



Universitatea  
Transilvania  
din Braşov

ŞCOALA DOCTORALĂ INTERDISCIPLINARĂ

Facultatea: Inginerie Electrică și Știința Calculatoarelor

Ing. Andrei PUIU

**Valorificarea învățării auto-supervizate, a  
datelor sintetice și a soluțiilor de inteligență  
artificială sigură pentru gestionarea  
inovativă a cancerului**

**Leveraging Self-Supervised Learning,  
Synthetic Data, and Trustworthy Artificial  
Intelligence for Innovative Cancer Care**

REZUMAT

Conducător științific

Prof. univ. dr. ing. Constantin SUCIU

BRAȘOV, 2024





Universitatea  
Transilvania  
din Braşov

D-lui (D-nei) .....

## Componenta

### Comisiei de doctorat

Numită prin ordinul Rectorului Universităţii Transilvania din Braşov

Nr. .... din .....

<b>PREŞEDINTE:</b>	Prof. univ. dr. ing. MORARU Sorin-Aurel Director de departament Universitatea Transilvania din Braşov
<b>CONDUCĂTOR ŞTIINŢIFIC:</b>	Prof. univ. dr. ing. SUCIU Constantin Universitatea Transilvania din Braşov
<b>REFERENŢI:</b>	Prof. univ. dr. ing. MICLEA Liviu Universitatea Tehnică din Cluj Napoca Prof. univ. dr. ing. NECOARĂ Ion Universitatea Naţională de Ştiinţă şi Tehnologie Politehnica Bucureşti Prof. univ. dr. fiz. URSUŢIU Doru Universitatea Transilvania din Braşov

**Data, ora şi locul susţinerii publice a tezei de doctorat: 13.09.2024, ora 9:00, sala VIII9**

**Eventualele aprecieri sau observaţii asupra conţinutului lucrării vor fi transmise electronic, în timp util, pe adresa andrei.puiu@unitbv.ro**

**Totodată, vă invităm să luaţi parte la şedinţa publică de susţinere a tezei de doctorat.**

**Vă mulţumim.**



# Cuprins

<b>1</b>	<b>Introducere</b>	<b>1</b>
1.1	Învățarea profundă pe date medicale restricționate: potențial și provocări . . . . .	1
1.2	Scopul tezei de doctorat . . . . .	5
1.3	Obiective specifice . . . . .	5
1.4	Structura tezei . . . . .	6
<b>2</b>	<b>Învățare auto-supervizată pentru extrapolarea și coregistrarea imaginilor medicale</b>	<b>7</b>
2.1	Introducere . . . . .	7
2.2	Extrapolarea imaginilor CT subțiri pentru intervențiile asupra ficatului . . . . .	8
2.3	Extrapolarea asimetrică . . . . .	9
2.3.1	Metodologie . . . . .	10
2.3.1.1	Setul de date . . . . .	11
2.3.1.2	Metoda propusă . . . . .	11
2.3.1.3	Cuantificarea performanței . . . . .	13
2.3.2	Rezultate . . . . .	14
2.3.2.1	Testul de detecție a reperelor . . . . .	14
2.3.2.2	Testul de coregistrare . . . . .	14
2.4	Extrapolarea simetrică . . . . .	15
2.4.1	Metodologie . . . . .	16
2.4.1.1	Setul de date . . . . .	16
2.4.1.2	Metoda propusă . . . . .	17
2.4.2	Rezultate . . . . .	17
2.4.2.1	Testul de detecție a reperelor . . . . .	17
2.4.2.2	Testul de coregistrare . . . . .	17
2.4.3	Concluziile abordărilor de extrapolare . . . . .	18
2.4.4	Discuții asupra abordărilor prezentate . . . . .	19
2.5	Concluzii . . . . .	19
2.6	Discuții . . . . .	20
<b>3</b>	<b>Generarea de date sintetice pentru stratificarea cancerului de prostată</b>	<b>21</b>
3.1	Introducere . . . . .	21
3.2	Stratificarea cancerului de prostată . . . . .	22
3.3	Metode . . . . .	23
3.3.1	Generarea de date sintetice . . . . .	23
3.3.2	Predicția stadiului TNM clinic . . . . .	26
3.3.3	Descrierea experimentală . . . . .	27
3.4	Rezultate . . . . .	27
3.4.1	Fidelitatea datelor sintetice . . . . .	27
3.4.2	Predicția stadiului TNM clinic . . . . .	29
3.5	Concluzii . . . . .	30
3.6	Discuții . . . . .	31

<b>4</b>	<b>Inteligența Artificială sigură pentru aplicațiile clinice în diagnosticarea și tratamentul cancerului.</b>	<b>33</b>
4.1	Introducere . . . . .	33
4.2	Rețea de asociere clinică pentru predicția cancerului de prostată clinic semnificativ . . . . .	35
4.2.1	Introducerea scenariului clinic . . . . .	35
4.2.2	Metode . . . . .	35
4.2.2.1	Setul de date . . . . .	36
4.2.2.2	Rețeaua de calificare a leziunilor . . . . .	36
4.2.2.3	Rețeaua de asociere clinică . . . . .	37
4.2.3	Rezultate . . . . .	38
4.2.3.1	Explicabilitate și interpretabilitate . . . . .	38
4.2.4	Concluzii . . . . .	39
4.2.5	Discuții . . . . .	39
4.3	Clasificarea subtipurilor de cancer pulmonar fără celule mici . . . . .	40
4.3.1	Introducerea scenariului clinic . . . . .	40
4.3.2	Rezultate . . . . .	41
4.3.3	Discuții . . . . .	43
4.3.4	Materiale și metode . . . . .	45
4.3.4.1	Date . . . . .	45
4.3.4.2	Metodologie . . . . .	46
4.4	Concluzii . . . . .	47
4.5	Discuții . . . . .	48
<b>5</b>	<b>Concluzii finale</b>	<b>49</b>
5.1	Concluzii . . . . .	49
5.1.1	Învățare auto-supervizată pentru extrapolarea și coregistrarea imaginilor medicale . . . . .	50
5.1.2	Generarea de date sintetice pentru stratificarea cancerului de prostată . . . . .	50
5.1.3	Inteligența artificială sigură . . . . .	51
5.2	Contribuții originale . . . . .	51
5.2.1	Sinteza contribuțiilor . . . . .	53
5.3	Diseminarea rezultatelor cercetării . . . . .	55
5.4	Discuții . . . . .	57
	<b>Bibliografie</b>	<b>58</b>
	<b>Rezumat</b>	<b>69</b>

# 1. Introducere

---

Învățarea profundă pe date medicale restricționate: potențial și provocări  
Scopul tezei de doctorat  
Obiective specifice  
Structura tezei

---

## 1.1 Învățarea profundă pe date medicale restricționate: potențial și provocări

### Cancerul: Criza modernă de sănătate

Cancerul este una dintre cele mai răspândite și devastatoare boli la nivel mondial, afectând milioane de oameni în fiecare an. Conform Organizației Mondiale a Sănătății (OMS), cancerul este a doua cauză principală de deces la nivel global, reprezentând aproape unul din șase decese [1]. Din punct de vedere terminologic, cancerul se referă la un grup de boli caracterizate printr-o creștere și răspândire necontrolată a celulelor anormale. Datorită complexității sale, gestionarea cancerului implică o abordare multidisciplinară care include prevenirea, detectarea timpurie, diagnosticarea eficientă, tratamentul și îngrijirea paliativă, fiecare aspect jucând un rol semnificativ în obținerea unor rezultate favorabile pentru pacienți.

În ciuda multor necunoscuturi cu care ne confruntăm în prezent în lupta împotriva cancerului, s-a demonstrat că prevenția joacă un rol crucial în reducerea incidenței sale. Cu toate acestea, detectarea în timp util a cancerului este esențială pentru maximizarea șanselor unui rezultat favorabil. Acest lucru se realizează de obicei prin screening regulat, permițând un diagnostic în stadiu incipient al bolii care crește eficacitatea tratamentului [3]. Procedurile de diagnostic includ adesea o combinație de imagistică, teste de laborator, biopsii și examinări histopatologice. Tehnicile avansate de imagistică, precum MRI, CT și PET, oferă vizualizări detaliate ale tumorilor, esențiale pentru stadializare și planificarea tratamentului. Cu toate acestea, biopsiile rămân în continuare standardul în confirmarea prezenței și a tipului de cancer. Totuși, apariția diagnosticului molecular a revoluționat procesul de diagnosticare a cancerului, permițând identificarea mutațiilor genetice specifice. Această abordare de medicină de precizie permite personalizarea planurilor de tratament pe baza caracteristicilor unice ale cancerului unei persoane, sporind eficacitatea terapiilor și minimizând efectele adverse.

Tratamentul cancerului cuprinde diverse modalități, fiecare adaptată tipului, stadiului și locației cancerului, precum și stării generale de sănătate și preferințelor pacientului. Opțiunile principale de tratament includ chirurgie, radioterapie, chimioterapie, imunoterapie și terapie țintită. De asemenea, îngrijirea paliativă este o componentă integrală a gestionării cancerului, concentrându-se pe îmbunătățirea calității vieții pacienților cu cancer avansat. Aceasta abordează simptome fizice precum durerea, greața și oboseala, precum și nevoile emoționale, sociale și spirituale.

Gestionarea eficientă a cancerului necesită o echipă multidisciplinară de profesioniști din domeniul sănătății, incluzând oncologi, chirurghi, radiologi, patologi, asistenți medicali și specialiști în îngrijire psihologică. Această abordare colaborativă asigură faptul că toate aspectele îngrijirii pacientului sunt abordate, de la diagnosticare până la tratament.

## Învățarea profundă în medicină

În ultimele decenii, progresele tehnologice remarcabile au remodelat practicile medicale, plasând industria sănătății într-o fază de transformare. Creșterea exponențială a puterii de calcul și a capacității de stocare a datelor a condus la multe idei inovatoare care reprezintă în prezent elemente de bază pentru furnizarea de servicii de sănătate superioare. Acest lucru se reflectă bine în rezultatele pacienților, jucând astfel, alături de alți factori, un rol semnificativ în menținerea unei tendințe pozitive în media globală a speranței de viață [1].

De exemplu, imagistica medicală, care a devenit utilizată de rutină în practicile clinice, joacă un rol major în fiecare aspect al sănătății, contribuind semnificativ la diagnosticarea precisă, planificarea tratamentului, administrarea și urmărirea acestuia [4]. Bruls și Kwee [5] au investigat numărul de studii imagistice efectuate în timpul orelor de gardă între 2006 și 2020, raportând o creștere dramatică: numărul absolut de radiografii (XR), ecografi (US) și tomografii computerizate (CT) efectuate într-o lună a crescut de la 1105 la 1805 (63%), de la 36 la 118 (227%) și respectiv, de la 112 la 817 (629%). Cu toate acestea, deși beneficii incontestabile vin din utilizarea intensivă a imagisticii în practicile clinice, radiologii se confruntă în prezent un volum de muncă enorm, fiind necesar să interpreteze imaginile și să furnizeze un raport clinicianului referent într-un timp cât mai scurt [5, 6].

În același timp, inteligența artificială (AI) a cunoscut o creștere remarcabilă, beneficiind de o atenție enormă din partea comunității științifice. Datorită potențialului său de a modela concepte complexe neliniare, fezabilitatea sa se extinde în diverse domenii, permițând o gamă largă de aplicații care anterior erau considerate inaccesibile. Prin urmare, integrarea tehnologiilor AI în industria sănătății este în prezent explorată pe scară largă, generând multe promisiuni în optimizarea fluxurilor de lucru clinice, și astfel, în îmbunătățirea îngrijirii pacienților [7, 8].

Sistemele de diagnosticare asistată de calculator bazate pe învățare profundă (DL-CAD) sunt concepute pentru a sprijini furnizorii de servicii medicale în diagnosticarea eficientă a pacienților pe baza unei anumite modalități, menținând în același timp niveluri rezonabile de muncă. Prin urmare, integrarea AI în sănătate ar putea acoperi lacunele existente în practicile clinice actuale, sprijinind furnizorii de îngrijiri medicale în oferirea celor mai bune servicii posibile, îmbunătățind acuratețea diagnosticului și, prin urmare, rezultatele actului medical. Deși s-a observat deseori o diferență de opinii între clinicienii cu diferite niveluri de experiență [9, 10, 11], adoptarea sistemelor DL-CAD în rutinele clinice ar putea duce la o abordare simplificată, avantajoasă pentru personalul medical cu diferite niveluri de senioritate și pentru pacienții acestora.

Din punct de vedere al procesului de dezvoltare, toți algoritmi de învățare automată se bazează exclusiv pe date suficiente, calitative și complete pentru a produce rezultate fiabile. Paradigma de antrenare a acestor algoritmi poate fi clasificată în mare în învățare supravegheată (supervizată) și nesupravegheată (nesupervizată). În abordarea nesupravegheată, modelarea se referă în principal la gruparea datelor de antrenare pe baza similarităților lor într-un număr predefinit de așa numite clustere. Mai departe, când modelului i se prezintă un caz nou, acesta îl va atribui unui anumit grup, indicând astfel apartenența la o clasă specifică. Cu toate acestea, paradigma de învățare supravegheată a fost stabilită ca metodologie preferată în antrenarea rețelelor neuronale. În această abordare, perechile de intrări și ieșiri sunt utilizate pentru a ghida optimizarea, după cum urmează: modelul este antrenat iterativ prin propagarea derivatelor parțiale ale unui semnal de eroare prin rețea, optimizând astfel modelul pentru a realiza o anumită sarcină de predicție.

Cu toate acestea, învățarea supervizată depinde de o serie de cerințe:

- Suficiența datelor. Aceeași capacitate de modelare care a impus algoritmi de învățare automată ca soluții de ultimă generație pentru rezolvarea problemelor complexe îi face susceptibili la o generalizare slabă a cunoștințelor pe cazuri noi, în special atunci când datele de antrenare sunt insuficiente (un fenomen denumit pe scară largă „overfitting”).
- Acuratețea etichetelor. Modelele de învățare automată profundă sunt concepute pentru a modela tipare și corelații în date, cu scopul de a le folosi ulterior în predicția etichetelor corespunzătoare. Prin urmare, adnotările precise sunt esențiale pentru un proces de antrenament corect,



și în final, pentru obținerea unui model robust și precis.

- **Completitudinea datelor.** Setul de caracteristici de intrare furnizat modelului ar trebui să posedă o capacitate predictivă suficientă în raport cu eticheta.
- **Calitatea datelor.** Datele afectate de zgomot sau artefacte ar trebui excluse din seturile de antrenament, deoarece acestea ar putea împiedica modelul să ajungă la o stare optimă.

## **Provocări în dezvoltarea soluțiilor bazate pe AI pentru domeniul medical**

În ciuda potențialului incontestabil pe care AI îl deține în oferirea de soluții pentru depășirea problemelor emergente din industria sănătății, dezvoltarea produselor în acest domeniu este adesea împiedicată de un set de obstacole.

Pentru început, deși datele biomedicale sunt abundente, circulația lor este restricționată din cauza principiilor etice care provin din preocupările legate de încălcarea confidențialității pacienților [12, 4]. Constrângerile impuse de Regulamentul General privind Protecția Datelor (GDPR) în Europa și Legea privind Portabilitatea și Responsabilitatea Asigurărilor de Sănătate (HIPAA) în SUA restricționează drastic accesul la datele biomedicale pentru a proteja confidențialitatea pacienților. Deși sunt fundamental corecte, aceste constrângeri obstrucționează semnificativ calea inovațiilor bazate pe AI care ar putea îmbunătăți considerabil practicile clinice, ducând în cele din urmă la un sistem de sănătate îmbunătățit. Pentru a utiliza datele pacienților în scopuri de cercetare și dezvoltare de produse, trebuie obținut un consimțământ informat. Cu toate acestea, o eventuală permisiune privește doar scopurile inițiale, ceea ce împiedică orice utilizare sau explorare ulterioară.

În al doilea rând, paradigma de învățare supervizată se bazează exclusiv pe adnotări precise și calitative pentru a ghida în mod robust procesul de învățare către un punct optim. Cu toate acestea, de la un caz de utilizare la altul, adnotarea datelor poate necesita un nivel ridicat de expertiză, fiind astfel accesibilă exclusiv practicienilor din domeniul sănătății [13]. Având în vedere volumul crescut de muncă deja menționat [5, 6], crearea prospectivă de etichete pentru un set de date suficient de mare este destul de nerealistă, implicând alocarea unui timp suplimentar. Cu toate acestea, în cazurile relativ mai simple, adnotatori non-profesioniști ar putea fi instruiți în etichetarea datelor, prezentând totuși un risc relativ crescut de eroare.

În al treilea rând, în ciuda îmbunătățirilor recente în capacitățile de stocare a datelor, a serviciilor de tip „cloud” și adoptării sistemelor de dosare electronice de sănătate (EHR) în multe spitale, partajarea datelor pacienților între diferiți furnizori de servicii medicale este în prezent obstrucționată de lipsa unui set clar de standarde și soluții de interoperabilitate. Prin urmare, deoarece pacienții ar putea avea consultații la diferite instituții pentru aceeași afecțiune, achiziționarea de date longitudinale este adesea afectată de incompletitudine, împiedicând utilizarea unui portret complet al pacientului în dezvoltarea soluțiilor bazate pe învățare automată.

În final, algoritmi de învățare profundă tind să fie excesiv de încrezători în predicțiile lor, ceea ce poate duce la furnizarea de rezultate eronate pentru cazurile nesigure. Datorită complexității lor relativ crescute, algoritmi de învățare profundă sunt adesea considerați cutii negre, ridicând scepticism în jurul adoptării lor în rutinele clinice. Indiscutabil, mulți clinicieni, împreună cu pacienții lor, sunt reticenți în a urma sugestii venite dintr-o sursă abstractă fără o înțelegere adecvată a proceselor de raționament utilizate. În consecință, lipsa inerentă de transparență manifestată de aceste metode împiedică adoptarea lor pe scară largă în practicile clinice.

## **Soluții de ultimă generație pentru depășirea provocărilor specifice din domeniul sănătății**

Ca răspuns la provocările menționate anterior, comunitatea științifică din domeniul tehnologic a propus mai multe soluții pentru a aborda diferite aspecte ale reformei sistemului de sănătate, incluzând (1) anonimizarea, învățarea descentralizată [14], antrenarea pe date criptate [15, 16, 17, 12] și generarea de date sintetice pentru a aborda problemele legate de insuficiența datelor, (2) paradigmele de învățare nesupervizată, semi-supervizată și auto-supervizată pentru a aborda provocările

legate de etichetare, și (3) estimarea importanței relative a caracteristicilor și cuantificarea incertitudinii pentru a crește gradul de încredere.

Generarea de date sintetice a fost explorată extensiv ca o soluție alternativă sau adiacentă pentru colectarea seturilor de date medicale la scară largă. Ideea de a fabrica date virtuale pentru a sprijini dezvoltările tehnologice în domeniul sănătății are un potențial remarcabil datorită proprietăților sale de protecție a confidențialității. Prin urmare, au fost propuse multe strategii pentru a aborda compromisurile între confidențialitate și utilitate, clasificate pe larg în metode sintetice parțiale, complete și hibride [18, 19, 20]. Conceptul de generare de date sintetice parțiale (strâns legat de metodele de anonimizare neperturbativă) se referă la limitarea generării sintetice a caracteristicilor identificabile ale pacienților, păstrând informațiile nesensibile, maximizând astfel gradul de utilitate a datelor. În schimb, generarea completă de date sintetice pune mai mult accent pe proprietățile de conservare a confidențialității, compromițând într-o oarecare măsură utilitatea. Prin urmare, generarea de date hibride urmărește găsirea unui compromis optim între calitatea datelor și proprietățile de asigurare a confidențialității, combinând informații reale cu entități fabricate.

Auto-supervizarea este în prezent una dintre cele mai utilizate tehnici pentru gestionarea seturilor de date parțial etichetate. Propusă inițial pentru problemele de procesare a limbajului natural (NLP), unde datele textuale sunt abundente, dar de obicei lipsesc adnotările [21], paradigma de învățare auto-supervizată a fost larg adoptată în dezvoltările tehnologice din domeniul sănătății datorită potențialului său remarcabil de a crește semnificativ numărul de exemple de antrenare utilizabile, și, astfel, puterea de generalizare. Ideea generală a acestei abordări este de a utiliza date neetichetate pentru a învăța reprezentări abstracte utile care pot fi reutilizate ulterior în rezolvarea sarcinilor biomedicale specifice pe baza unei porțiuni relativ reduse de exemple etichetate. De obicei, auto-supervizarea este utilizată împreună cu tehnicile de transfer de cunoștințe [13]: în primul rând, reprezentările latente cuprinzătoare sunt obținute prin utilizarea problemelor surogat, adesea denumite sarcini de pretext, unde informațiile de referință ar putea fi fie deduse, fie sintetizate din intrările neetichetate; ulterior, episodul principal de învățare utilizează modelele pre-antrenate bazate pe auto-supervizare ca ponderi inițiale, astfel încât optimizarea să înceapă dintr-un punct avantajos. Prin urmare, deoarece reprezentările abstracte ale datelor sunt deja produse de ponderile modelului pre-antrenat, se poate obține o soluție robustă de învățare profundă pe un set relativ redus de mostre de date etichetate, reducând astfel drastic povara procesului de adnotare. Cu toate acestea, în unele scenarii, auto-supervizarea este adesea suficientă pentru a atinge soluția finală, fără a necesita pași suplimentari de optimizare. De exemplu, completarea imaginii (cunoscută și sub denumirea de „inpainting”) poate fi antrenată prin eliminarea aleatorie de regiuni din intrare și utilizarea rețelelor neuronale convoluționale pentru a restaura informațiile într-un mod consistent [22, 23] din punct de vedere semantic.

