bibliografie selectivă**

1. V. M. Ancusa and L. Broasca, "A Method to Pinpoint Undiscovered Links in Genetic and Protein Networks," *Stud. Health Technol. Inform.*, vol. 210, pp. 771–775, 2015.
2. Y. Li and L. Chen, "Big biological data: challenges and opportunities.," *Genomics Proteomics Bioinformatics*, vol. 12, no. 5, pp. 187–189, Oct. 2014, doi: 10.1016/j.gpb.2014.10.001.
3. H. Jeong, Tombor, R. Albert, Z. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, pp. 651–654, 2000.
4. R. Albert and A.-L. Barabási, "Topology of Evolving Networks: Local Events and Universality," *Phys. Rev. Lett.*, vol. 85, pp. 5234–5237, 2000.
5. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, vol. 2008, no. 10, p. P10008, Oct. 2008, doi: 10.1088/1742-5468/2008/10/p10008.
6. L. Broască, V. Ancușă, and H. Ciocârlie, "Bioinformatics Visualisation Tools: An Unbalanced Picture," *Stud. Health Technol. Inform.*, vol. 228, pp. 760–764, 2016.
7. O.-H. Kwon, T. Crnovrsanin, and K.-L. Ma, "What Would a Graph Look Like in this Layout? A Machine Learning Approach to Large Graph Visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 478–488, Jan. 2018, doi: 10.1109/tvcg.2017.2743858.
8. L. Broasca, V.-M. Ancusa, and H. Ciocarlie, "A Qualitative Analysis on Force Directed Network Visualization Tools in the Context of Large Complex Networks," in *2019 23rd International Conference on System Theory, Control and Computing (ICSTCC)*, Oct. 2019, pp. 656–661. doi: 10.1109/ICSTCC.2019.8885641.
9. L. Broască, V.-M. Ancușă, and H. Ciocârlie, "Towards a Hybrid Layout for Complex Networks Visualization," in *2020 24th International Conference on System Theory, Control and Computing (ICSTCC)*, 2020, pp. 118–123. doi: 10.1109/ICSTCC50638.2020.9259656.
10. L. Broasca, V. Ancusa, and H. Ciocarlie, "A 3D Surface Fitting Layout for Complex Networks Visualization," *Stud. Health Technol. Inform.*, vol. 272, pp. 362–365, Jun. 2020, doi: 10.3233/SHTI200570.
11. A. Christe et al., "Computer-Aided Diagnosis of Pulmonary Fibrosis Using Deep Learning and CT Images.," *Invest. Radiol.*, vol. 54, no. 10, pp. 627–632, Oct. 2019, doi: 10.1097/RLI.0000000000000574.
12. L. Vulliard and J. Menche, "Complex Networks in Health and Disease," *Syst. Med.*, pp. 26–33, 2021, doi: 10.1016/B978-0-12-801238-3.11640-X.
13. A. Truşculescu, L. Broască, V. M. Ancușă, D. Manolescu, E. Tudorache, and C. Oancea, "Managing Interstitial Lung Diseases with Computer-Aided Visualization," in *Hybrid Artificial Intelligence and IoT in Healthcare*, A. Kumar Bhoi, P. K. Mallick, M. Narayana Mohanty, and V. H. C. de Albuquerque, Eds. Singapore: Springer Singapore, 2021, pp. 245–271. doi: 10.1007/978-981-16-2972-3\_12.
14. J.-Z. Cheng et al., "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans," *Sci. Rep.*, vol. 6, no. 1, p. 24454, Apr. 2016, doi: 10.1038/srep24454.
15. W. Shi et al., "A deep learning-based quantitative computed tomography model for predicting the severity of COVID-19: a retrospective study of 196 patients.," *Ann. Transl. Med.*, vol. 9, no. 3, p. 216, Feb. 2021, doi: 10.21037/atm-20-2464
16. L. Li et al., "Using Artificial Intelligence to Detect COVID-19 and Community-acquired

- Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy,” *Radiology*, vol. 296, no. 2, pp. E65–E71, Aug. 2020, doi: 10.1148/radiol.2020200905.
17. 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.
  18. 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.
  19. 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.
  20. T. Zrimec and S. Busayarat, “Computer-aided Analysis and Interpretation of HRCT Images of the Lung,” 2011. doi: 10.5772/14507.
  21. A. Depeursinge, T. Zrimec, S. Busayarat, and H. Müller, “3D Lung Image Retrieval Using Localized Features,” *SPIE Med. Imaging 2011 Lake Buena Vista Orlando Fla. U. S.*, Oct. 2011, doi: 10.1117/12.877943.
  22. 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.
  23. L. Broască et al., “A Novel Method for Lung Image Processing Using Complex Networks,” *Tomogr. Ann Arbor Mich*, vol. 8, no. 4, pp. 1928–1946, Jul. 2022, doi: 10.3390/tomography8040162.
  24. R. Grassi et al., “COVID-19 pneumonia: computer-aided quantification of healthy lung parenchyma, emphysema, ground glass and consolidation on chest computed tomography (CT),” *Radiol. Med. (Torino)*, vol. 126, no. 4, pp. 553–560, Apr. 2021, doi: 10.1007/s11547-020-01305-9.
  25. A. A. Truşculescu et al., “Enhancing Imagistic Interstitial Lung Disease Diagnosis by Using Complex Networks,” *Med. Kaunas Lith.*, vol. 58, no. 9, Sep. 2022, doi: 10.3390/medicina58091288.