Cu toate acestea, deși abordările menționate anterior ar putea acționa ca factori facilitatori pentru integrarea AI în practicile clinice, trebuie asigurată o funcționare robustă și transparentă pentru a crește încrederea furnizorilor de îngrijiri medicale și a pacienților lor în astfel de soluții. Adesea denumite cutii negre, modelele de învățare profundă necesită mecanisme care, pe lângă oferirea de modalități de interpretare a proceselor lor de raționament, identifică cazurile în care predicțiile lor ar putea fi susceptibile la erori. Prin urmare, au fost propuse diverse metode pentru a aborda aceste cerințe.

În contrast cu abordările clasice de învățare automată, algoritmi de învățare profundă manifestă o lipsă inerentă de transparentă datorită complexității lor. Pentru a elucida modul în care rețelele neuronale capturează și folosesc tiparele din date pentru a deduce anumite rezultate, au fost propuse mai multe abordări [24, 25, 26, 27, 28], având în principal scopul de a atribui scoruri de importanță relativă fiecărei caracteristici de intrare în raport cu predicția finală. De exemplu, Lundberg et. al. [27] au propus o metodă pentru a clasifica intrările modelului pe baza contribuției lor la rezultat, estimând valorile Shapley. Provenind din teoria jocurilor, analiza Shap a fost inițial concepută pentru a împărți o recompensă între jucători pe baza contribuției lor relative la rezultat. În ceea ce privește învățarea automată, setul de caracteristici de intrare sunt considerate jucători, în timp ce predicția modelului

reprezintă recompensa, permițând astfel calcularea unui scor de importanță relativă pentru fiecare predictor în parte. Valorile Shapley reprezintă explicații excelente despre cum un model a ajuns la o anumită concluzie, crescând astfel semnificativ transparența acestuia. Mai mult, analiza Shap poate fi utilizată pentru a oferi explicații la nivel global sau la nivel de instanță. Analiza generală este foarte potrivită pentru evaluarea și caracterizarea inițială a modelului, permițând dezvoltatorilor și experților din domeniu să identifice și să remedieze potențialele defecțiuni într-o fază de incipientă, crescând semnificativ încrederea în sistem. La fel de important, explicațiile la nivel de instanță ar putea însoți predicțiile în timp real, permițând furnizorilor de îngrijiri medicale să valideze sau să respingă eficient raționamentul modelului.

Mai mult, o serie de abordări au fost propuse ca măsuri de siguranță, prevenind rețelele neuronale să facă predicții atunci când nu posedă suficientă expertiză. Deoarece algoritmi de învățare profundă se bazează puternic pe seturile de date utilizate în antrenamentul lor, predicțiile făcute pe cazuri noi, sub-reprezentate în distribuția exemplilor folosite la antrenare, sunt destul de nesigure. Prin urmare, mecanismele care permit rețelelor neuronale să ofere un răspuns de tipul „Nu știu” în loc de a încerca să deducă sunt extrem de importante pentru crearea unui cadru de funcționare sigur, prevenind astfel consecințele potențial dezastruoase asupra bunăstării pacienților. Multe abordări eficiente au fost propuse pentru a îndeplini aceste cerințe [29, 30, 31, 32], clasificate pe larg în metode probabilistice și metode bazate pe ansamblu de modele. De exemplu, pe lângă furnizarea de soluții de ultimă generație în multe aplicații clinice [33], ansamblurile de modele [34], prin construcția lor, pot oferi și estimări ale incertitudinii prin cuantificarea nivelului de concordanță între diverse instanțe, cu costul creșterii necesităților computaționale. În general, încorporarea acestor mecanisme în toate soluțiile bazate pe învățare automată ameliorează semnificativ preocupările legate de siguranță, conducând la un sistem general fiabil și de încredere.

## 1.2 Scopul tezei de doctorat

Această teză de doctorat își propune să studieze modul în care provocările întâmpinate în dezvoltarea soluțiilor bazate pe învățarea adâncă pentru industria sănătății pot fi abordate prin utilizarea unei serii de tehnici propuse în literatură. În ciuda potențialului remarcabil al inteligenței artificiale în rezolvarea diverselor probleme clinice actuale, adoptarea sa imediată este împiedicată de un set de obstacole care provin din calitatea și completitudinea datelor, disponibilitatea adnotărilor și lipsa inerentă de transparență. Prin urmare, investigăm aici cum tehnici precum auto-supervizarea, generarea de date sintetice, estimarea importanței relative a caracteristicilor și cuantificarea incertitudinii ar putea acoperi aceste lacune și permite dezvoltări într-un set de scenarii de utilizare clinic relevante. Datorită capacității lor de a modela concepte complexe neliniare, rețelele neuronale profunde sigure ar putea eficientiza semnificativ rutinele clinice actuale, îmbunătățind în același timp acuratețea diagnosticului, conducând astfel la rezultate superioare pentru pacienți. Prin urmare, explorările în acest domeniu sunt esențiale pentru a sprijini aducerea industriei medicale într-o nouă eră, în care tehnologiile recente sunt înțelept puse în slujba pacienților.

## 1.3 Obiective specifice

În mod specific, următoarele obiective au fost urmărite pe parcursul tezei:

- Îmbunătățirea acurateței coregistrării într-un sistem de ghidare a intervențiilor prin utilizarea unei abordări de învățare auto-supervizată în antrenarea modelelor AI pentru extrapolarea imaginilor medicale.
  - Dezvoltarea unui cadru pentru a simula achiziții de imagini subțiri din volume CT, astfel încât să se creeze perechi de intrări-ieșiri pentru antrenarea supervizată.

- Producerea de seturi de date adecvate pentru a rezolva (1) probleme de extrapolare asimetrică și (2) probleme de extrapolare simetrică.
- Utilizarea unui cadru generativ adversarial pentru antrenarea unui model de învățare profundă în extrapolarea CT-urilor intra-operatorii simulate.
- Dezvoltarea unei metode de stratificare a pacienților cu cancer de prostată pe baza dosarelor electronice de sănătate longitudinale create sintetic (EHR).
  - Definirea proceselor pentru diagnosticarea, tratamentul și monitorizarea cancerului de prostată.
  - Colectarea diferitelor proprietăți statistice din literatură, sau derivarea acestora din datele reale limitate disponibile, pentru a asigura coerența variantelor sintetice.
  - Construirea unui generator de date sintetice care poate produce dosare electronice de sănătate longitudinale realiste.
  - Studierea fezabilității antrenării unei rețele de predicție a stadiului TNM pe datele pacienților generate sintetic.
- Utilizarea tehnicilor de estimare a importanței caracteristicilor de intrare și a cuantificării incertitudinii pentru a spori transparența și fiabilitatea rețelelor neuronale.
  - Dezvoltarea unui model pentru a îmbunătăți acuratețea predicției unui sistem DL-CAD de ultimă generație, specializat în detectarea malignității, prin asocierea informațiilor clinice și demografice.
  - Dezvoltarea unui model pentru a clasifica cancerul pulmonar non-microcelular (NSCLC) în subtipurile sale, respectiv adenocarcinom pulmonar (LUAD) și carcinom cu celule scuamoase (LUSC), pe baza statutului mutațional al unui set heterogen de gene.
  - Efectuarea unei analize post-hoc a importanței caracteristicilor pentru a spori explicabilitatea și interpretabilitatea celor două modele menționate.
  - Investigarea modului în care estimările de incertitudine ar putea identifica cazurile nesigure unde modelul de clasificare a subtipului este predispus la erori, crescând astfel fiabilitatea acestuia.

## 1.4 Structura tezei

Această teză este structurată după cum urmează: Capitolul 2 subliniază importanța auto-supervizării prin prezentarea unei abordări de antrenare a rețelelor neuronale generative adversariale în extrapolarea volumelor subțiri de CT, simetric și asimetric. Abordabilă exclusiv datorită paradigmei de învățare auto-supervizată, metoda noastră a îmbunătățit semnificativ acuratețea unei faze ulterioare de coregistrare, conducând la o robustețe superioară a întregului flux. Capitolul 3 prezintă o metodologie sistematică pentru generarea de dosare electronice de sănătate sintetice realiste. Fiind lipsite de orice preocupări legate de confidențialitate, datele noastre generate sintetic nu poartă restricții de circulație, fiind astfel adecvate pentru dezvoltarea soluțiilor AI pentru diverse nevoi în gestionarea cancerului de prostată. Am exemplificat acest lucru prin antrenarea unei rețele neuronale în atribuirea stadiului TNM pe baza informațiilor longitudinale randomizate în dosarele electronice ale pacienților. Pașii către crearea inteligenței artificiale sigure sunt prezentați în capitolul 4, unde sporim transparența a două modele de învățare adâncă specializate în (1) identificarea cancerului de prostată semnificativ clinic pe baza imaginilor bp-MRI și a parametrilor clinici și/sau demografici suplimentari și (2) clasificarea leziunilor pulmonare fără celule mici în subtipurile lor pe baza profilurilor genomice ale pacienților. Tehnicile de explicabilitate, interpretabilitate și cuantificare a incertitudinii sunt utilizate pentru a crește fiabilitatea și încrederea în modele. În cele din urmă, concluziile finale și direcțiile viitoare ale acestei lucrări sunt prezentate în capitolul 5, rezumând toate rezultatele prezentate pe parcursul tezei și impactul lor potențial în deschiderea drumului inteligenței artificiale în sănătate.

## 2. Învățare auto-supervizată pentru extrapolarea și coregistrarea imaginilor medicale

---

Introducere

Extrapolarea imaginilor CT subțiri pentru intervențiile asupra ficatului

Extrapolarea asimetrică

Extrapolarea simetrică

Concluzii

Discuții

---

### 2.1 Introducere

Dezvoltările recente în algoritmi de învățare automată (ML) au dezvăluit un potențial remarcabil în crearea de soluții utile pentru a sprijini procesele din industria sănătății, având ca scop îmbunătățirea rezultatelor actului medical, reducerea volumului de muncă a clinicienilor și optimizarea costurilor [4]. De exemplu, radiologii sunt în prezent foarte susceptibili la a face erori [35] deoarece experimentează o sarcină de muncă enormă, timpul mediu necesar pentru citirea imaginilor fiind în prezent de aproximativ 3,5 secunde [6]. Prin urmare, soluțiile bazate pe inteligența artificială ar putea fi utilizate în mod sustenabil pentru automatizarea acestui proces, reducând astfel costurile de timp, cele financiare și făcând întregul proces mai puțin predispus la erori.

Dintre toate paradigmele de învățare, cea mai consacrată rămâne abordarea supervizată, unde perechile de intrări - ieșiri sunt utilizate iterativ pentru a ghida algoritmi în modelarea caracteristicilor și corelațiilor din date. Antrenamentul se realizează prin minimizarea erorii produse de model pe baza unui set de intrări la o anumită iterație în raport cu adnotările corespunzătoare.

Cu toate acestea, adesea în practică, adunarea seturilor de date la scară largă cu adnotări de calitate este extrem de dificilă din cauza complexității, costurilor și motivelor legate de timp. De la un caz de utilizare la altul, etichetarea datelor poate necesita anumite niveluri de expertiză și cunoștințe de specialitate, solicitând astfel adnotatori instruiți sau experți clinici [13]. Complexitatea inerentă în adnotarea seturilor de date mari împiedică adesea comunitatea științifică să dispună de suficiente exemple de antrenare pentru a dezvolta în mod fiabil soluții bazate pe învățare automată pentru a rezolva diverse provocări clinice, restricționând semnificativ beneficiile potențiale pentru clinicieni și pacienții lor.

Datorită potențialului mare deținut de algoritmi de învățare automată în îngrijirea sănătății, comunitatea științifică a propus o serie de metode pentru a depăși provocările menționate anterior. Printre toate, paradigma de învățare auto-supervizată a fost sugerată pentru a produce reprezentări semnificative ale datelor neetichetate [13] care pot fi reutilizate ulterior pentru a antrena rețelele neuronale în realizarea unor sarcini relevante pe baza unei cantități limitate de date adnotate, prin ajustare. Inițial a fost concepută pentru a gestiona problemele de procesare a limbajului natural (NLP) [21], unde datele textuale sunt abundente, dar de obicei nu sunt structurate și nu dispun de etichete. Datorită impactului remarcabil al auto-supervizării în domeniul limbajului natural, a apărut un mare interes de adaptare a acestei tehnici în domeniul procesării de imagini.

Completarea imaginilor (cunoscută și sub denumirea de interpolare a imaginilor) și extinderea imaginilor (de asemenea, referită ca extrapolare a imaginilor) reprezintă exemple ale strategiei de auto-predicție. În acest capitol demonstrăm beneficiile acestei abordări prezentând o metodă de învățare auto-supervizată pentru a extrapola volume subțiri simulate care imită achizițiile CT intra-operative (iCT) pentru o coregistrare superioară cu volumele pre-operatorii de înaltă rezoluție. Pentru acest scenariu clinic, o abordare supervizată clasică nu reprezintă o opțiune, deoarece, prin natura sa, nu pot fi stabilite exemple reale de antrenare.

Acest capitol este organizat după cum urmează: Secțiunea 2.2 prezintă introducerea cazului de utilizare urmărit în tot acest capitol, adresat prin auto-supervizare, care este componenta esențială pentru fezabilitatea unei soluții bazate pe învățare automată. Secțiunile 2.3 și 2.4 prezintă și compară două abordări diferite de extrapolare, unde volumele sunt extinse asimetric, respectiv simetric. Concluziile lucrării curente sunt trase în secțiunea 2.5, în timp ce limitările și perspectivele viitoare sunt discutate în secțiunea 2.6.

## 2.2 Extrapolarea imaginilor CT subțiri pentru intervențiile asupra ficatului<sup>1</sup>

În ultimii ani, utilizarea imagisticii medicale în intervențiile asistate de calculator a devenit din ce în ce mai populară, sprijinind clinicienii în fluxul lor de lucru și reducând astfel riscurile asociate procedurilor [37].

Acest capitol este concentrat pe creșterea siguranței în intervențiile asupra ficatului, cum ar fi Ablatia prin Radiofrecvență (RFA) sau biopsia, unde imagistica în timp real joacă un rol principal în ghidarea sigură. Deși este bine cunoscut că există un compromis între doza de radiație, timpul de achiziție și calitatea imaginii, în timpul unor astfel de intervenții chirurgicale, toate procedurile trebuie efectuate cât mai rapid și precis posibil. O posibilă soluție la această problemă este achiziționarea intraoperatorie de imagini subțiri care oferă vizualizări de rezoluție scăzută ale unei mici regiuni a ficatului, urmată de coregistrarea acestora cu imagini preoperatorii complete de înaltă rezoluție [38].

Coregistrarea este o tehnică utilizată pentru a alinia două imagini în raport cu structurile interne ale pacientului. Formal, având o imagine de referință și o imagine șablon  $R, T : \mathbb{R}^d \rightarrow \mathbb{R}$ , obiectivul înregistrării este de a găsi o transformare  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  astfel încât  $R \approx T \circ \varphi$  [39]. Prin urmare, tehnicile de înregistrare sunt folosite pentru a recupera informații preoperatorii de înaltă rezoluție, cum ar fi localizarea și aspectul leziunii, și pentru a le integra cu imaginile intraoperatorii subțiri, care dezvăluie localizarea instrumentelor chirurgicale în timp real, crescând astfel încrederea în navigare. În funcție de operanzi, există multiple tipuri de coregistrare, incluzând coregistrarea între felii și volum, proiecție către volum, volum către volum, etc. [40]. Aici ne concentrăm pe acesta din urmă, cu scopul de a îmbunătăți performanța înregistrării rigide între două volume CT. Două volume pot fi coregistrate folosind o abordare bazată pe puncte cheie, o abordare bazată pe intensități sau o combinație a celor două tehnici. În coregistrarea bazată pe puncte de control, un set de repere corespondente sunt folosite pentru a calcula transformarea  $\varphi$ , cu scopul de a alinea un volum (numit volumul mobil sau șablon),  $T$ , în spațiul celui alt volum (fix sau de referință,  $R$ ) [41]. Abordarea bazată pe intensități poate fi formulată ca o problemă de optimizare, căutând cel mai bun set de parametri pentru transformarea  $\varphi$  pentru a minimiza o măsură de distanță predefinită:  $\operatorname{argmin}_{\varphi} [D(R, T \circ \varphi)]$  [39, 42]. Cu toate acestea, această abordare per se nu este robustă din cauza prezenței potențiale a minimelor locale provocate de artefactele imaginii și metricile de distanță suboptimale. Combinații ale celor două abordări pot fi folosite pentru a îmbunătăți acuratețea și robustețea coregistrării (de exemplu, utilizând alinierea bazată pe intensități ca un pas de rafinare pentru celei bazată pe repere).

După știința noastră, coregistrarea imaginilor subțiri a fost trecută cu vederea până acum. Deoarece toate tehnicile de coregistrare depind foarte mult de cantitatea de informații mutuale (date comune

---

<sup>1</sup> Această secțiune descrie experimentele prezentate în [36], reprezentând diseminări anterioare ale autorului ca parte a acestui program de doctorat.

prezentate de ambele imagini din perspective diferite), analiza imaginilor subțiri este foarte provocatoare din cauza câmpului vizual redus (FOV). Cu toate acestea, în timpul intervențiilor chirurgicale se achiziționează volume subțiri CT cu rezoluție redusă pentru a reduce riscul de expunere al pacientului. În acest context, în ciuda efectuării unei alinieri inițiale bazate pe centrul de masă sau centrul geometric, înregistrarea bazată pe intensități este predispusă la eșec din cauza câmpurilor vizuale distincte ale operanzilor. Pentru a folosi în mod robust datele preoperatorii de înaltă rezoluție, trebuie luată în considerare o abordare bazată pe puncte cheie. Totuși, algoritmi de detectare a reperelor ar putea fi, de asemenea, afectați de calitatea volumului subțire, ceea ce duce la o performanță slabă a coregistrării.

Prin urmare, propunem o metodă de extrapolare a volumelor CT subțiri, generând secțiuni suplimentare din cele câteva existente, oferind astfel informații de context îmbogățite necesare algoritmilor de înregistrare pentru funcționarea robustă.

Rețelele generative adversariale (GANs)[43] sunt o metodă de ultimă generație pentru rezolvarea sarcinilor precum generarea de imagini sintetice[44, 45, 46], segmentare [47], super-rezoluție [48], eliminarea zgomotului [49, 50], transfer de stil [51, 52] și completarea imaginilor [22, 23].

Interpolarea imaginilor, cunoscută și sub denumirea de completarea imaginilor sau „inpainting”[53], are ca obiectiv completarea regiunilor lipsă dintr-o imagine, cu conținut coerent și realist bazat pe informațiile din jur. Astfel, în interpolarea imaginilor, câmpul vizual este bine definit. În contrast, extrapolarea imaginilor[54, 55, 56] este o sarcină mai dificilă, deoarece câmpul vizual trebuie extins prin halucinaarea unui conținut coerent și realist în afara limitelor informațiilor existente.

În acest capitol, introducem o metodologie de extrapolare bazată pe o rețea generativă care crește câmpul vizual al volumelor CT intraoperatorii subțiri și îmbunătățește acuratețea și robustețea unui proces de coregistrare efectuat ulterior. Pentru a demonstra eficiența metodei propuse, ne concentrăm pe zona ficatului și presupunem că o achiziție subțire ar avea o grosime de aproximativ 5 cm. Cu toate acestea, aceasta poate fi ușor ajustată pentru alte grosimi sau scenarii de utilizare.

## 2.3 Extrapolarea asimetrică<sup>2</sup>

Prima opțiune pe care am explorat-o a fost o abordare de extrapolare asimetrică care vizează reconstrucția întregului câmp vizual al ficatului pe baza imaginii subțiri, indiferent de structurile pe care le afișează (de exemplu, zona vezicii biliare care este situată chiar sub ficat). Prin urmare, în contextul ghidării intervențiilor invazive, imaginile extrapolate afișează întotdeauna vizualizări complete ale regiunii de interes, făcând astfel ca pasul ulterior de coregistrare să se bazeze în mare măsură pe acea zonă. În funcție de anatomia tratată, această proprietate ar putea fi benefică în ceea ce privește minimizarea posibilelor artefacte (de exemplu, mișcarea, care este mai proeminentă în plămâni sau intestine comparativ cu ficatul). Cu toate acestea, extrapolarea asimetrică a volumelor generează un set de provocări care trebuie abordate din motive de utilitate, crescând în același timp complexitatea algoritmului.

În primul rând, dificultatea problemei poate fi formulată ca o funcție a distanței dintre zonele care trebuie extrapolate și zona actuală de informații. Figura 2.1 arată trei exemple de imagini subțiri sintetizate, fiecare prezentând diferite zone ale ficatului, care este complet vizibil în ultima coloană. În a doua imagine, secțiunea subțire este localizată central în cadrul casetei de delimitare a ficatului, caz în care problema extrapolării este mai degrabă simetrică - aceeași cantitate de informații trebuie generată sintetic în fiecare direcție. În schimb, imaginile 1 și 3 arată cazuri extreme, unde locația secțiunii subțiri este exact în partea superioară sau inferioară a ficatului. În acest caz, problema devine mai dificilă, deoarece trebuie să producă informații sintetice în locații relativ îndepărtate de cele existente.

Prin urmare, dificultatea problemei de extrapolare crește odată cu distanța dintre pixeli sau voxeli care sunt regresați într-o anumită locație și zona de informații adevărate. În figura 2.1, aceste distanțe sunt explicate calitativ prin lungimile săgeților roșii. De exemplu, în primul scenariu (stânga) regresia

<sup>2</sup>Această secțiune descrie experimentele prezentate în [36], reprezentând diseminări anterioare ale autorului ca parte a acestui program de doctorat.



Figura 2.1: Scenarii de extrapolare asimetrică. Primele trei coloane descriu posibile simulări ale imaginilor subțiri, în timp ce ultima coloană afișează întregul câmp vizual al ficatului.

intensităților pixelilor care reprezintă partea superioară a ficatului este relativ mai dificilă decât în al doilea scenariu (stânga-mijloc).

În plus, atunci când se utilizează rețele neuronale convoluționale (CNNs) pentru a realiza extrapolarea, fezabilitatea acestora este condiționată de alegerea arhitecturii potrivite pentru fiecare problemă specifică. În special, câmpul vizual receptiv al codificatorului trebuie să fie suficient de mare pentru a captura suficiente informații reale pentru toate zonele care trebuie completate sintetic.

Secondly, in real world applications thin slab's relative position in the liver's grid is not available (e.g. black areas in the first three columns of figure 2.1). Therefore, although this information can be calculated and/or randomized at training time in a self-supervised setting, at inference time it is impossible to establish the extent of extrapolation before feeding the thin slab to the network. Possible approaches to overcome this limitations exist, but at the cost of increasing the pipeline complexity through addition of extra processing steps.

In this section we propose a self-supervised generative-adversarial approach to increase the thin slab's field of view by means of extrapolation, of which extent is only specified at training time through a conditional discriminator. Moreover, to infer the metadata of extrapolated volumes we employ an extra-registration step allowing us compute their spatial information.

În al doilea rând, în aplicațiile reale, poziția relativă a imaginii subțiri în spațiul ficatului nu este disponibilă (de exemplu, zonele negre din primele trei coloane ale figurii 2.1). Prin urmare, deși această informație poate fi calculată și/sau randomizată în timpul antrenării într-un cadru auto-supervizat, în timpul inferenței este imposibilă determinarea extinderii extrapolării. Abordări posibile pentru depășirea acestor limitări există, dar cu costul creșterii complexității algoritmului prin adăugarea de pași suplimentari de procesare.

În această secțiune propunem o abordare generativ-adversarială auto-supervizată pentru a crește câmpul vizual al volumelor subțiri prin intermediul extrapolării, a cărei extindere este specificată doar în timpul antrenării prin valorificarea unui discriminator condițional. În plus, pentru a deduce informațiile spațiale ale volumelor extrapolate, utilizăm un pas suplimentar de coregistrare.

### 2.3.1 Metodologie

În această secțiune, introducem o abordare auto-supervizată pentru extrapolarea secțiunilor axiale, îmbogățind astfel informațiile de context necesare algoritmilor de coregistrare în obținerea unei alinieri corecte. Din cauza lipsei de date intraoperatorii reale, sintetizăm imagini subțiri prin extragerea de sub-regiuni cu grosimea de aproximativ 5 cm (vezi Secțiunea 2.3.1.1) din câmpurile vizuale complete ale CT-ului.

Conform figurii 2.2, având un volum CT  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , creem în primă fază o mască binară  $m : \mathbb{R}^d \rightarrow \{0, 1\}$  cu scopul de a șterge în mod aleatoriu 75% din conținut, generând astfel o imagine  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ . Ne referim în continuare la această imagine ca fiind imaginea grilă, care definește extinderea extrapolării. În continuare, simulăm o achiziție subțire  $t : \mathbb{R}^d \rightarrow \mathbb{R}$  prin extragerea unei regiuni de interes, iar ulterior antrenăm o rețea neuronală generativă în reconstrucția volumului inițial,



extrapolând astfel de-a lungul axei Z.

### 2.3.1.1 Setul de date

Setul de date a constat în 1400 de imagini CT de înaltă rezoluție, fiecare dintre acestea oferind o vizualizare completă a ficatului. În plus, din fiecare dintre aceste imagini am considerat doar o regiune de interes determinată de caseta de delimitare a ficatului în raport cu axa z (ne referim în continuare la aceasta ca fiind imaginea completă). Pentru a obține o bună putere de generalizare a modelului, grosimea imaginii complete este setată la înălțimea casetei de delimitare a ficatului, adăugând  $\pm 25$  mm în fiecare direcție. Toate aceste imagini au o rezoluție constantă de  $512 \times 512$  în planul x-y, cu o spațiere de 0.8 mm, în timp ce dimensiunea medie pentru axa z este de 179.2 voxeli (variiind de la 24 la 796, cu o spațiere medie de 1.49 mm). Toate imaginile au fost redimensionate la o spațiere de [3,3,1.5] mm. În plus, pentru a crea o dimensiune izotropică de  $128 \times 128 \times 128$  voxeli, s-a efectuat fie umplere (cu valoarea 0), fie decupare. Pentru a evita instabilitatea numerică atunci când calculăm varianța, am normalizat datele noastre folosind algoritmul online al lui Welford [57].

Datele au fost utilizate pentru a dezvolta un cadru de învățare auto-supervizată, creând automat perechi intrări-ieșiri din imaginile complete: în timpul antrenării, un sfert din volumul complet a fost extras aleatoriu, simulând un volum intraoperator de grosime variabilă. În continuare, o rețea neuronală profundă a fost utilizată pentru a reconstrui volumul original, extrapolând astfel volumele subțiri de-a lungul axei z.

Am împărțit aleatoriu datele într-un set de antrenare, care reprezenta 80% dintre acestea, și un set de testare care reprezenta restul de 20% din date. În plus, am folosit 100 de perechi CT pentru a cuantifica performanța coregistrării.

### 2.3.1.2 Metoda propusă

Am antrenat rețeaua noastră de extrapolare (denumită și generator) într-un cadru adversarial, optimizând-o pentru a „induce în eroare” o altă rețea neuronală (numită critic sau discriminator) în ceea ce privește autenticitatea mostrelor generate.

Rețeaua generativă realizează în primă fază o repetare a imaginii subțiri de-a lungul axei z, crescând grosimea intrării cu un factor de patru, definind astfel nivelul de extindere al extrapolării. Această repetare adaptează hărțile de caracteristici ale codificatorului la dimensiunile decodificatorului, astfel încât să putem profita de conexiunile de lungă durată care propagă informațiile prin rețea. În plus, această strategie este benefică în ceea ce privește extinderea câmpului vizual receptiv, utilizând astfel eficient cantitatea limitată de informații reale. Restul generatorului este o variație a cunoscutei rețele de tip „U-net”, unde fiecare bloc constă dintr-o secvență de straturi de convoluție, funcție de activare și normalizare la nivel de instanță [58]. În partea de codificator, reducerea dimensiunii a fost realizată folosind convoluții cu pas de 2, până când s-a obținut un câmp vizual receptiv de  $255 \times 255 \times 255$  voxeli la nivelul global. Nonlinearitățile sunt furnizate de activări de tip „LeakyReLU”, în timp ce decodificatorul utilizează activări de tip „ReLU”. Creșterea dimensiunii a fost realizată prin straturi de interpolare urmate de convoluții cu pas de 1 voxel.

Am folosit blocuri similare cu cele din generator pentru a crea un discriminator de sub-regiuni [59] condiționat pe imaginea grilă (Figura 2.2—g), care, pe lângă imaginea extrapolată, a fost furnizată ca intrare. Această imagine a ajutat criticul să penalizeze generatorul în ceea ce privește determinarea extinderii corecte a extrapolării. În loc să producă o singură valoare, criticul produce o hartă de caracteristici de dimensiuni  $8 \times 8 \times 8$ , în care fiecare element discriminează sub-regiuni de  $31 \times 31 \times 31$  voxeli din intrare.

## Strategia de optimizare

Am antrenat criticul să distingă mostrele false ( $\tilde{e}$ ) de cele reale ( $f$ ), maximizând astfel distanța Wasserstein dintre distribuția datelor reale ( $P_r$ ) și cele false ( $P_g$ )[60]:

$$L_{critic} = E_{\tilde{e} \sim P_g} [D(\tilde{e}, g)] - E_{f \sim P_r} [D(f, g)] + \lambda E_{\tilde{e} \sim P_{\tilde{e}}} [(\|\nabla_{\tilde{e}} (D(\tilde{e}, g))\|_2 - 1)^2] \quad (2.1)$$

Ecuția (2.1) descrie funcția de cost utilizată în antrenarea criticului, unde al treilea termen reprezintă o penalizare a magnitudinii gradientilor pentru a obține un proces de antrenare stabil[61].

Concomitent, am antrenat rețeaua generator să producă imagini ce nu pot fi discriminate de cele reale prin minimizarea  $L_{critic}$ , optimizând astfel:

$$L_{adv} = -E_{\tilde{e} \sim P_g} [D(\tilde{e}, g)] \quad (2.2)$$

Pentru a stimula în continuare generarea de detalii ale imaginii și structuri interne consistente, pe lângă componenta adversarială, am folosit și o penalizare a pierderii de caracteristici (feature loss)[62]. Această componentă are ca scop minimizarea distanței  $\ell_1$  între caracteristicile  $F$  extrase din mostre reale, respectiv cele false. Hărțile de caracteristici sunt furnizate de al treilea strat de convoluție al unei rețele 3D antrenate pentru segmentarea tumorilor cerebrale[63].

$$L_{feat} = E_{\tilde{e}, f} [\|F(\tilde{e}) - F(f)\|_1] \quad (2.3)$$

Așa cum este ilustrat în Figura 2.2, informațiile grilei (volumul  $g$ ) au fost folosite doar în timpul antrenării de către critic pentru a constrânge generatorul să găsească poziția corectă a volumelor subțiri în cadrul câmpului vizual ținută.

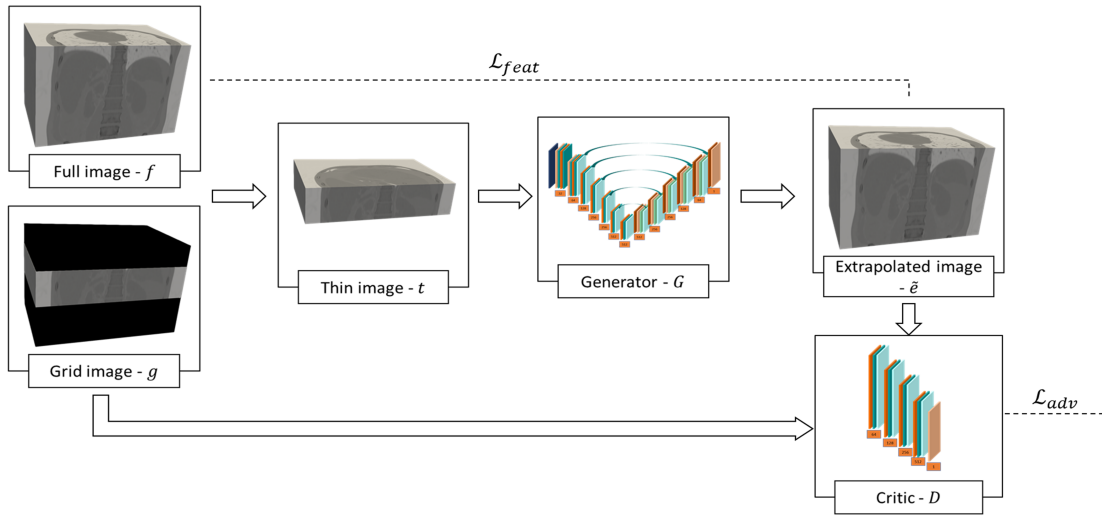


Figura 2.2: Generator optimization workflow. A conditional GAN was employed in extrapolating thin input volumes, expanding their FOV with a factor of 4.

Obiectivul generatorului reprezintă o combinație ponderată a celor doi termeni din ecuațiile (2.2) și (2.3). Ponderile au fost alese empiric astfel încât componentele să ia valori în același interval:  $\lambda_{adv} = 1$  și  $\lambda_{feat} = 1$ , ceea ce s-a dovedit a conduce la o performanță mai bună a modelului. Când s-a utilizat o pondere mai mare pentru semnalul de supervizare, așa cum este sugerat în [59], componenta adversarială a devenit instabilă în primele etape ale antrenării, împiedicând o îmbunătățire a imaginilor generate în timp.

$$L_{gen} = \lambda_{adv} L_{adv} + \lambda_{feat} L_{feat} \quad (2.4)$$

Având în vedere că funcția de cost utilizată pentru a antrena rețelele generative adversariale provine dintr-o altă rețea neuronală antrenată concomitent, funcția de cost în sine poate fi înșelătoare atunci când se încearcă identificarea celui mai performant model. Prin urmare, pentru experimentul curent, selecția modelului a fost realizată printr-o inspecție vizuală a mostrelor produse de generator în timp.

## Determinarea informațiilor spațiale

Deoarece rețelele neuronale convoluționale (CNN) operează doar pe informații de intensitate a voxelilor, a fost necesar să efectuăm un pas suplimentar pentru a recupera informațiile spațiale ale imaginilor extrapolate.

Intuitiv, imaginea extrapolată va avea aceeași dimensiune discretă și orientare ca cea subțire. Cu toate acestea, originea și dimensiunea spațială a imaginii se schimbă datorită adăugării de informații sintetice. Determinarea dimensiunii grilei volumului extins este simplă, deoarece întotdeauna extrapolăm câmpul vizual de intrare cu un factor de 4, de-a lungul axei z:

$$(d\tilde{x}, d\tilde{y}, d\tilde{z}) = (dt_x, dt_y, dt_z \times 4) \quad (2.5)$$

Pentru a calcula originea volumului extrapolat, a trebuit mai întâi să determinăm locația sub-regiunii subțiri în cadrul grilei de extrapolare. În lucrarea de față, am abordat această problemă în faza de post-procesare, efectuând un pas suplimentar de coregistrare pentru a determina extinderea extrapolării, după cum urmează: am suprapus imaginea subțire (glisând-o de-a lungul direcției z) în fiecare locație posibilă a volumului extrapolat, calculând eroarea pătratică medie la nivel de voxel (Figura 2.3 -  $d_{1..k}$ ). În continuare, am determinat extinderea extrapolării alegând indexul care a minimizat această metrică.

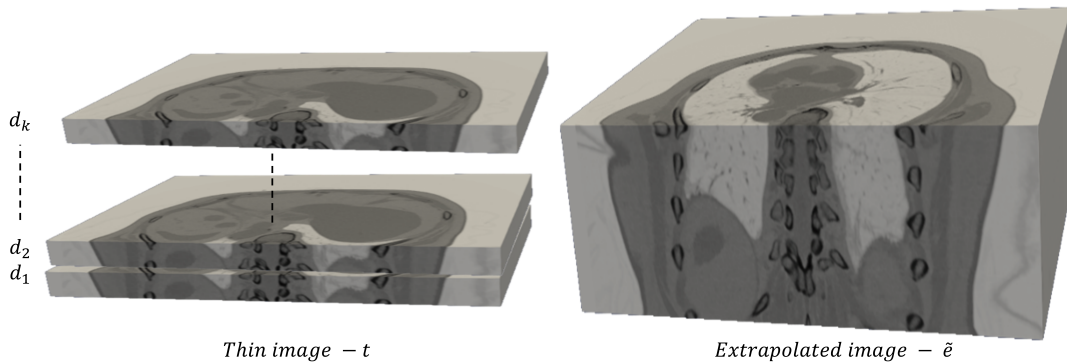


Figura 2.3: Regresia poziției; stânga - volumul subțire  $t$ ; dreapta - volumul extrapolat  $\tilde{e}$ .

În continuare, originea spațială a imaginii extrapolate a fost calculată folosind următoarea expresie:

$$(o\tilde{x}, o\tilde{y}, o\tilde{z}) = (ot_x, ot_y, ot_z - \text{argmin}_{i=1..k}(d_i) \times st_z) \quad (2.6)$$

unde  $st_z$  reprezintă spațierea volumului subțire de-a lungul axei z.

### 2.3.1.3 Cuantificarea performanței

Una dintre marile provocări în sarcinile de generare a imaginilor este lipsa unei metode standard pentru a cuantifica performanța modelelor generative. Prin urmare, propunem o metodă de cuantificare orientată spre obiective constând în două teste: detectarea reperelor [64, 65] și măsurarea erorilor de coregistrare [66].

Deoarece dorim să efectuăm o coregistrare bazată pe repere a două volume, trebuie să încurajăm detectarea precisă pe imaginile sintetice. Prin urmare, evaluăm mai întâi modelele noastre de extrapolare pe baza distanței euclidiene dintre adnotările manuale și reperele detectate pe volumele subțiri, extrapolate și complete.

Pentru testul de coregistrare, cele 100 de perechi CT suplimentare menționate în Secțiunea 2.3.1.1 au fost utilizate după cum urmează: am extras aleatoriu volume subțiri din imaginile fixe și apoi am folosit modelele noastre pentru extrapolare. În continuare, am comparat performanța între coregistrarea imaginilor mobile complete cu imaginile fixe, în 3 variante: imagini fixe subțiri, imagini fixe

complete și imagini fixe extrapolate. Am folosit două metrice pentru această evaluare: distanța dintre suprafațe și DICE, ambele calculate pe măștile binare ale ficatului, obținute folosind același model de segmentare utilizat pentru preprocesarea datelor.

## 2.3.2 Rezultate

### 2.3.2.1 Testul de detecție a reperelor

Am rulat un model de detectare a reperelor preantrenat [65] pe trei variante ale fiecărei imagini de test: completă, subțire și extrapolată. Apoi, am calculat distanța euclidiană între fiecare reper detectat și adnotarea manuală corespunzătoare. Rezultatele sunt prezentate în Figura 2.4: metoda propusă reduce eroarea mediană de detectare cu aproximativ 40% (de la 19.51 mm la 12.08 mm, valoarea  $p = 7.38e^{-37}$ ), în timp ce intervalul intercuartil (IQR) este redus la mai mult de jumătate, ceea ce înseamnă că metoda noastră crește semnificativ robustețea detectării reperelor (Tabel 2.1).

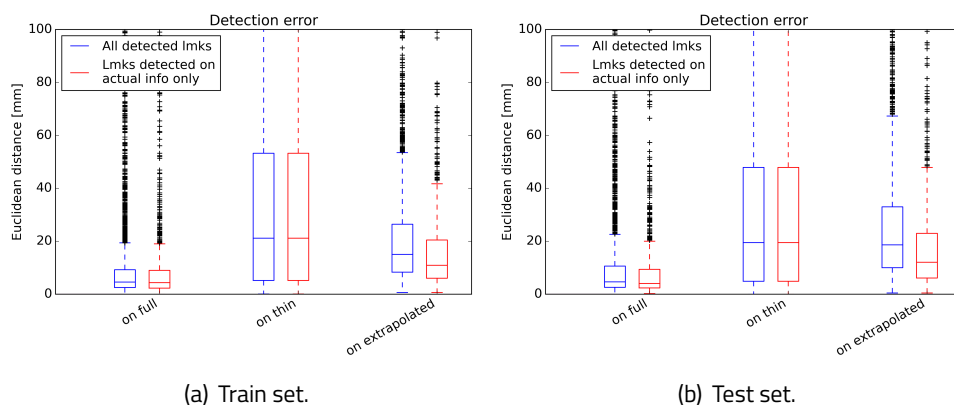


Figura 2.4: Erori de detectare a reperelor. Pentru boxplot-urile albastre, toate reperele detectate au fost luate în considerare, în timp ce boxplot-urile roșii iau în considerare doar reperele detectate în regiunea care conține informațiile reale.

Tabelul 2.1: Erorile de detecție a reperelor.

Imagini	Median ( $\pm$ IQR) [mm]	
	Toate reperele	Doar repere detectate în regiunea reală
Volume complete	4.64( $\pm$ 8.02)	4.04( $\pm$ 7.05)
Volume subțiri	19.51( $\pm$ 43.0)	19.51( $\pm$ 43.0)
Volume extrapolate	18.62( $\pm$ 22.96)	12.08( $\pm$ 16.86)

Deoarece un sfert din grosimea volumului complet este întotdeauna utilizat ca intrare, fiecare imagine extrapolată ar trebui să conțină (1) acel sfert din câmpul vizual (ne vom referi la acesta ca regiunea de informații reale) și (2) trei sferturi de informații extrapolate (halucinate). Toate reperele detectate au fost luate în considerare pentru boxplot-urile albastre, inclusiv cele detectate în regiunea extrapolată. Pe de altă parte, boxplot-urile roșii afișează eroarea de detectare doar pe informațiile reale, ceea ce este mai relevant, deoarece am folosit extrapolarea doar pentru a oferi mai mult context pentru algoritmi de detecție.

### 2.3.2.2 Testul de coregistrare

Figura 2.5 afișează rezultatele coregistrării imaginilor complete în mișcare cu toate cele trei variante ale imaginilor fixe: completă, subțire și extrapolată. Boxplot-urile albastre afișează rezultatele

coregistrării bazate pe repere, care este apoi utilizată ca inițializare pentru înregistrarea bazată pe intensități, reprezentată în roșu.

Așa cum era de așteptat, cea mai bună performanță a fost obținută atunci când imaginile complete în mișcare sunt alinate cu imagini complete fixe (având o eroare mediană de  $0.20(\pm 0.08)$  mm după coregistrarea bazată pe intensitate), iar cele mai slabe rezultate au fost obținute atunci când imaginile complete în mișcare au fost coregistrate cu imagini subțiri fixe ( $5.66(\pm 20.56)$  mm). Cu toate acestea, am obținut o performanță de coregistrare comparabilă cu cea corespunzătoare imaginilor fixe complete ( $0.57(\pm 2.05)$  mm) folosind metoda de extrapolare propusă ca pas prealabil, reducând astfel eroarea de coregistrare a imaginilor subțiri de 10 ori (valoarea  $p = 4.18e^{-6}$ ). Același lucru este valabil și atunci când se consideră scorul DICE (Figura 2.5b), care a crescut datorită extrapolării de la 0.67 la 0.88 (mediană).

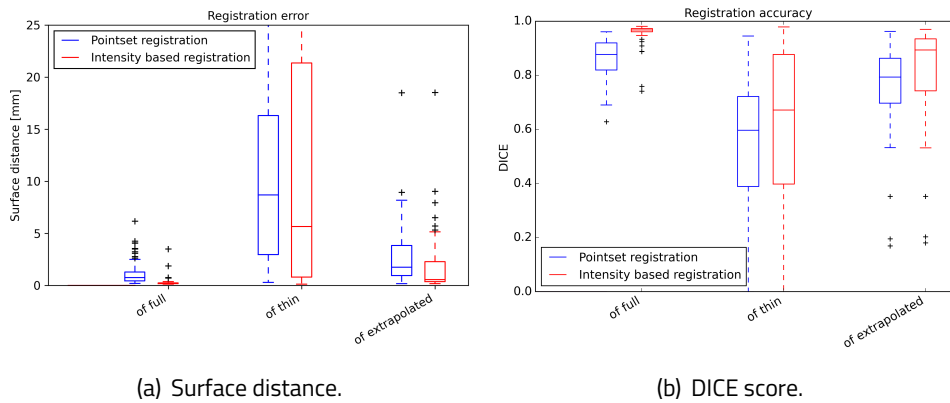


Figura 2.5: Rezultatele coregistrării: coregistrare bazată pe repere în albastru, coregistrare bazată pe intensitate în roșu. Fiecare figură are trei grupuri: stânga - coregistrarea imaginilor complete fixe cu imagini complete în mișcare; mijloc - coregistrarea imaginilor subțiri fixe cu imagini complete în mișcare; dreapta - coregistrarea imaginilor extrapolate fixe cu imagini complete în mișcare.

## 2.4 Extrapolarea simetrică

A doua strategie pe care am explorat-o a fost să formulăm sarcina ca pe o problemă de extrapolare simetrică. Ideea principală a acestei abordări este de a genera aceeași cantitate de informații sintetice în fiecare direcție de extrapolare (vezi a doua imagine din figura 2.1), minimizând astfel distanța dintre pixelii/voxelii halucinați și informațiile reale. Considerând calitatea extrapolării ca o funcție a decalajului dintre intensitățile regresate și cele existente, ne așteptăm la o performanță mai bună a acestei abordări comparativ cu cea descrisă în secțiunea 2.3. Prin urmare, câmpurile vizuale ale volumelor subțiri care descriu regiunile ficatului sunt îmbogățite, neținând cont de depășirea limitelor acestuia. Dacă un volum subțire descrie partea superioară a ficatului, extrapolarea se va extinde în zona toracică, în timp ce extrapolările volumelor care descriu partea inferioară vor expune regiunea abdominală inferioară.

În toate experimentele descrise aici, extrapolarea este concepută pentru a quadrupla întotdeauna câmpul vizual al volumelor subțiri. Prin urmare, cel mai mare decalaj posibil între zonele halucinate și cele reale este  $1.5 \times t_{thin}$ , unde  $t_{thin}$  reprezintă grosimea secțiunii subțiri. Același raționament ar putea fi aplicat problemei de extrapolare asimetrică, unde cel mai mare decalaj posibil ar putea ajunge până la  $3 \times t_{thin}$  în cazuri extreme. Generalizând exemplul de mai sus pentru orice extindere a extrapolării, distanța maximă între voxelii sintetici și cei reali pentru abordările simetrică și asimetrică ar fi  $\frac{(e_r - 1)}{2} \times t_{thin}$ , respectiv  $(e_r - 1) \times t_{thin}$ , unde  $e_r$  reprezintă gradul de extindere a extrapolării ( $e_r = 4$  pentru cuadruplarea câmpului vizual al volumului subțire). Pe lângă minimizarea nivelului de extrapolare, un alt beneficiu al acestei abordări provine din furnizarea tuturor informațiilor necesare

pentru a calcula caracteristicile spațiale ale volumelor, cum ar fi originea, nefiind necesari pași suplimentari de coreregistrare în fluxul de lucru. Evitarea pașilor suplimentari predispuși la erori crește robustețea întregului proces, ceea ce poate duce la o încredere sporită, susținând adoptarea unor astfel de soluții în rutina clinică. Prin urmare, această secțiune prezintă o metodă auto-supervizată de extrapolare simetrică care vizează îmbunătățirea performanței coregistrării prin extinderea câmpului vizual al volumelor subțiri ca o fază de preprocesare.

## 2.4.1 Metodologie

### 2.4.1.1 Setul de date

Deoarece extrapolarea simetrică necesită volume care nu sunt limitate la câmpul vizual al ficatului, pentru această abordare dispunem de un set de date relativ mai mic, de 983 de volume care descriu, pe lângă abdomenul superior, regiunile toracică și abdominală. Cu toate acestea, cadrul de învățare auto-supervizată ne permite să creăm un număr mare de exemple de antrenare prin randomizarea simultană a grosimii achizițiilor subțiri simulate și a locației exacte a acestora în raport cu întreaga casetă de delimitare a ficatului. Rândul superior din figura 2.6 arată cum volumele subțiri de grosimi diferite pot fi eșantionate din același volum complet: în timp ce abordarea asimetrică era constrânsă să producă o versiune sintetică a întregului ficat pe baza unui sfert din acesta (grosimile plăcilor variau doar pe baza anatomiei, de exemplu, înălțimea ficatului împărțită la 4), contrapartea simetrică ne-a permis să eșantionăm orice grosime dintr-o distribuție uniformă,  $t_{thin} \approx U[30mm, 50mm]$ . Al doilea punct de randomizare este locația exactă a plăcii subțiri în cadrul grilei ficatului, așa cum este ilustrat în partea de jos a figurii 2.6: o regiune subțire de o anumită grosime ar putea afișa diferite regiuni ale ficatului. Prin urmare, un număr mare de perechi intrări-ieșiri pot fi create din același volum original, permițându-ne să creăm sintetic suficiente mostre de antrenament.

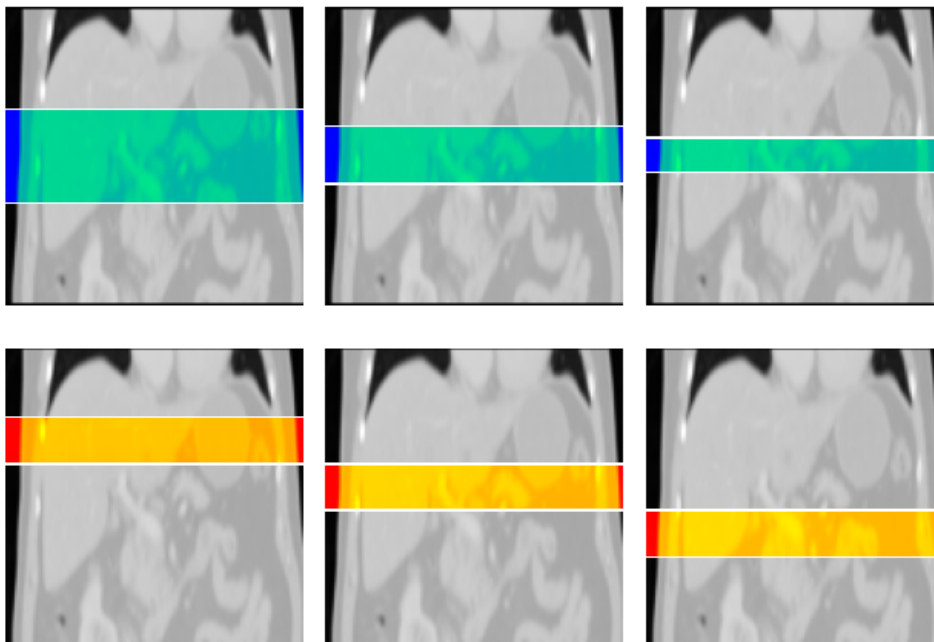


Figura 2.6: Generarea perechilor sintetice intrări-ieșiri pentru antrenarea modelului de extrapolare simetrică.

Experimentele descrise în acest capitol au folosit 5 perechi sintetice diferite intrări-ieșiri din fiecare imagine, creând astfel un set de date de 4915 exemple. Toate volumele au o dimensiune constantă a grilei de 512 în planul x-y, cu o spațiere medie între voxelii de  $0.8mm$  (variind de la  $0.45mm$  la  $0.97mm$ ). Numărul mediu de voxelii pe direcția z este de 473 (de la 59 la 2241) cu o spațiere medie de  $1.44mm$  ( $0.3mm$  la  $8mm$ ). Toate imaginile au fost re-eșantionate la o distanță constantă de

$3mm \times 3mm \times 1.5mm$  și s-a obținut o dimensiune discretă constantă de  $128 \times 128 \times 128$  pentru imaginile complete prin umplere sau decupare.

Datele au fost împărțite la nivel de cohortă în seturi de antrenare și testare, cuprinzând 80%, respectiv 20% dintre pacienți. Similar cu experimentele efectuate pentru extrapolarea asimetrică, am folosit același set suplimentar de 100 de perechi CT pentru a evalua performanța coregistrării, obținând astfel o comparație corectă a celor două abordări.

#### 2.4.1.2 Metoda propusă

În ceea ce privește cadrul de optimizare și arhitectura modelului, am folosit aceeași abordare ca în secțiunea 2.3.1.2. Cu toate acestea, determinarea informațiilor spațiale ale imaginii prezentată acolo, în special coregistrarea volumului subțiri la cel extrapolat, nu mai este necesară în contextul extrapolării simetrice, unde toate informațiile spațiale pot fi calculate în mod direct. În timp ce spațierea volumelor și dimensiunile acestora pot fi determinate direct, calcularea originii este simplă, așa cum este ilustrat în ecuația 2.7.

$$(o\tilde{e}_x, o\tilde{e}_y, o\tilde{e}_z) = (ot_x, ot_y, ot_z - \frac{gsf_z - gst_z}{2} \times st_z) \quad (2.7)$$

unde  $st_z$  reprezintă spațierea volumelor subțiri,  $gsf_z$  reprezintă dimensiunea discretă a imaginilor complete și  $gst_z$  semnifică dimensiunea discretă a imaginilor subțiri.

### 2.4.2 Rezultate

#### 2.4.2.1 Testul de detecție a reperelor

Tabelul 2.2 arată erorile mediane și intervalele intercuartile (IQR) corespunzătoare pentru reperele detectate pe 1) imagini extrapolate întregi și 2) zonele ce conțin informație reală. În comparație cu extrapolarea asimetrică, erorile de detecție pe imagini subțiri sunt mai mari, în principal datorită grosimilor distribuite uniform, variind de la 3 la 5 cm în abordarea curentă. Cu toate acestea, extrapolarea reduce aceste erori de două ori, scăzând distanța euclidiană mediană de la 21.73 la 10.86 mm, atunci când se iau în considerare doar reperele detectate pe regiunea reală. Mai mult, atunci când se consideră toți markerii corporali, indiferent de zona în care au fost detectați, eroarea mediană este încă îmbunătățită cu 30% comparativ cu detectarea pe imaginea subțiri, în timp ce extrapolarea asimetrică a produs doar o îmbunătățire de 5%.

Tabelul 2.2: Erorile de detecție a reperelor pentru extrapolarea simetrică.

Imagine	Mediana ( $\pm$ IQR) [mm]	
	Toate reperele	Doar repere detectate în regiunea reală
Volume complete	4.68( $\pm$ 8.49)	4.60( $\pm$ 9.21)
Volume subțiri	21.73( $\pm$ 41.46)	21.73( $\pm$ 41.46)
Volume extraolate	15.52( $\pm$ 21.48)	10.86( $\pm$ 16.45)

#### 2.4.2.2 Testul de coregistrare

Așa cum este ilustrat în tabelul 2.3, coregistrarea este superioară atunci când achizițiile subțiri simulate sunt extrapolate în faza de preprocesare, indiferent de tipul de coregistrare. Pentru a cuantifica acuratețea alinierii, am folosit aceeași abordare ca cea prezentată în varianta asimetrică.

Din perspectiva ambelor metrici folosite aici, distanța dintre suprafețe și DICE, volumele extrapolate sintetic oferă o calitate superioară a coregistrării. Alinierea bazată pe intensități a volumelor subțiri sunt inferioare celor extrapolate coregistrate pe baza seturilor de puncte, care, după ajustarea bazată pe intensități depășesc chiar coregistrarea bazată pe repere a volumelor complete.

Tabelul 2.3: Rezultatele testului de coregistrare pentru extrapolarea simetrică.

Volume	Pe baza de repere		Pe baza de intensități	
	SD mediană (±IQR)[mm]	DICE median(±IQR)	SD mediană (±IQR)[mm]	DICE median (±IQR)
Volume complete	0.49(±1.43)	0.85(±0.14)	0.04(±0.24)	0.93(±0.02)
Volume subțiri	11.55(±19.12)	0.52(±0.32)	10.37(±24.90)	0.54(±0.52)
Volume extrapolate	<b>1.67(±3.73)</b>	<b>0.76(±0.17)</b>	<b>0.14(±1.77)</b>	<b>0.9(±0.18)</b>

Faptul că setul de test utilizat pentru a rula analiza prezentată aici este același folosit în secțiunea 2.3 ne permite să comparăm direct cele două abordări prin intermediul metricilor de performanță a coregistrării. În ceea ce privește distribuția grosimii, setul de date folosit în experimentele de extrapolare simetrică este mai provocator, deoarece câmpul vizual al imaginilor subțiri de-a lungul axei z a fost eșantionat uniform pentru a expune regiuni între 30 și 50 mm, în timp ce volumele subțiri sintetizate în varianta asimetrică erau constrânse să arate un sfert din înălțimea ficatului, depinzând astfel în mare măsură de anatomie (grosimile urmează o distribuție Gaussiană cu un mod de 5 cm). Această observație este confirmată de rezultatele curente, volumele subțiri fiind coregistrate cu o acuratețe inferioară, indiferent de tipul de aliniere. De exemplu, abordarea bazată pe intensități a produs un scor DICE median de 0.54, comparativ cu 0.67 obținut în secțiunea anterioară. Cu toate acestea, în ciuda operării pe intrări relativ mai dificile, extrapolarea simetrică a depășit abordarea asimetrică, scăzând SD-ul median de la 0.57 la 0.14 mm și crescând scorul DICE median de la 0.88 la 0.9.

### 2.4.3 Concluziile abordărilor de extrapolare

Acest capitol a prezentat o abordare de extrapolare simetrică auto-supervizată pentru a extinde câmpul vizual al imaginilor subțiri intraoperatorii ca un pas premergător coregistrării. Comparativ cu abordarea asimetrică, problema este simplificată atunci când se extrapolează aceeași cantitate de informații în fiecare direcție prin 1) minimizarea distanței dintre voxelii ce trebuie estimați și cei reali și 2) nefiind necesare soluții alternative pentru a estima informațiile spațiale ale imaginilor extrapolate. În timp ce abordarea generală în ceea ce privește arhitectura și strategia de antrenare a rămas neschimbată, așa cum este prezentată în secțiunea 2.3.1.2, beneficiile extrapolării simetrice sunt confirmate de rezultatele curente, obținându-se îmbunătățiri în metricile specifice sarcinii evaluate.

În primul rând, în mod consecvent cu rezultatele asimetrice, metodologia propusă de extrapolare a îmbunătățit performanța și stabilitatea generală a fluxului de lucru. Acuratețea detectării reperelor este crucială pentru a oferi o aliniere inițială bună a volumelor, deoarece abordarea bazată pe intensitate este o problemă slab definită (multiplele suprapuneri ar putea produce aceeași eroare). Am arătat că extrapolarea a redus eroarea mediană de detectare de la 21.73 obținută pe volume subțiri, la 10.86 mm, încurajând astfel o aliniere inițială superioară. Indubitabil, aceste îmbunătățiri se reflectă în performanța ulterioară a coregistrării setului de puncte, crescând semnificativ scorul DICE median de la 0.52 la 0.76. Cu toate acestea, îmbunătățirile în calitatea transformărilor inițiale sunt propagate în ajustarea bazată pe intensități, care produce un scor DICE median final de 0.9, fiind relativ aproape de cel produs de volumele complete (0.93).

În al doilea rând, metoda de extrapolare simetrică propusă a redus eroarea mediană de detectare de la 12.08 la 10.86 mm atunci când sunt considerate doar reperele detectate pe regiunea reală a imaginii și de la 18.62 la 15.52 mm atunci când toți markerii corporali au fost utilizați, indiferent de regiunea de origine. Acest lucru indică o calitate mai bună a imaginilor extrapolate în setarea simetrică, obținând astfel o aliniere inițială mai bună. Orice îmbunătățire a calității coregistrării bazate pe repere are beneficii substanțiale în rezultatele ajustării bazate pe intensități, lucru confirmat de rezultatele curente, care arată îmbunătățiri în scorurile mediane ale SD și DICE, de la 0.57 la 0.14 mm și respectiv de la 0.88 la 0.9.



Prin urmare, rezultatele prezentate în acest capitol sunt extrem de încurajatoare în ceea ce privește creșterea robusteții sistemelor de ghidare a intervențiilor chirurgicale.

#### 2.4.4 Discuții asupra abordărilor prezentate

Am constatat că metodologia de extrapolare simetrică propusă în acest capitol este superioară abordării asimetrice, ducând la o detectare mai bună a reperelor și, astfel, la o aliniere inițială îmbunătățită a imaginilor, ceea ce determină pasul final de ajustare bazată pe intensități să fie mai precis. Problema extrapolării simetrice este relativ mai accesibilă comparativ cu versiunea asimetrică datorită următoarelor proprietăți:

1. Dimensiunea extrapolării este stabilită apriori, depinzând doar de factorul de extindere (de exemplu, 4 în cadrul experimentelor prezentate aici). Prin urmare, nu necesită pași suplimentari de coregistrare pentru a permite calcularea informațiilor spațiale pentru volumele extrapolate.
2. Prin natura ei, minimizează distanța dintre zonele ce trebuie estimate și regiunea de informații reale, îmbunătățind astfel consistența structurală între cadrele axiale halucinate.

În contextul ghidării procedurilor invazive, extrapolarea simetrică pare o alegere rezonabilă, deoarece imaginile intraoperatorii achiziționate vor fi întotdeauna considerate ca fiind centrul extrapolării, posibil îmbunătățind coregistrarea în timp real cu achizițiile preoperatorii. Cu toate acestea, proprietatea menționată anterior implică, de asemenea, extrapolarea în afara limitelor structurii corporale care se află sub intervenție, ceea ce poate duce la provocări suplimentare. De exemplu, de la un caz de utilizare la altul, fluxul de lucru necesită modele antrenate pentru a detecta precis reperele pe întregul domeniu de extrapolare. Disponibilitatea unor astfel de modele ar putea reprezenta o provocare, deoarece necesită seturi de date la scară largă cu markeri corporali adnotați, care sunt dificili de construit din cauza complexității, cerințelor de timp și a costurilor.

Din perspectiva modelului, arhitecturile de tip UNET au fost recent depășite de alte arhitecturi care utilizează, de exemplu, mecanismul de atenție în partea de codificare pentru a captura mai eficient corelațiile între regiuni mai îndepărtate [67]. Deși nu a fost explorat în lucrarea curentă, limitările legate de câmpul vizual receptiv al modelului ar putea fi abordate cu astfel de metode, ceea ce ar putea duce la o calitate îmbunătățită a volumelor generate. Prin urmare, utilizarea arhitecturilor bazate pe atenție reprezintă o posibilă direcție promițătoare pentru dezvoltările viitoare.

Cu toate acestea, metodele propuse în acest capitol reprezintă o dovadă clară că rețelele generative adversariale pot fi utilizate cu succes pentru a spori robustețea sistemelor de ghidare a intervențiilor, îmbunătățind semnificativ calitatea coregistrării.

## 2.5 Concluzii

Acest capitol a descris o abordare auto-supervizată ce permite dezvoltarea algoritmilor de învățare automată care vizează creșterea robusteții ghidării imagistice în intervențiile chirurgicale. În special, deoarece coregistrarea imaginilor intraoperatorii este o sarcină foarte provocatoare, acestea au fost simulate pentru a crea perechi de intrări-ieșiri, care pot fi utilizate în antrenarea rețelelor neuronale în extrapolarea volumelor intraoperatorii subțiri ca un pas premergător coregistrării.

Am cuantificat beneficiile acestei abordări prin două metrice specifice, și anume eroarea de detectare a reperelor și performanța coregistrării. Detectarea precisă a reperelor este esențială în calcularea transformărilor inițiale care pot alinia câmpurile vizuale ale celor două volume într-o măsură decentă, facilitând astfel coregistrarea rigidă bazată pe intensități. Rezultatele prezentate aici dovedesc că metoda propusă crește semnificativ robustețea coregistrării, ducând la o ghidare imagistică îmbunătățită.

Cu toate acestea, fezabilitatea acestei abordări este puternic condiționată de disponibilitatea perechilor de intrări subțiri - ieșiri complete, care nu ar putea fi produse în practică în timpul intervențiilor. Prin urmare, auto-supervizarea este componenta principală în disponibilitatea aceste

dezvoltări, permițându-ne să creăm un număr mare de exemple de antrenare, și astfel, să antrenăm modele de extrapolare a volumelor. Strategia utilizată aici a fost o schemă de auto-predicție, în care volumele complete CT au fost alterate prin eliminarea a 75% din secțiunile axiale pentru a imita o achiziție subțire, cu o grosime aleatorie de 30 până la 50 mm. În continuare, o rețea neuronală convoluțională cu o arhitectură de tip codificator-decodificator a fost utilizată pentru a restaura volumele inițiale, extrapolând astfel secțiunile axiale rămase de-a lungul direcției z.

## 2.6 Discuții

Învățarea auto-supervizată a fost componenta cheie care a permis experimentele prezentate aici. Urmând o strategie de auto-predicție, am reușit să creăm sintetic perechi de volume CT pentru a antrena o rețea neuronală profundă în extrapolarea achizițiilor subțiri. Rezultatul principal al acestei abordări este o coregistrare îmbunătățită și, prin urmare, o ghidare superioară a intervențiilor chirurgicale.

Am cuantificat îmbunătățirile prin metrici specifice sarcinilor ulterioare, raportând distanța euclidiană mediană între markerii corporali detectați pe volume și adnotările corespunzătoare, precum și distanța mediană dintre suprafețe și scorul DICE al coregistrării ulterioare. Îmbunătățirile evidente generate de abordarea propusă pot fi atribuite rețelei de extrapolare, al cărei antrenament a fost exclusiv posibil prin auto-supervizare. Cu toate acestea, deși oferă dovezi clare că metoda propusă a îmbunătățit semnificativ alinierea volumelor, analiza curentă nu a putut cuantifica impactul auto-supervizării per se, deoarece datele reale de testare care ar servi ca bază de referință sunt imposibil de obținut: câștigul în performanță ar putea fi atribuit alegerii arhitecturii, cadrului de antrenament și, în cele din urmă, modelelor de detectare a reperelor utilizate și motoarelor de coregistrare rigidă. Cu toate acestea, nicio analiză din toate acestea nu ar fi fost posibilă fără auto-supervizare, care a condus la experimentele prezentate în acest capitol.

De obicei, auto-supervizarea este utilizată în preantrenarea rețelelor neuronale profunde pe seturi de date mari, dar neetichetate, creând astfel așa-numitele modele fundamentale capabile să producă reprezentări de nivel înalt ale datelor într-un spațiu abstract. În continuare, de la un caz de utilizare la altul, poate fi folosit un transfer de cunoștințe prin ajustarea fină a acestor modele fundamentale pe cantitatea limitată de date etichetate disponibile. În general, această abordare este superioară paradigmei clasice în care modelele sunt antrenate de la zero doar pe baza unui număr relativ mic de exemple de antrenament [13]. Cu toate acestea, îmbunătățirile în performanță generate de auto-supervizare depind, de asemenea, de sarcina utilizată ca pretext în faza de preantrenare. De la o aplicație la alta, o sarcină de pretext ar putea fi preferată în detrimentul altora.

Alte abordări pentru a crește robusteții ghidării imagistice a intervențiilor ar putea fi utilizate, unde paradigma clasică de antrenare este încă o opțiune viabilă. De exemplu, modelele de detecție ar putea fi specializate în detectarea robustă a reperelor pe volume subțiri. Cu toate acestea, adnotarea seturilor mari de date în mod corespunzător ar fi extrem de dificilă, făcând astfel această abordare mai puțin scalabilă comparativ cu metoda propusă în acest capitol.

În concluzie, rezultatele prezentate în acest capitol reprezintă o dovadă clară că auto-supervizarea ar putea fi utilizată cu succes în domenii unde colectarea seturilor de date etichetate este problematică, sau chiar imposibilă. Principalele beneficii provin din facilitarea răspunsurilor rapide din partea comunității științifice la cerințele venite din partea clinică, promovând astfel desfășurarea studiilor de fezabilitate pentru diverse cazuri de utilizare, care pot avea un mare potențial pentru clinicieni și pacienții lor.

## 3. Generarea de date sintetice pentru stratificarea cancerului de prostată

---

Introducere  
Stratificarea cancerului de prostată  
Metode  
Rezultate  
Concluzii  
Discuții

---

### 3.1 Introducere

Când sunt disponibile date suficiente și de înaltă calitate, inteligența artificială este, în prezent, cea mai promițătoare abordare pentru rezolvarea oricăror probleme emergente, având astfel un potențial extrem de ridicat în creșterea calității vieții, îmbunătățirea proceselor, reducerea costurilor etc. Cu toate acestea, colectarea și explorarea seturilor de date calitative la scară largă poate fi provocatoare în unele industrii din cauza unei serii de preocupări specifice. De exemplu, informațiile medicale sunt protejate pe scară largă de GDPR în Europa și HIPAA în SUA pentru a menține confidențialitatea pacienților. Ca urmare, datele din domeniul sănătății suferă adesea de incompletitudine, calitate slabă sau insuficientă din cauza constrângerilor de confidențialitate [12, 68], împiedicând dezvoltarea efektivă a soluțiilor bazate pe învățare automată să fie adoptate în rutinele clinice[4].

Generarea de date sintetice (SGD) este una dintre cele mai promițătoare abordări pentru a depăși provocările impuse de constrângerile de reglementare în accesarea datelor clinice restricționate. Un set de date sintetice calitative nu conține nicio indicație despre o persoană reală, menținând în același timp distribuția parametrilor realistă, precum și corelațiile naturale între caracteristici. Un set de date sintetic care imită și păstrează proprietățile statistice ale cohortelor reale ar putea fi utilizat pentru modelare, scopuri educaționale, simulare și predicție, testarea ipotezelor și algoritmilor, dezvoltări în tehnologia informației (IT) etc. [20, 70].

Prin urmare, acest tip de date poate fi partajat pe scară largă cu părți terțe pentru a permite dezvoltări rapide în domeniul sănătății, ceea ce ar putea duce la o îngrijire mai bună pentru pacienți și, de asemenea, la o reducere a volumului de muncă a clinicienilor. Cu toate acestea, deși dezvoltarea eficientă a unor modele de încredere pentru a fi adoptate în practica clinică este puternic condiționată de calitatea și realismul datelor sintetice, de obicei există un compromis între proprietățile de protecție a confidențialității și utilitate. Seturile de date pur sintetice, care sunt complet lipsite de amenințări legate de confidențialitate, au de obicei o utilitate mai redusă în contrast cu seturile de date sintetice care încă păstrează părți din informațiile reale, dar care prezintă riscuri de re-identificare a pacienților.

Deși unii oameni de știință consideră că termenul "sintetic" ar trebui să se refere exclusiv la date complet fabricate [20], SGD ar putea fi împărțit în trei categorii principale [18, 19]:

- Datele complet sintetice se referă la date fabricate integral, care nu au nicio legătură cu indivizii, deci neavând riscuri de încălcare a confidențialității.

- Datele parțial sintetice folosesc informații nesensibile din exemple reale, în timp ce valorile parametrilor care ar putea sprijini identificarea subiecților sunt generate sintetic.
- Datele sintetice hibride se referă la potrivirea unui înregistrări reale cu una complet fabricată prin instrumente statistice și apoi combinarea caracteristicilor din cele două înregistrări pentru a crea un eșantion sintetic.

În domeniul sănătății, generarea de date sintetice a fost utilizată cu succes pentru o varietate de tipuri și modalități de date, inclusiv imagini [71, 72], text [73, 74], dosare electronice de sănătate (EHRs) [69, 75], electrocardiografe (ECG) [76, 77] și chiar genomică [78, 79] (cu anumite limitări [80] în ceea ce privește protejarea confidențialității și compromisurile de utilitate). Alte studii au avut ca scop crearea de imagini medicale sintetice de anumite modalități pe baza anatomiei relevate de o tehnică de imagistică diferită [81] prin transfer de stil [82].

Synthea [69] este tehnologia de vârf actuală în generarea de dosare medicale electronice complet sintetice și realiste, oferind date clinice fabricate de înaltă calitate complet lipsite de riscuri legate de confidențialitate. Metoda utilizează ghiduri generale de îngrijire clinică și informații de sănătate publică, cum ar fi incidența bolilor și diferite atribute statistice, pentru a crea module de boli clinice care generează populații sintetice.

În acest capitol propunem o metodologie bazată pe Synthea pentru a crea module de boli clinice. Scopul nostru este de a genera dosare electronice de sănătate realiste pentru pacienții care suferă de cancer de prostată, variind de la boli cu risc scăzut până la risc ridicat, tratate corespunzător cu diverse opțiuni de tratament care fluctuează de la monitorizare activă, la terapie definitivă, sau chiar îngrijire paliativă. Un set de date sintetic, generat corect, care imită distribuțiile și corelațiile datelor reale între diferiți parametri, are un potențial extraordinar în a permite dezvoltarea rapidă a aplicațiilor medicale menite să sprijine calitatea serviciilor medicale. Prin urmare, fezabilitatea datelor generate în acest fel este evaluată prin utilizarea rețelelor neuronale profunde pentru stratificarea pacienților cu cancer de prostată în categorii de risc, atribuind un stadiu al bolii pe baza informațiilor clinice prezentate în dosarele medicale electronice complet sintetice.

Acest capitol este organizat după cum urmează: Secțiunea 3.2 prezintă o introducere generală a scenariului clinic urmărit pe parcursul acestui capitol, subliniind importanța automatizării fiabile în atribuirea unui stadiu pentru cancerul de prostată. În continuare, secțiunea 3.3.1 oferă detalii despre generatorul de date sintetice propus, împreună cu o descriere a conceptelor și elementelor de date disponibile în cohorta rezultată și o analiză a fidelității datelor sintetice. Secțiunea 3.3.2 descrie o abordare de procesare a limbajului natural (NLP) pentru a atribui stadiile TNM pacienților pe baza descrierilor codurilor medicale și observațiilor rezultate în dosarele medicale sintetice. Rezultatele acestei abordări sunt prezentate în secțiunea 3.4.2, în timp ce concluziile și discuțiile generale sunt prezentate în secțiunile 3.5 și 3.6.

## 3.2 Stratificarea cancerului de prostată

Cancerul de prostată este foarte comun în rândul bărbaților la nivel mondial. De exemplu, 1 din 8 bărbați din SUA va fi diagnosticat cu cancer de prostată în timpul vieții sale. Se estimează că vor exista aproape 250.000 de noi cazuri de cancer de prostată în Statele Unite în acest an [83], similar cu incidența cancerului pulmonar și mamar.

Datele agregate din Europa sunt foarte similare cu cele din Statele Unite [84], iar incidența și mortalitatea cancerului de prostată sunt în creștere în Asia. La nivel mondial, cel puțin 360.000 de bărbați mor de cancer de prostată în fiecare an [85].

„National Comprehensive Cancer Network” (NCCN) este o alianță formată din 31 de centre oncologice care publică și actualizează periodic orientări privind tratamentul cancerului, care sunt acum acceptate ca recomandări de ultimă oră în majoritatea centrelor de tratament al cancerului. Recomandările NCCN privind tratamentul cancerului de prostată depind de care dintre cele cinci grupuri de risc de cancer de prostată descrie cel mai bine un anumit pacient. Aceste grupuri de risc sunt definite

în funcție de stadiul clinic (inclusiv constatările radiologice), markerii tumorali din sânge (de exemplu, antigenul specific prostatei - PSA), gradul histologic al tumorii determinat în urma unei biopsii și, mai recent, subtipul (subtipurile) genetic(e).

Stadializarea clinică surprinde cantitatea și răspândirea cancerului în anatomia unui pacient. Stadializarea constă de obicei în trei componente, T, N și M, denumite sistemul TNM. T descrie dimensiunea tumorii și orice răspândire a cancerului în țesutul din apropiere; N descrie răspândirea cancerului în ganglionii limfatici din apropiere; iar M descrie metastazele (răspândirea cancerului în alte părți ale corpului). Acest sistem a fost creat și este actualizat de American Joint Committee on Cancer (AJCC)[86] și International Union (UI)[87].

Stadializarea clinică este adesea determinată pe baza mai multor teste de diagnostic clinic care sunt administrate longitudinal. Atribuirea unui stadiu clinic este, în general, reprezintă o sarcină dificilă și consumatoare de timp, deoarece constatările pertinente din istoricul și examinarea fizică a pacientului, precum și interpretarea radiografică sunt, de obicei, înregistrate într-un text clinic nerestricționat, care ar putea include o sugestie privind stadiul clinic al pacientului exclusiv pe baza rezultatelor testelor de diagnostic specifice. Din acest motiv, ambiguitățile de stadializare nu sunt neobișnuite[88] și sunt de obicei rezolvate de un consiliu tumoral instituțional (TB).

De obicei, sistemele care extrag stadializarea clinică din fișele medicale fac acest lucru prin combinarea concluziilor diferiților medici care se ocupă de pacient, care sunt adesea afectate de ambiguități și declarații contradictorii, precum și de o examinare incompletă la acel moment. Sistemul nostru diferă de altele prin faptul că nu extrage concluziile privind stadializarea ale diferiților medici care se ocupă de pacient și nu își bazează rezultatele pe acestea. Mai degrabă, acesta deduce stadiul adecvat din notele clinice și radiologice originale. Pentru a face acest lucru, utilizăm o rețea neuronală, bine antrenată pentru această sarcină.

### 3.3 Metode

Constrângerile de confidențialitate limitează accesul la datele clinice longitudinale, necesare pentru formarea sistemelor de inteligență artificială (AI). În plus, instituțiile devin din ce în ce mai restrictive în a permite utilizarea datelor clinice pentru cercetare în general. Chiar dacă accesul la datele retrospective anonimizate ale pacienților este permis, acesta este adesea limitat din cauza lipsei consimțământului pacientului, care este solicitat de majoritatea consiliilor de revizuire instituționale (IRB). Aceste măsuri necesare pentru a proteja confidențialitatea pacienților fac dificilă formarea sistemelor AI care necesită în mod inerent cantități mari de date clinice etichetate.

Pentru a depăși această problemă, am conceput o abordare pentru a genera și utiliza un set de date sintetice de 10000 de înregistrări pentru antrenare și 4000 de înregistrări pentru testare, așa cum este descris în secțiunile următoare.

#### 3.3.1 Generarea de date sintetice

Am utilizat cadrul Synthea și am adăugat module specifice cancerului de prostată pentru a crea date sintetice coerente. Pentru crearea modulelor sunt necesare două intrări: hărți de îngrijire clinică și statistici privind boala. Îndrumați cu atenție de un oncolog cu experiență, am întocmit hărți de îngrijire clinică pentru diagnostic, terapie localizată și avansată, precum și pentru monitorizarea ulterioară.

### Recomandări pentru diagnosticarea cancerului de prostată

Figura 3.1 prezintă o imagine de ansamblu a procesului de diagnosticare, care este o combinație de stări stohastice și deterministe. Stările stohastice surprind variațiile posibile care sunt așteptate în datele clinice reale, în timp ce stările deterministe iau în considerare reguli bazate pe practica obișnuită de ghidare a actului medical, de exemplu, metodologia de diagnosticare pentru stadializarea

clinică. În final, stadiul clinic al cancerului și categoria de risc sunt stabilite pe baza rapoartelor de diagnosticare generate în mod stocastic pe parcursul examinării. Am utilizat distribuții statistice pentru eșantionarea rapoartelor de diagnostic care au fost colectate fie din literatură [89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105], fie derivate din studiul de screening al cancerului de prostată, pulmonar, colorectal și ovarian (PLCO)[106], care relevă informații longitudinale de screening, cum ar fi nivelurile de antigen specific prostatei (PSA), rezultatele examinărilor rectale digitale (DRE), gradul histologic (scorul Gleason) și informațiile privind stadializarea[107].

Tabelul 3.1: Proprietățile setului intern de date.

Parametru	Media ( $\pm$ std)	Minimul	Maximul
Vârsta [ani]	66.3 ( $\pm$ 7.75)	43	89
PSA [ng/mL]	20.3 ( $\pm$ 69.3)	0.29	1545
Volumul [cc]	42.2 ( $\pm$ 22.82)	0.57	164
PI-RADS	4.2 ( $\pm$ 0.88)	2	5
Grupul Gleason	2.3 ( $\pm$ 1.42)	1	5

În plus, am utilizat un set intern de date pentru a obține proprietăți statistice care nu au putut fi obținute din PLCO (de exemplu, scorurile PI-RADS), care conține 768 de adnotări ale scorului PI-RADS, grupului Gleason, PSA-ului, volumul glandei și vârstei. Tabelul 3.1 prezintă o descriere a acestei colecții, denumită în continuare set de date intern.

Procesul de diagnosticare a pacientului simulat începe întotdeauna cu un examen rectal digital și o măsurare a PSA-ului. Mai întâi, o valoare pentru PSA este eșantionată dintr-o distribuție derivată din datele de screening PLCO utilizând metoda „Kernel density estimation” (KDE). Pentru măsurătorile preoperatorii ulterioare PSA-ului, eșantionarea se face în așa fel încât să asigure o dinamică realistă. În mod similar, unele dintre rezultatele DRE, precum estimările de volum, sunt eșantionate în mod consecvent cu vârsta pacientului și nivelurile PSA, în urma analizei pe subgrupuri prezentată în [108]. Orice anomalie identificată în cadrul unuia dintre aceste teste de screening inițiale declanșează proceduri de diagnosticare ulterioare, cum ar fi imagistica sau biopsia. În schimb, în cazul în care nu apar rezultate suspecte, screeningul va continua, repetând măsurătorile PSA și/sau procedura DRE la aproximativ 1 an, interval de screening recomandat pentru cancerul de prostată[95].

Pentru a obține coerență, toate rapoartele medicale ulterioare sunt eșantionate pe baza corelațiilor expuse în literatura de specialitate între procedura (sau testul) de diagnosticare curentă și cele sintetizate anterior. De exemplu, în faza de screening, toate măsurătorile ulterioare ale PSA vor fi constrânse de cele anterioare, pe baza dinamicii PSA citeDamico2004, Loeb2011 identificate în diferite categorii de risc, sau a timpului de dublare a PSA [101]. În mod similar, pentru pacienții care au nevoie de proceduri suplimentare de diagnosticare, constatările se bazează fie pe descrierile din literatura de specialitate, fie pe statisticile derivate din seturi de date reale, după cum se descrie în continuare.

În funcție de măsurătorile PSA deja sintetizate, unui pacient i se pot recomanda proceduri de diagnosticare suplimentare - cum ar fi RMN și/sau biopsie - în cazul unor rezultate anormale (de exemplu, PSA > 4 ng/mL sau un DRE anormal), sau continuarea screeningului dacă nu apar anomalii (de exemplu, DRE negativ și o valoare scăzută a PSA). În caz de suspiciune de cancer de prostată prin screening, pacientul va fi supus în mod aleatoriu unei biopsii sistematice sau unui RMN multi-parametric. În primul caz, rezultatele biopsiei vor fi eșantionate în conformitate cu singura informație cunoscută din dosarul electronic de sănătate, și anume valorile PSA. Figura 3.2 prezintă distribuția gradului Gleason în funcție de nivelul PSA discretizat, astfel cum rezultă din setul de date PLCO. Acest peisaj distribuțional este încorporat în modulele propuse de noi, permițându-le să genereze cohorte care imită proprietățile statistice ale datelor clinice reale.

În mod similar, am utilizat aceeași metodologie pentru a eșantiona un scor PI-RADS în funcție de nivelurile PSA discretizate atunci când un pacient este randomizat pentru a fi investigat prin imagistică, mai degrabă decât prin biopsie. Cu toate acestea, leziunile maligne evidențiate prin mpMRI (PI-RADS  $\geq$ 3) trebuie confirmate prin biopsii țintite, prin atribuirea consecventă a unui grad Gleason

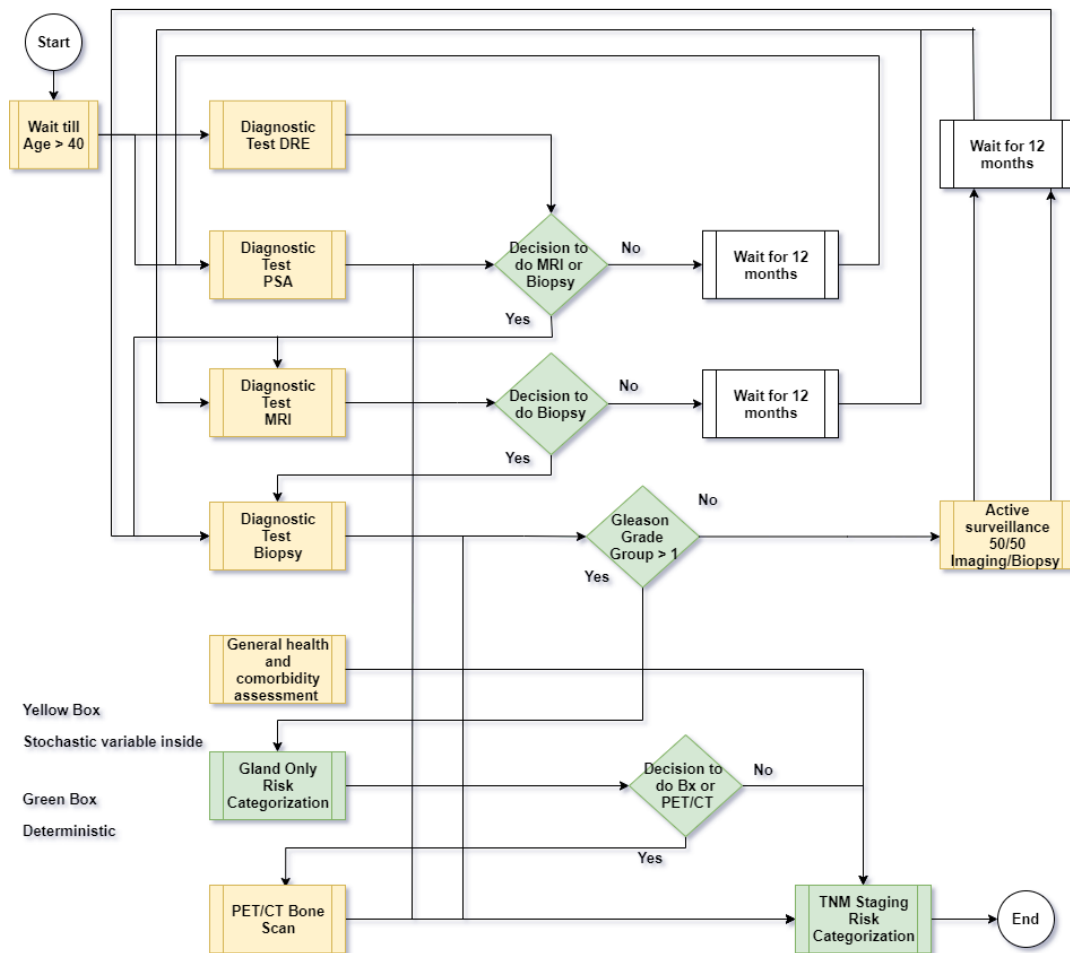


Figura 3.1: Diagnosticarea cancerului de prostată: Stările galbene au în interior variabile stohastice, în timp ce cele verzi sunt deterministe. Abrevieri: DRE, examen rectal digital; PSA, antigen specific prostatic; RMN, imagistică prin rezonanță magnetică; BX: biopsie; PET, tomografie cu emisie de pozitroni; CT, tomografie computerizată;

fiecărei zone suspecte expuse prin imagistică. În afară de randomizarea inițială, un mpMRI ar putea fi utilizat în alte scenarii, cum ar fi urmărirea unei biopsii sistematice negative, caz în care un scor PI-RADS este selectat pe baza cotelor derivate pentru fiecare grup Gleason separat. Deoarece scorul PI-RADS are o tendință de supradiagnosticare (sensibilitate ridicată cu specificitate relativ scăzută), este posibil ca acesta să dezvăluie leziuni foarte suspecte chiar și la pacienții cu un grup Gleason de 1, atribuit sistematic. În această măsură, o biopsie țintită ar putea fi utilizată în stabilirea unui diagnostic final, dezvăluind histopatologia zonelor suspecte.

La un nivel mai granular, numărul de probe pozitive rezultate din biopsie este eșantionat pe baza distribuției cohorței prezentate în Vallette et. al.[90], în timp ce locațiile corespunzătoare sunt alese aleatoriu. Pentru fiecare locație pozitivă, un indicator de răspândire a cancerului, și anume procentul de cancer din probă, este eșantionat utilizând proprietățile statistice din [91]. Atunci când un grup Gleason este eșantionat conform figurii 3.2, cel puțin una dintre leziuni i se va atribui un scor Gleason corespunzător, în timp ce altele ar putea fi la fel sau mai puțin agresive.

Toate rezultatele imagistice (de exemplu, mpMRI sau PET/CT) sunt corelate cu rapoartele de diagnostic anterioare prin intermediul tabelor Partin[92], oferind scoruri de risc pentru extensiile extra-capsulare (ECE), invaziile veziculei seminale (SVI) și implicarea ganglionilor limfatici (LN+) stratificate pe baza valorilor PSA, a scorurilor Gleason și a stadiului clinic T. Prin urmare, eșantionăm în mod aleatoriu caracteristicile pe baza probabilității de boală delimitată în capsula prostatei, în conformi-

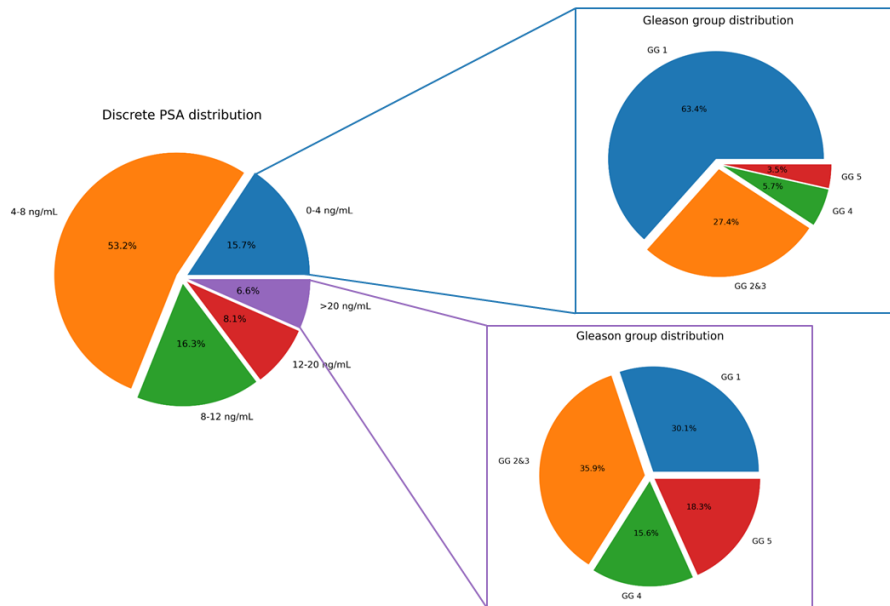


Figura 3.2: Distribuția gradului Gleason în funcție de intervalele PSA discrete derivate din setul de date PLCO.

tate cu seturile de date din lumea reală. În cazul în care imagistica relevă o boală nelocalizată, se efectuează o biopsie de confirmare pentru a identifica posibilele rezultate fals-pozitive ale PET/CT [93, 94].

Prin urmare, datele sintetice propuse de noi imită modelele și corelațiile din lumea reală, acoperind o gamă largă de variabilități, de la pacienții cu niveluri scăzute de PSA și grade Gleason ridicate la pacienții cu niveluri ridicate de PSA și grade Gleason scăzute.

Atunci când toate procedurile de diagnostic necesare sunt finalizate, aplicăm recomandările NCCN pentru a stabili stadiul clinic al cancerului și categoria de risc pe baza rapoartelor clinice eșantionate aleatoriu și exportate pe parcursul stadializării, creând astfel valori de adevăr pentru modelele de stratificare.

Merită subliniat faptul că traseele pacienților sunt foarte diverse datorită proceselor stohastice impuse în majoritatea etapelor (casetele galbene din figura 3.1). De exemplu, faza de diagnosticare ar putea produce pacienți care au fost evaluați doar prin măsurători repetate ale PSA și DRE (dacă nu există rezultate suspecte, sau nivelurile PSA sunt foarte scăzute), pacienți care au fost evaluați prin imagistică cu sau fără proceduri ulterioare în funcție de rezultate, pacienți care au trecut prin biopsii sistematice sau direcționate, pacienți care au urmat prima linie de tratament direct după o suită de proceduri de diagnostic sau pacienți care au fost monitorizați activ înainte de terapia definitivă.

### 3.3.2 Predicția stadiului TNM clinic

În timp ce datele sintetice descrise în secțiunea anterioară prezintă o variabilitate și o coerență bune la nivel de cohortă, la nivel de pacient acestea prezintă încă unele dezavantaje. EHR-urile generate de Synthea sunt foarte bine structurate și utilizează mereu aceeași codificare. Cu toate acestea, datele din lumea reală nu sunt întotdeauna atât de organizate, diferite instituții pot utiliza sisteme de codificare diferite și, prin urmare, anumite informații pot fi indicate prin coduri multiple, nume sau chiar ortografii diferite. Așadar, o simplă încercare de a extrage toate ingredientele necesare stadializării ar putea eșua. Pe de altă parte, în ciuda acestei variabilități a sistemelor de codificare, descrierile codurilor aceluiași element de date au întotdeauna un anumit grad de suprapunere. Încercăm să exploatăm aceste elemente comune prin utilizarea unui model de prelucrare a limbajului natural care operează direct pe textul simplu din descrieri, mai degrabă decât pe codurile structurate dintr-un



EHR.

Mai întâi trebuie să traducem resursele pacienților (de exemplu, reprezentate în standarde FHIR sau HL7 - „Health Level Seven”) în text clar. Prin urmare, extragem descrierile codurilor procedurilor de diagnosticare împreună cu data la care au fost efectuate. Extragem apoi toate observațiile înregistrate în cadrul acestor proceduri de diagnostic prin intermediul unui extractor de perechi cheie-valoare bazat pe reguli, în care cheia este descrierea codului observației, iar valoarea este reprezentată din constatări. În cele din urmă, concatenăm textul pentru a obține un paragraf care să cuprindă toate ingredientele generate pe parcursul stagiului de lucru sub formă de text clar, păstrând, de asemenea, dimensiunea temporală a datelor longitudinale. Pentru a spori și mai mult variabilitatea textului, augmentăm setul de date prin înlocuirea aleatorie a anumitor cuvinte sau fraze cu sinonime.

### 3.3.3 Descrierea experimentală

În faza de antrenare, utilizăm resursele exportate de Synthea ca set de date adnotate pentru învățarea supravegheată. După cum se arată în figura 3.3, datele fiecărui pacient sunt preprocesate de un program extractor de perechi cheie-valoare bazat pe reguli pentru a crea un bloc de text care descrie toate procedurile de diagnosticare împreună cu rezultatele acestora. Antrenăm sistemul de clasificare a stadiului clinic ca o sarcină de ajustare prin transfer de cunoaștere pentru un model BERT („Bidirectional Encoder Representations from Transformers”) [21], pornind de la modelul preantrenat clinic [109]. BERT este un model lingvistic bidirecțional preantrenat în furnizarea de reprezentări contextuale pe 2 sarcini de antrenare nesupervizată: modelarea limbajului mascat (din eng. „Masked language modelling” - MLM) și predicția propoziției următoare (din eng. „next sentence prediction” - NSP).

În continuare, ajustăm ponderile modelului pentru 5 epoci - hiper-parametru determinat empiric din curbele de învățare - pentru a minimiza o funcție de cost de tip entropie categorică, utilizând o rată de învățare de  $10^{-5}$ . Deși nu este necesară în majoritatea cazurilor, lungimea secvenței de intrare a BERT este setată la 512 cuvinte pentru a maximiza șansele de a capta toate informațiile necesare pentru atribuirea stadiului clinic. Deoarece memoria GPU necesară pentru procesul de antrenare este disproporționat de mare în funcție de lungimea secvenței, dimensiunea maximă a pachetelor de antrenare pe care o putem seta este 16, pe un GPU NVIDIA Volta V100 cu 32 GB de memorie. Pentru a evita dezechilibrul sever al claselor (de exemplu, pacienții T1cM1 - cei cu boală metastatică - fiind relativ foarte rari în comparație cu pacienții T3M1), am antrenat clasificatori diferiți pentru fiecare componentă de stadializare, dar ne referim la aceștia ca la un singur predictor pentru simplitate. Cu toate acestea, deoarece datele noastre sintetice sunt în concordanță cu statisticile cohortelor reale extrase din literatura de specialitate, ne confruntăm în continuare cu distribuții de clasă dezechilibrate în cadrul fiecărei componente de stadializare. Prin urmare, utilizăm ponderi în cadrul funcției de cost pentru a oferi semnale mai puternice pentru eșantioanele din clasele subreprezentate.

Odată antrenate, modelele pot fi utilizate pentru inferență, după cum se arată în partea de jos a figurii 3.3. Dosarul electronic medical al pacientului, stocat în baza de date FHIR a spitalului, va fi convertit în text clar cu ajutorul modulului de preprocesare și apoi trecut prin model pentru a obține o estimare a stadiului clinic, care poate fi apoi introdus în mod corespunzător înapoi în sistemul informatic.

## 3.4 Rezultate

### 3.4.1 Fidelitatea datelor sintetice

Pentru a evalua plauzibilitatea datelor generate sintetic, am comparat diverse caracteristici pur sintetice cu omologii din lumea reală disponibili în PLCO, precum și cu setul de date intern descris în tabelul 3.1.

În primul rând, am comparat distribuțiile măsurătorilor PSA și ale volumelor glandelor generate de modulele noastre Synthea cu cele înregistrate în setul intern de date. După cum este ilustrat în

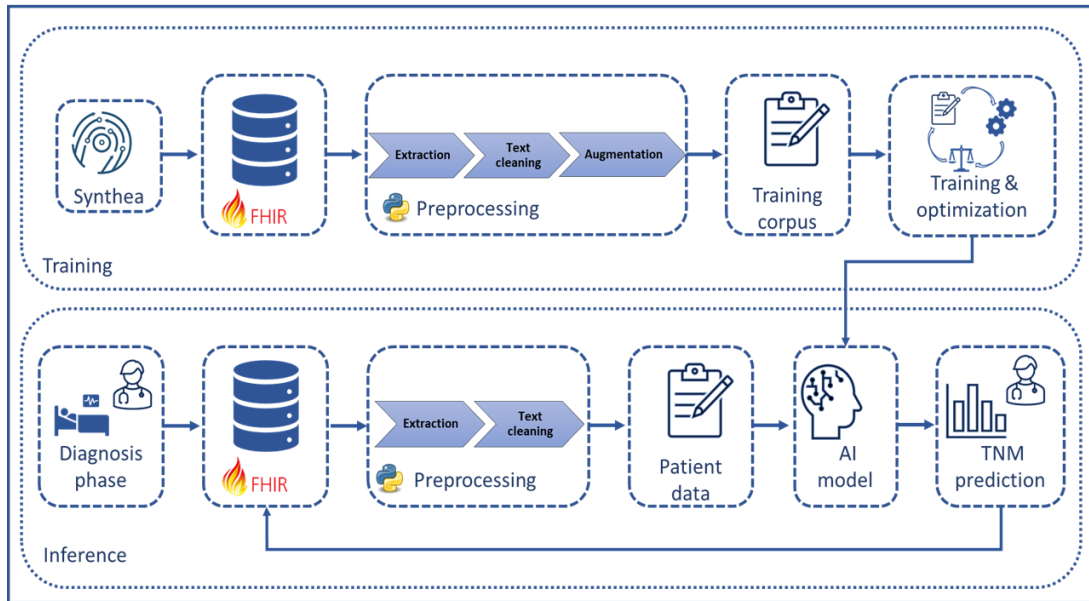


Figura 3.3: Prezentare generală a metodei: Faza de antrenare (sus) și cea de inferență (jos).

figura 3.4, în ciuda faptului că valorile PSA au fost eșantionate pe baza PLCO, iar volumul prostatei a fost randomizat în funcție de PSA și vârstă [108], există o asemănare distributivă foarte puternică între măsurătorile sintetice și cele reale ale PSA (figura 3.4(a)), precum și între volumele prostatei (figura 3.4(b)).

Cu toate acestea, evaluarea similarității doar dintr-o perspectivă univariată ar putea induce în eroare, deoarece datele ar trebui să fie coerente între diferitele teste de diagnostic, prezentând corelații plauzibile între diferiți parametri. În special, este bine cunoscut faptul că nivelurile PSA sunt corelate cu volumele prostatei, glandele mai mari tinzând să producă mai mulți antigeni. Prin urmare, figura 3.4(c) demonstrează coerența între cele două caracteristici menționate anterior, prezentând, de asemenea, distribuții similare ale densității PSA, care este calculată prin împărțirea nivelului PSA la volumul prostatei.

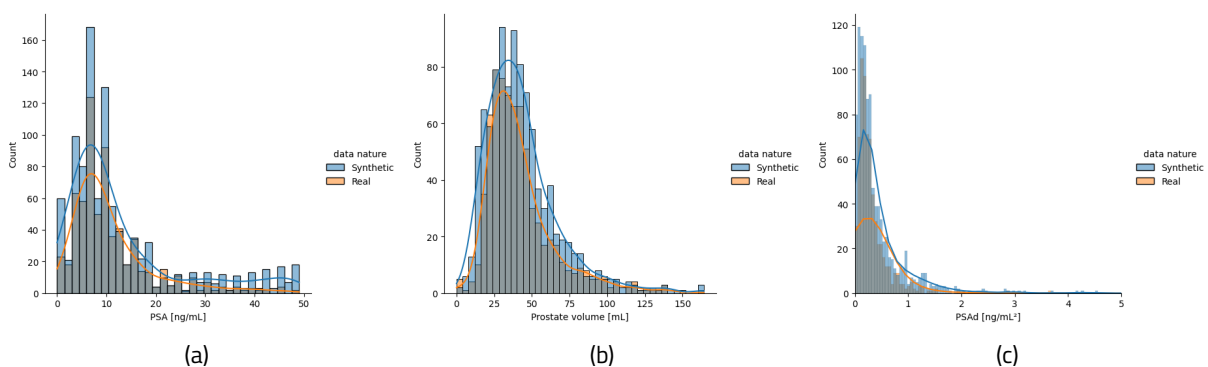


Figura 3.4: Evaluarea plauzibilității PSA, a volumului glandei și a densității PSA.

În mod similar, am studiat coerența gradelor Gleason (GG) generate sintetic și distribuțiile PI-RADS dintr-o perspectivă univariată și, respectiv, bivariată. Din perspectiva prevalenței, am putut concluziona că datele noastre sintetice au o fidelitate perfectă.

Sistemul PI-RADS a devenit principalul instrument utilizat în reducerea numărului de biopsii inutile, dar are încă o rată de specificitate relativ scăzută în comparație cu sensibilitatea sa. Cu alte cuvinte, acesta tinde să solicite în exces biopsii la pacienții cu cancer de prostată nesemnificativ din

punct de vedere clinic. Această capacitate predictivă inerentă a scorurilor PI-RADS este bine reflectată în datele noastre generate sintetic, demonstrând încă o dată gradul de fidelitate al rezultatelor longitudinale.

Toate procesele stohastice utilizate în întregul parcurs de stadializare se bazează fie pe statistici derivate din seturi de date reale, fie extrase din descrieri bibliografice ale anumitor comportamente și corelații între parametri. Prin urmare, nu se poate aștepta o fidelitate perfectă față de o anumită sursă. Cu toate acestea, deoarece generatorul nostru de date este proiectat într-un cadru causal puternic (rapoartele la un anumit moment sunt întotdeauna eșantionate în mod consecvent față de constatările randomizate în momentele anterioare), peisajul general al cohortei noastre sintetice este de așteptat să corespundă într-o anumită măsură distribuției datelor clinice reale. Stadiul TNM rezultat în urma stadializării oferă o intuiție foarte bună a modului realist în care diferitele rapoarte au fost randomizate în timpul stadializării, deoarece se bazează pe mai multe rezultate cheie, fiecare eșantionat pe baza unei surse statistice diferite (de exemplu, pe baza PLCO, a setului de date interne sau a literaturii). Figura 3.5 ilustrează fiabilitatea generatorului de date propus prin prezentarea comună a distribuției stadiilor TNM ale cohortelor sintetice și a celor din PLCO. Deși datele Synthea au o granularitate mai bună a stadializării (de exemplu, T2 este clasificat ulterior în T2a, T2b sau T2c), am agregat stadiile similare pentru a corespunde formatului PLCO, obținând astfel o analiză semnificativă. În comparație cu PLCO, datele fabricate par să reprezinte în exces stadiile înalte pentru componentele T și N: T3 și T4 sunt relativ mai frecvente în datele Synthea în comparație cu PLCO, precum și în cazul pacienților N1. În plus, evaluările componentelor non-localizate (N și M) se realizează mai frecvent în PLCO comparativ cu Synthea. Având în vedere metodologia utilizată pentru colectarea proprietăților statistice încorporate în generator, este de așteptat un anumit nivel de discrepanță între cohortele reale și cele sintetice din cauza prejudecăților inductive. Cu toate acestea, după cum se arată în figura 3.5, cohorta sintetică rezultată reprezintă în mod fiabil toate stadiile, menținând în același timp clasamentul prevalenței relative în raport cu PLCO, cu excepția stadiilor T1 și T2. În plus, este demn de remarcat faptul că studiul PLCO a fost conceput pentru a evalua eficacitatea protocoalelor de screening, prin urmare nu prezintă în mod specific o populație cu risc de cancer de prostată. Dimpotrivă, cohorta noastră sintetică a fost menită să reflecte mai bine diferitele niveluri de agresivitate ale bolii, de la risc foarte scăzut la risc foarte ridicat și de la forme localizate la cele metastatice.

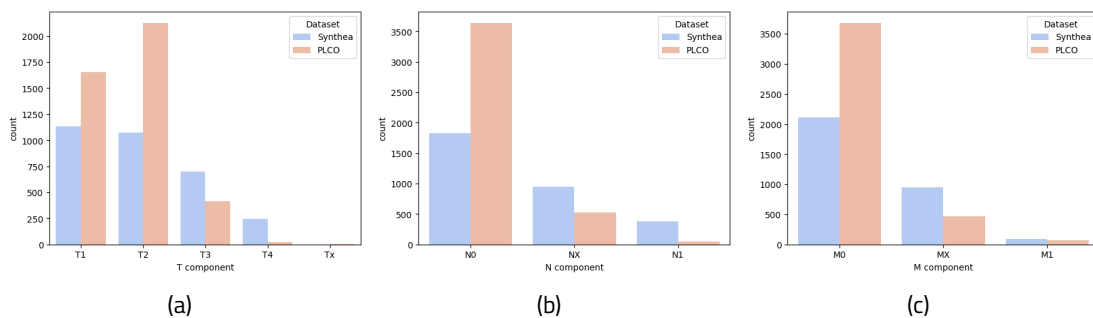


Figura 3.5: Evaluarea plauzibilității distribuției stadiilor TNM pe baza datelor PLCO.

### 3.4.2 Predicția stadiului TNM clinic

Prezentăm aici scorurile de precizie, de „recall” și F1 pentru fiecare clasă în parte, precum și mediile ponderate și neponderate, dar și precizia generală calculată pe setul de teste sintetice. Predicția componentei T a sistemului de stadializare este cea mai dificilă din cauza necesității de a interpreta și corela un număr mai mare de rapoarte longitudinale și, de asemenea, din cauza unui nivel mai ridicat de granularitate în clasificare (problema clasificării în 8 clase). Prin urmare, predictorul stadiului T a avut o acuratețe globală de numai 98,7% în comparație cu predictorii N și M (problema clasificării în 3 clase), care au atins peste 99,5%, după cum se arată în tabelul 3.2.

Tabelul 3.2: Evaluarea modelului de predicție a stadiului TNM.

Componenta stadiului	Precizia	Scorul „Recall”	Scorul F1	Numărul de apariții	
T	Tx	1	0.997	0.998	590
	T1c	0.997	0.984	0.991	1062
	T2a	0.971	0.962	0.966	208
	T2b	0.972	0.991	0.981	428
	T2c	0.978	0.994	0.986	311
	T3a	0.990	0.986	0.988	417
	T3b	0.994	0.991	0.992	333
	T4	0.952	0.983	0.967	240
	Acuratețea	0.987	0.987	0.987	3589
	Media	0.982	0.986	0.984	3589
	Media ponderată	0.987	0.987	0.987	3589
N	Nx	0.992	0.996	0.994	1643
	N0	0.996	0.992	0.994	1849
	N1	1	1	1	399
	Acuratețea	0.995	0.995	0.995	3891
	Media	0.996	0.996	0.996	3891
	Media ponderată	0.995	0.995	0.995	3891
M	Mx	0.993	0.998	0.995	1643
	M0	0.998	0.995	0.997	2176
	M1	1	1	1	72
	Acuratețea	0.996	0.996	0.996	3891
	Media	0.997	0.998	0.997	3891
	Media ponderată	0.996	0.996	0.996	3891

### 3.5 Concluzii

Acest capitol prezintă o metodă de estimare a stadiului clinic al cancerului din rapoartele de diagnostic stocate în dosarele medicale electronice, utilizând date sintetice realiste generate de Synthea. Din cunoștințele noastre, precizarea stadiului clinic al cancerului direct din EHR-uri nu a fost încă întreprinsă de alții, probabil din cauza provocărilor majore din procesele de adnotare a datelor și a constrângerilor privind confidențialitatea.

Prin urmare, am dezvoltat module Synthea pentru a genera fișe de sănătate realiste pentru pacienții cu cancer de prostată, acoperind o gamă largă de variații și agresivitate a bolii, variind de la risc scăzut la risc foarte ridicat, și, de la cancer localizat la cancer regional sau chiar metastatic. Având în vedere că deficitul de date reprezintă un obstacol foarte frecvent în implementarea unor soluții robuste bazate pe inteligență artificială care vizează diferite domenii ale asistenței medicale, un generator fiabil de date sintetice are potențialul de a permite dezvoltarea timpurie a prototipurilor și demonstrarea conceptelor care pot duce la îmbunătățirea experienței și a rezultatelor pacienților, reducând în același timp volumul de muncă al personalului clinic. Pe lângă capacitatea de a genera un număr nelimitat de eșantioane, datele noastre fabricate dețin un set de proprietăți extrem de importante:

- **Protejarea confidențialității.** Datorită naturii sale pur sintetice, setul de date propus de noi este lipsit de orice preocupări legate de păstrarea confidențialității sau de riscuri de re-identificare, fiind astfel potrivit pentru o gamă largă de aplicații, cum ar fi formularea și testarea ipotezelor,

dezvoltarea și testarea de software, scopuri educaționale și dezvoltări timpurii de aplicații inteligente pentru a sprijini medicii în tratarea eficientă a cancerului de prostată.

- **Completitudinea datelor.** Prin definirea atentă a hărților de îngrijire pentru diagnostic, tratament și monitorizare, datele noastre sintetice conțin toate testele și procedurile utilizate în mod curent în identificarea și gestionarea pacienților care suferă de cancer de prostată, la momentul redactării acestei teze.
- **Variabilitatea datelor.** Cohortele noastre sintetice prezintă o mare variabilitate în ceea ce privește agresivitatea bolii, precum și parcursul pacienților. Am utilizat diverse proprietăți statistice fie derivate din seturi de date reale disponibile, fie publicate în literatură, pentru a asigura un peisaj distribuțional realist al coortei noastre, inclusiv incidența stadiilor, distribuția opțiunilor de tratament, comportamentele de monitorizare și riscurile de recidivă.
- **Coerența datelor.** În ciuda coerenței distribuționale, natura longitudinală a datelor EHR în-deamnă la adoptarea de tehnici de menținere a coerenței între toate observațiile randomizate. Prin urmare, corelațiile dintre diferite caracteristici au fost extrase din publicații științifice și utilizate în toate etapele stohastice ale generatorului nostru de pacienți.

O evaluare calitativă a coortei sintetice generate de modulele Synthea propuse a fost efectuată în două moduri diferite. În primul rând, am inspectat proprietățile distribuționale ale diferitelor caracteristici și le-am comparat cu incidențele și statisticile, colectate din literatura de specialitate, sau derivate din seturi de date reale. După cum s-a demonstrat în acest capitol, datele noastre sintetice prezintă o variabilitate realistă atât în ceea ce privește specificul bolii, cât și parcursul pacienților. În al doilea rând, am îmbunătățit iterativ generatorul de date sintetice pe baza observațiilor furnizate de Dr. J.R., un oncolog cu experiență. Acest lucru a jucat un rol crucial în creșterea fiabilității și plauzibilității fișelor medicale electronice generate, dezvăluind inconsecvențe în observații, în distribuția în timp, în opțiunile de tratament, în dinamica de monitorizare a PSA, a efectelor secundare, etc.

În cele din urmă, am demonstrat că metodologia propusă poate fi utilizată cu succes în dezvoltarea de predictorii de înaltă performanță prin furnizarea de evaluări pe seturi de date sintetice: Stadializarea TNM a obținut o precizie de 98,7% pentru stadiul clinic T și o precizie de peste 99,5% pentru componentele nelocalizate pe un set de testare sintetic.

## 3.6 Discuții

### Generarea de date sintetice

Am propus aici o metodă sistematică de a genera seturi de date la scară largă utilizând Synthea. În special, am dezvoltat module pentru a crea înregistrări electronice de sănătate ale pacienților cu cancer de prostată care pot fi apoi utilizate pentru a dezvolta, de exemplu, funcții inteligente pentru a sprijini medicii în fluxurile lor de lucru. Datele noastre generate nu ridică probleme de confidențialitate, deoarece sunt complet sintetice prin natura lor, oferind în același timp (1) completitudine în ceea ce privește investigațiile și procedurile, (2) variabilitate între observații, opțiuni de tratament și traseul pacientului și (3) coerență între diferitele proceduri de diagnosticare, tratamente și monitorizare.

Cu toate acestea, din punctul de vedere al completitudinii, activitatea prezentată în acest capitol are o serie de limitări. În primul rând, datele sintetice obținute sunt limitate la fazele de diagnostic, terapie și monitorizare, dar unele aplicații ar putea necesita informații ce depășesc prima linie de tratament. De exemplu, predicția celei mai bune opțiuni de tratament pe baza datelor de monitorizare în caz de recidivă ar putea avea un potențial uriaș în furnizarea celor mai bune soluții personalizate pentru pacienți. Cu toate acestea, generatorul propus de noi acoperă doar tratamentul primului cancer apărut și vizitele de monitorizare ulterioare, astfel încât extinderea modulelor reprezintă o direcție viitoare a lucrării prezentate. În al doilea rând, deși Synthea generează rapoarte pentru biopsii sau studii imagistice precum mpMRI sau PET/CT, imaginile asociate acestor rezultate nu există. Acest

lucru împiedică adoptarea imediată a setului de date sintetice propus de noi în dezvoltarea de aplicații care implică fuziunea diferitelor modalități și tipuri de date, cum ar fi scanările mpMRI și PET, imaginile patologice și informațiile clinice. Totuși, o opțiune simplă de depășire a acestei limitări este potrivirea cu rapoartele sintetice pe baza constatărilor (de exemplu, cuplarea imaginilor mpMRI la raportul sintetic prin potrivirea numărului de leziuni, a scorurilor PI-RADS ale acestora, a localizării acestora etc.). Devenind un set de date sintetice hibride, această soluție ar spori și mai mult utilitatea cohortelor noastre fabricate, dar cu prețul slăbirii proprietăților de păstrare a confidențialității deținute de setarea actuală

În ceea ce privește variabilitatea și coerența, modulele Synthea propuse de noi ar fi putut încorpora prejudecăți inductive din sursele pe care le-am utilizat pentru a colecta proprietățile statistice. De exemplu, am utilizat datele PLCO, care au fost colectate în cadrul unui studiu de screening, pentru a obține distribuții comune ale măsurătorilor PSA și ale grupurilor Gleason. Prin urmare, am moștenit distribuția PSA a unei cohorte care nu este suspectă în mod natural la momentul screeningului, care ar putea fi diferită de datele disponibile într-un centru clinic specializat în tratarea cancerului de prostată.

În plus, am conceput modulele Synthea pentru a genera cohorte în toate scopurile, imitând astfel cât mai mult posibil peisajele distribuționale ale datelor clinice reale. Prin urmare, implementarea actuală nu permite o generare selectivă a pacienților care corespund anumitor filtre (de exemplu, generarea exclusivă de pacienți cu cancer agresiv tratați cu radioterapie și terapie hormonală).

În cele din urmă, deși îndeplinesc așteptările noastre privind plauzibilitatea din perspectiva unei cohorte, la nivel de pacient, datele Synthea sunt nerealist de bine structurate, stocând întotdeauna informațiile într-un anumit format și folosind un sistem de codificare specific. În schimb, datele clinice reale ar putea fi stocate diferit de la un spital la altul, folosind diverse sisteme de codificare pentru a stoca informații în diverse formate, sau chiar descriind constatările sub formă de text liber. Cu toate acestea, având în vedere progresele recente ale modelelor generative de limbaj (cum ar fi GPT), acestea ar putea fi utilizate pentru a spori plauzibilitatea structurală a datelor sintetice prin traducerea constatărilor din dosarele electronice în text liber și, astfel, pentru a spori utilitatea datelor noastre sintetice.

## **Stratificarea pacienților cu cancer de prostată**

În ciuda rezultatelor foarte promițătoare raportate aici, am oferit doar o dovadă a conceptului, nu un flux de lucru dezvoltat și stabilit. Deși testul de generalizare a avut succes pe un set de testare sintetic, evaluarea pe date clinice reale este obligatorie pentru a face o declarație definitivă cu privire la fezabilitatea adoptării modelului în aplicațiile reale. Pe lângă eventualele părtiniri ale modelului, tehnica de preprocesare care traduce pachetele FHIR în text clar reprezintă o posibilă sursă de erori, unde se utilizează euristici suplimentare pentru a filtra rapoartele care nu sunt legate de cancerul de prostată, asigurând o lungime maximă a secvenței de text de 512 cuvinte. Prin urmare, toate aceste elemente ar trebui să fie evaluate în mod extensiv pe o cohortă reală colectată din mai multe instituții clinice pentru a determina robustețea modelului.

În ciuda faptului că oferă doar o analiză preliminară prin predicția stadiului TNM, modulele noastre Synthea ar putea furniza datele de bază pentru explorarea diverselor cazuri de utilizare în care modelele de predicție ar putea acoperi lacunele din rutinele clinice și ar putea reduce semnificativ volumul de muncă al clinicienilor. De exemplu, s-ar putea utiliza toate informațiile disponibile în dosarele electronice pentru a prezice cea mai bună opțiune de tratament personalizată unui anumit pacient.

În concluzie, am arătat că datele sintetice generate cu Synthea pot fi utilizate pentru a dezvolta o varietate de sisteme și aplicații inteligente utilizabile în practica clinică, evitând în același timp constrângerile legate de confidențialitate.

## 4. Inteligența Artificială sigură pentru aplicațiile clinice în diagnosticarea și tratamentul cancerului.

---

Introducere

Rețea de asociere clinică pentru predicția cancerului de prostată clinic semnificativ

Clasificarea subtipurii de cancer pulmonar fără celule mici.

Concluzii

Discuții

---

### 4.1 Introducere

Integrarea tehnologiilor AI în industria medicală este în prezent unul dintre cele mai active domenii pentru cercetătorii din întreaga lume, în principal datorită potențialului său remarcabil de a crește acuratețea diagnosticului, de a oferi planuri de îngrijire personalizate și, în cele din urmă, de a îmbunătăți rezultatele pentru pacienți.

Sistemele de diagnostic asistat de calculator (DL-CAD) bazate pe învățarea profundă sunt concepute pentru a automatiza procedurile care sunt consumatoare de timp și susceptibile la erori, având astfel ca rezultat o reducere semnificativă a sarcinilor clinicienilor și o îmbunătățire a nivelului general de acuratețe a diagnosticului [110]. Cu toate acestea, integrarea imediată a acestor sisteme DL-CAD în rutina clinică este împiedicată de o serie de preocupări legate de fiabilitate și siguranță în funcționare. Datorită nivelului lor de complexitate, modelele de inteligență artificială sunt adesea percepute ca fiind cutii negre, fapt ce contravine principiilor etice care stau la baza prestației serviciilor medicale. Deciziile adoptate în domeniile asistenței medicale reprezintă riscuri iminente pentru starea de sănătate a pacienților, din cauza consecințelor inadvertente. Lipsa inerentă de transparență specifică modelelor de învățare profundă determină o creștere a scepticismului la nivelul cadrelor medicale și al pacienților acestora în ceea ce privește urmarea sugestiilor venite din partea AI, fără o înțelegere solidă a raționamentului din spatele procesului decizional. Complexitatea sporită a algoritmilor DL, comparativ cu abordările tradiționale de învățare automată, îi face mai opaci, împiedicându-le adoptarea pe scară largă, în ciuda potențialului lor. Prin urmare, dilema actuală derivă dintr-un compromis între: (1) complexitatea arhitecturală și capacitatea de a soluționa probleme complexe și (2) complexitatea algoritmului și transparența procesului decizional (Winkel, 2020).

În plus, algoritmi AI se bazează exclusiv pe învățarea automată a datelor de formare [4], ceea ce îi face foarte susceptibili la prejudecăți inductive. Prin urmare, ar trebui efectuată o evaluare și o caracterizare riguroasă a modelului pentru a determina robustețea acestuia, incluzând, de exemplu, testarea pe date colectate din mai multe instituții și, de asemenea, considerarea tuturor scenariilor posibile care ar putea apărea în practica clinică. Cu toate acestea, modelele de predicție DL, prin natura lor, tind să fie prea încrezătoare în raționamentul lor atunci când li se furnizează eşantioane de date sub-reprezentate în baza lor de date de formare, ceea ce ar putea duce la decizii incorecte, care ar putea dăuna bunăstării pacienților.

Înțelegerea raționamentului unui model de învățare automată pentru a ajunge la o anumită predicție este extrem de importantă pentru sporirea fiabilității acestuia și, astfel, pentru a ne apropia de inte-

grarea unor astfel de soluții în diferite etape ale rutinelor clinice. În ultimul deceniu, mai multe studii au încercat să atribuie scoruri de importanță caracteristicilor de intrare în ceea ce privește predicțiile modelului, ca indicator al impactului relativ pe care îl au asupra procesului de raționament [24, 25, 26, 27, 28]. Lundberg et. al. [27] au propus o metodă stocastică pentru a atribui valori Shapely fiecărei intrări a unui model de învățare automată prin eșantionarea aleatorie a coalițiilor dintre caracteristici și evaluarea efectului acestora asupra predicției. Provenind din teoria jocurilor colaborative, valorile Shapely împart recompensa (predicția modelului) între jucători (caracteristicile de intrare) pe baza contribuției acestora la rezultat. Prin urmare, în contextul ML, ele reprezintă măsuri cantitative ale impactului relativ pe care fiecare caracteristică luată în considerare l-a avut asupra inferenței.

Prin încorporarea unor astfel de tehnici în procedurile clinice, predicțiile făcute de algoritmi AI ar putea fi însoțite de explicații care ar permite medicilor să evalueze cazurile relativ dificile, sau neobișnuite, în care modelele ar putea fi predispuse la eșec. Prin urmare, creșterea transparenței ar putea consolida încrederea în astfel de sisteme DL-CAD, ceea ce ar putea aduce beneficii potențiale în ceea ce privește acuratețea diagnosticului și rezultatele actului medical. Cu toate acestea, pe lângă proprietățile de interpretabilitate și explicabilitate, pentru o rețetă completă a inteligenței artificiale sigure, o soluție de învățare profundă ar trebui, de asemenea, să dispună de capacități de identificare a cazurilor dificile în care predicțiile sunt relativ de incerte. Eșantioanele de date ar putea avea niveluri ridicate de incertitudine din cauza mai multor factori, inclusiv achiziții bruiate sau corupte, etichetare greșită sau efecte ale diferențelor de opinie între adnotatori asupra modelului [4], cazuri noi în afara distribuției datelor folosite la antrenare, etc.

S-a demonstrat că ansamblurile de modele [34] ating cele mai bune performanțe în ceea ce privește o varietate de probleme de segmentare biomedicală [33] prin agregarea predicțiilor mai multor instanțe de model antrenate pe diferite fracțiuni de date. Pe lângă o performanță în general îmbunătățită în comparație cu antrenarea standard, ansamblurile sunt capabile în mod avantajos să cuantifice un scor de incertitudine prin calcularea variabilității predicțiilor diferitelor instanțe de model, sau a fracțiunii de predicții care sunt în acord [31, 32]. Prin urmare, o soluție bazată pe inteligență artificială ar putea identifica cazurile dificile pe baza acestor estimări ale incertitudinii și ar putea solicita asistența radiologilor, sporind astfel robustețea generală a întregului sistem.

În general, sistemele DL-CAD care dau dovadă de transparență prin furnizarea de predicții explicabile și interpretabile, precum și de estimări ale incertitudinii, ar putea reduce lacunele legate de încredere care împiedică în prezent adoptarea generală a acestor metode în rutina clinică, deblocând potențialul promis de inteligența artificială în îmbunătățirea îngrijirii pacienților, reducând în același timp în mod semnificativ volumul de muncă pe care îl au în prezent cadrele medicale. În acest capitol explorăm modul în care modelele standard de învățare profundă ar putea beneficia de îmbogățirea lor cu capacități de explicabilitate și cuantificare a incertitudinii, ceea ce ar putea să le apropie de conceptul de inteligență artificială sigură. Concret, contribuția noastră ar putea fi prezentată după cum urmează:

- Explorăm fezabilitatea utilizării unei analize Shap [27] în explicarea raționamentului modelelor DL în rezolvarea a două sarcini relevante din punct de vedere clinic, și anume predicția cancerului de prostată clinic semnificativ (csPCa) și clasificarea subtipurii de cancer pulmonar fără celule mici (din eng. „non-small cell lung cancer” - NSCLC).
- Propunem estimări ale incertitudinii bazate pe ansambluri pentru a identifica modele mutaționale neobișnuite sau contradictorii într-un set de date genomice disponibile în mod public și, astfel, să elucidăm limitările modelelor predictive actuale în reprezentarea datelor eterogene.

Acest capitol este organizat după cum urmează: Secțiunea 4.2 descrie modul în care un algoritm, de ultimă generație, de detectare a leziunilor poate beneficia de informații clinice suplimentare în identificarea cazurilor semnificative din punct de vedere clinic de cancer de prostată. Pe lângă o performanță generală îmbunătățită, valorile Shapely oferă o perspectivă excelentă asupra modului în care intrările suplimentare au influențat procesul de raționament al modelului. Secțiunea 4.3 prezintă o soluție de învățare profundă pentru clasificarea NSCLC în subtipuri pe baza datelor genomice.



Având în vedere că o clasificare precisă este extrem de importantă pentru planificarea tratamentului, estimările incertitudinii și explicațiile bazate pe analize Shapely au fost utilizate pentru a crește robustețea modelului. În cele din urmă, concluziile generale sunt formulate în secțiunea 4.4, în timp ce toate abordările prezentate aici sunt discutate în 4.5.

## 4.2 Rețea de asociere clinică pentru predicția cancerului de prostată clinic semnificativ

### 4.2.1 Introducerea scenariului clinic

Cancerul de prostată (din eng. „Prostate CAncer” - PCa) este a doua cauză principală de deces (prin cancer) în rândul bărbaților din SUA și prima în ceea ce privește numărul estimat de cazuri noi (1 din 8 bărbați va fi diagnosticat astfel pe parcursul vieții sale) [83]. Antigenul specific prostatic (PSA) este modalitatea stabilită de a identifica inițial un pacient ca fiind suspect de cancer de prostată, recomandând o biopsie prostatică confirmatorie, o procedură invazivă și riscantă. Cu toate acestea, deoarece PSA poate fluctua și din cauza unor factori non-maligni, specificitatea sa este scăzută [111], ceea ce conduce la un număr semnificativ de supra-diagnosticări. Imagistica prin rezonanță magnetică multi-parametrică (mpMRI) a fost adoptată pe scară largă în rutina clinică pentru a tria mai bine pacienții cu niveluri anormale ale PSA, ajutând 27% dintre aceștia să evite o biopsie primară inutilă, detectând în același timp cu până la 18% mai multe cancere de prostată semnificative din punct de vedere clinic (cs-PCa) [112]. „Prostate Imaging Reporting and Data System” (PI-RADS v2) [113] este un sistem frecvent utilizat pentru localizarea și raportarea leziunilor suspecte din prostată, clasificându-le cu un scor de la 1 la 5 (orice leziune cu un scor mai mare de 2 este considerată malignă). Cu toate acestea, citirea unei serii mpMRI este o sarcină consumatoare de timp, necesită un nivel ridicat de expertiză [114], iar acordul între radiologi în atribuirea unui scor PIRADS este adesea destul de mic [115]. În plus, în conformitate cu recomandările actuale, radiologii sunt instruiți să atribuie un scor PI-RADS fără a avea acces la parametrii clinici și demografici, care, în plus față de studiul mpMRI, ar putea oferi indicii valoroase în diagnosticarea cs-PCa.

Sistemele de diagnostic asistat de calculator (CAD) ar putea fi utilizate pentru a acoperi lacunele actuale din cadrul sistemului de notare PI-RADS prin automatizarea procesului, păstrând (sau chiar depășind) performanța radiologilor. Deși multe sisteme CAD candidate ating o sensibilitate comparabilă cu cea a clinicienilor, acestea nu sunt încă suficient de specifice [116].

Mai multe studii au demonstrat eficacitatea evaluării concomitente a parametrilor clinici și a informațiilor imagistice pentru o mai bună stratificare a pacienților [117, 118, 119]. Cu toate acestea, aceste modele depind în continuare de acuratețea interpretării imaginilor. Cu toate acestea, ideea de a utiliza informații clinice pentru a ghida interpretarea imagistică ar putea fi adoptată cu succes în sistemele CAD pentru a le spori robustețea [111].

În această secțiune, ne propunem să îmbunătățim performanța sistemelor CAD prin adăugarea unei faze suplimentare de corecție clinică în fluxul de lucru. În mod specific, utilizăm o rețea neuronală profundă pentru a ajusta predicția unui sistem de ultimă generație [116], pe baza unei serii de informații clinice (de exemplu, PSA, densitatea PSA, volumul glandelor) și demografice (de exemplu, vârsta). În plus, efectuăm o analiză Shap pentru a atribui o importanță relativă fiecărei caracteristici de intrare în ceea ce privește predicția finală pentru a crește transparența modelului DL.

### 4.2.2 Metode

Propunem aici o etapă de corecție clinică pentru o linie de procesare curentă [11], care automatizează procesul de diagnosticare a cancerului de prostată printr-o serie de algoritmi de învățare automată, care efectuează secvențial segmentarea glandelor, detectarea leziunilor, reducerea de fals pozitive și calificarea leziunilor. În primul rând, o fază de preprocesare extrage seriile de imagini T2 ponderate (T2W) și dinamice ponderate (DWI) din DICOM și calculează o hartă ADC. În continuare, așa

cum este prezentat în figura 4.1, se utilizează un model de segmentare a întregii glande [120] pentru a detecta prostata în seriile T2W și DWI, și pentru a crea o mască 3D care va servi ca intrare pentru modelul ulterior de detectare a leziunilor. În urma metodelor prezentate în [116, 110], se utilizează o rețea neuronală convoluțională pentru a detecta leziunile candidate în cadrul prostatei, iar apoi se utilizează o rețea de reducere de fals pozitive (FPR) pentru a ajusta predicția, păstrând în același timp sensibilitatea generală.

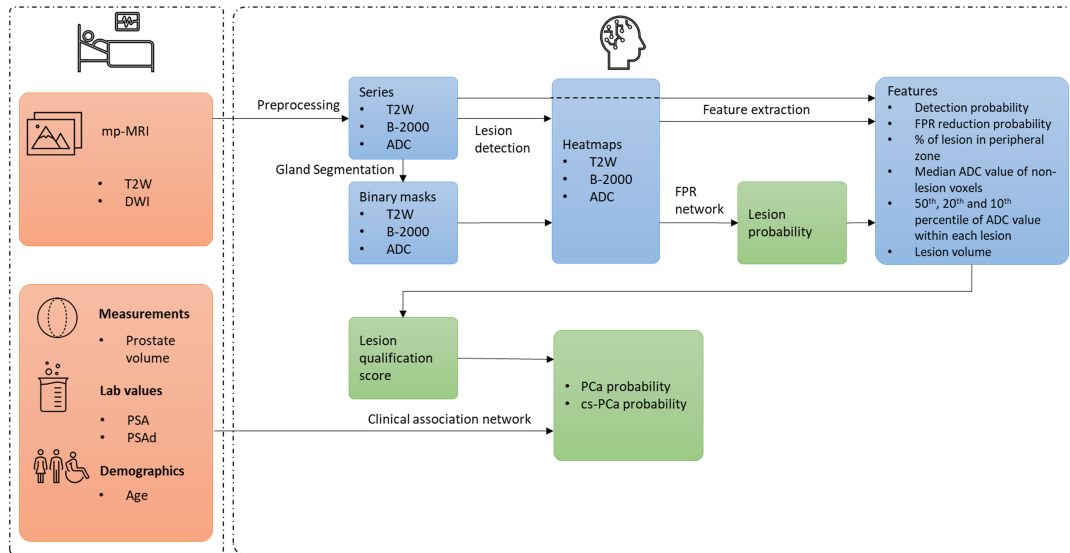


Figura 4.1: Fluxul de lucru propus.

Procesul complet automatizat descris mai sus generează un scor de calificare (QS) care descrie malignitatea unei leziuni exclusiv pe baza imaginilor RMN bi-parametrice. Având în vedere că informațiile clinice ar putea deține o capacitate predictivă suplimentară, propunem o fază de corecție în care scorul QS este combinat cu caracteristici adiționale pentru a ajusta predicția globală a malignității. În această măsură, combinăm rezultatele procesului inițial cu parametri clinici suplimentari și utilizăm o rețea neuronală, denumită ulterior rețea de asociere clinică, pentru a prezice dacă o anumită leziune este semnificativă din punct de vedere clinic.

#### 4.2.2.1 Setul de date

Am colectat date anonime de la 11 clinici din diferite țări, asigurând eterogenitatea, ceea ce duce la un potențial crescut de generalizabilitate a modelului final. Setul de date constă în 2261 de pacienți cu date clinice (de exemplu, PSA, PSAD), demografice (de exemplu, vârsta) și imagistice (serii RMN multi-parametrice împreună cu un scor PI-RADS atribuit de radiologi experimentați). 48,7% dintre acești pacienți au fost diagnosticați cu cancer de prostată printr-o biopsie țintită, ca urmare a unui mpMRI anormal (PIRADS  $\geq 3$ ), dintre care 59% au fost găsiți cu boală semnificativă din punct de vedere clinic (GG  $> 2$ ). Deoarece majoritatea centrelor nu efectuează biopsii sistematice ca parte a rutinei lor clinice, am considerat cazurile cu un scor PIRADS mai mic de 2 ca fiind negative. Au fost excluse toate cazurile în care PIRADS a fost mai mare de 2, dar nu a fost disponibil scorul Gleason. Tabelul 4.1 prezintă statisticile generale ale setului de date.

#### 4.2.2.2 Rețeaua de calificare a leziunilor

După cum se propune în [116], toate leziunile identificate și confirmate de modelele de detecție și FPR sunt apoi inspectate de un model de calificare, care atribuie un scor de calificare (QS) pe baza unei serii de 8 caracteristici de intrare: probabilitatea de detecție, probabilitatea de reducere a FPR,

Tabelul 4.1: Descrierea setului de date.

	Grup Gleason 0 N = 1160	Grup Gleason 1 N=451	Grup Gleason >= 2 N = 650
PSA [ng/dL]	9.08 ( $\pm$ 9.25)	8.56 ( $\pm$ 7.48)	16.55 ( $\pm$ 30.72)
PSAD [ng/dL/cc]	0.16 ( $\pm$ 0.21)	0.22 ( $\pm$ 0.28)	0.45 ( $\pm$ 0.74)
Volumul [cc]	66.14 ( $\pm$ 39.3)	46.88 ( $\pm$ 24.99)	41.65 ( $\pm$ 23.03)
Vârsta [ani]	63.83 ( $\pm$ 7.72)	64.65 ( $\pm$ 7.23)	67.06 ( $\pm$ 7.52)
PI-RADS	2.12 ( $\pm$ 0.83)	3.54 ( $\pm$ 1.15)	4.31 ( $\pm$ 0.82)
Grup Gleason	0	1	3.03( $\pm$ 1.13)

proporția de extindere a leziunii în zona periferică, valoarea mediană ADC a tuturor voxelurilor care nu sunt leziuni, percentilele 50, 20 și 10 ale valorilor ADC în cadrul fiecărei leziuni și volumul leziunii calculat din harta termică 3D produsă de rețeaua de detecție. Modelul de calificare are o arhitectură densă cu 2 straturi ascunse, a câte 32 de neuroni fiecare. Între al doilea strat ascuns și ieșire a fost utilizat un strat de „dropout” pentru regularizare. Scorul de calificare produs de acest model poate fi folosit în sine pentru a prezice PCa și/sau cs-PCa doar pe baza seriei mpMRI. Prin urmare, sistemul propus în această secțiune poate să funcționeze și atunci când nu sunt disponibile date clinice suplimentare, dar cu performanțe limitate, după cum se arată în secțiunea 4.2.3.

#### 4.2.2.3 Rețeaua de asociere clinică

Atât probabilitatea modelului FPR, cât și scorul de calificare pot fi folosite în mod direct pentru a prezice PCa sau cs-PCa. Cu toate acestea, propunem aici o fază suplimentară, în care adăugăm informațiile clinice disponibile pentru a îmbunătăți în continuare performanța sistemului. Ne referim la această etapă drept o fază de corecție clinică, în care folosim o altă rețea neurală pentru a ajusta predicția de malignitate pe baza datelor suplimentare furnizate rețelei. Aceasta este, de asemenea, o rețea densă cu 3 straturi ascunse activate de funcții tangente hiperbolice, în timp ce stratul de ieșire constă dintr-un neuron activat de o funcție sigmoidă. La fel ca în cazul rețelei de calificare, pentru regularizarea modelului a fost utilizat un strat de „dropout”, care întrerupe stohastic 20% din conexiunile dintre ultimul strat ascuns și ieșirea. Pentru a evalua impactul fiecărei caracteristici suplimentare în mod independent, efectuăm o analiză multivariată constând în 10 experimente diferite: QS + vârsta, QS + PSA, QS + densitatea PSA (PSAD), QS + volumul, QS + PSAD + volumul, QS + vârsta + PSA, QS + vârsta + PSAD, QS + vârsta + volumul, QS + vârsta + PSAD + volumul, QS + vârsta + PSA + PSAD + volumul.

Rețeaua de asociere clinică a fost antrenată pentru a minimiza o funcție de cost de tip entropie binară încrucișată (BCE) pentru 300 de epoci. Cu toate acestea, selecția modelului s-a realizat în raport cu eroarea pe setul de validare. Optimizarea a fost realizată utilizând optimizatorul Adam cu o rată inițială de învățare de 0,0001 și o dimensiune a loturilor de antrenare egală cu 64. Pentru lua în considerare dezechilibrul claselor (în special în cazul de utilizare a predicției cs-PCa), am definit ponderi de clasă pentru a pune accentul pe exemplele din clasa minoritară.

A fost utilizată o schemă stratificată de validare încrucișată, prin care loturi distincte de 20% din date sunt utilizate iterativ ca set de validare. Pentru a crește robustețea modelului, scorul final este obținut prin calcularea mediei celor 5 scoruri generate de fiecare model individual. Pentru a evita orice prejudecăți care decurg din diferențele instituționale, capacitatea finală de generalizare a ansamblului a fost realizată pe un set de testare separat, provenit de la un spital diferit.

### 4.2.3 Rezultate

#### Analiza multivariată

După cum este descris în tabelul 4.2, scorurile PI-RADS atribuite de radiologi pot prezice PCa cu un AUC de 0,904, în timp ce QS atinge doar 0,852. Cu toate acestea, rețeaua de corecție clinică îmbunătățește semnificativ QS în majoritatea cazurilor, în special atunci când se ia în considerare volumul total al glandei: atunci când se adaugă doar volumul, rețeaua de corecție atinge un AUC de 0,902, atingând astfel performanța radiologică (valoarea  $p = 7,87e-01$ , indicând faptul că nu există nicio diferență semnificativă din punct de vedere statistic între rezultatul modelului nostru și scorurile PI-RADS).

Tabelul 4.2: Rezultatele analizei multivariate.

	PCa			cs-PCa		
	PI-RADS AUC	QS AUC	DL AUC	PI-RADS AUC	QS AUC	DL AUC
QS + Vârsta	0.904	0.852	0.854	0.899	0.841	0.865
QS + Volum	0.904	0.852	0.902	0.899	0.841	0.899
QS + PSA	0.904	0.852	0.866	0.899	0.841	0.857
QS + PSAD	0.904	0.852	0.869	0.899	0.841	0.867
QS + PSAD + Volum	0.904	0.852	0.897	0.899	0.841	0.890
QS + Vârsta + Volum	0.904	0.852	0.895	0.899	0.841	0.904
QS + Vârsta + PSA	0.904	0.852	0.855	0.899	0.841	0.867
QS + Vârsta + PSAD	0.904	0.852	0.852	0.899	0.841	0.873
QS + Vârsta + PSAD + Volum	0.904	0.852	0.894	0.899	0.841	0.908
QS + Vârsta + PSA + PSAD + Volum	0.904	0.852	0.897	0.899	0.841	0.906

În ceea ce privește cs-PCa, PI-RADS clinic are un AUC de 0,899, în timp ce QS are un AUC de 0,841. În concordanță cu rezultatele privind predicția PCa, informațiile privind volumul par să ofere cea mai mare creștere în precizia predicției cs-PCa, îmbunătățind QS cu 5,8 puncte în AUC, comparabil cu performanța radiologilor. Mai mult, PSAD utilizat împreună cu QS, vârsta și volumul glandei produc o creștere a AUC la 0,908, depășind PIRADS clinic cu 0,9 puncte.

#### 4.2.3.1 Explicabilitate și interpretabilitate

Deși beneficiile unei astfel de soluții autonome sunt incontestabile, există un nivel ridicat de scepticism în ceea ce privește modul în care un model DL ajunge la o anumită concluzie. Deși eficacitatea lor a fost puternic demonstrată de comunitatea științifică în ultimul deceniu, algoritmi de învățare profundă sunt asimilați "cutiilor negre", împiedicând adesea adoptarea lor în rutina clinică [4]. Prin urmare, această secțiune prezintă rezultatele unei analize Shap [27] utilizate pentru a oferi explicații cu privire la modul în care rețeaua de asociere clinică propusă utilizează parametri suplimentari în procesul său de raționament.

Figura 4.2 prezintă o analiză Shap globală care descrie impactul relativ pe care fiecare parametru de intrare considerat îl are asupra rezultatelor modelului. Valorile diferiților parametri sunt reflectate de harta de culori, valorile scăzute fiind reprezentate în albastru, în timp ce punctele roșii reprezintă valori ridicate. Fiecare dintre parametrii utilizați aici poate influența predicțiile modelului într-o anumită măsură, care este indicată pe axa x. Prin urmare, în ceea ce privește impactul relativ pe care îl au asupra predicției cs-PCa, se obține următorul clasament al parametrilor propuși de noi: QS, volum, vârstă, PSAD și PSA.

În afară de scorul de calificare, cea mai predictivă caracteristică este volumul total al glandei, care poate modifica rezultatul modelului cu aproximativ  $\pm 20\%$ , după cum urmează: glandele mari par să reducă riscul de csPCa, în timp ce cele mici par să facă contrariul. În schimb, vârsta pacientului prezintă o corelație pozitivă cu QS, deplasând pozitiv predicția în funcție de vârstă. Același lucru este valabil și pentru PSA și PSAD, care sunt corelate pozitiv cu rezultatul, dar cu un impact relativ mai mic. În general, figura 4.2 demonstrează că fiecare parametru suplimentar luat în considerare are un anumit impact asupra predicțiilor și, astfel, câștigul de performanță produs de faza de corecție clinică are o bază stabilă.

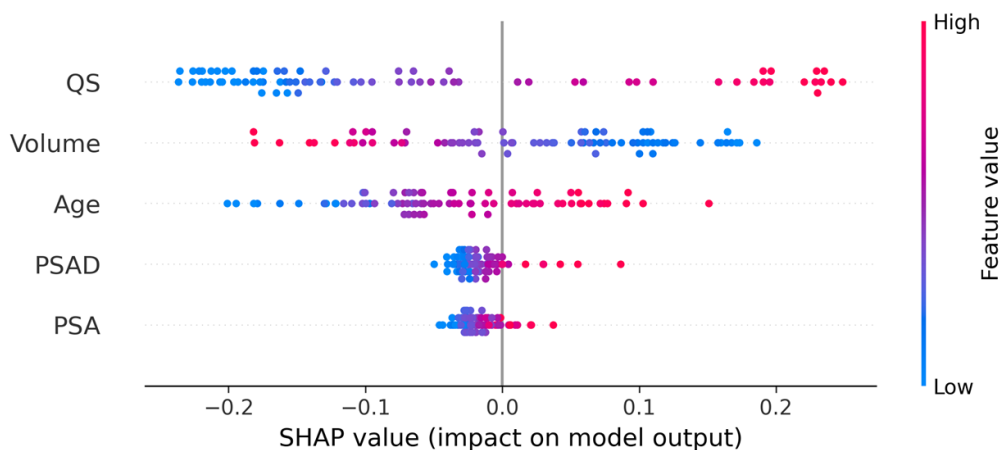


Figura 4.2: Analiza Shap la nivel global.

O astfel de analiză la nivel global sporește transparența prin furnizarea de explicații cu privire la modul în care fiecare intrare joacă rolul său în inferență. Prin evaluarea acestor rezultate, clinicienii ar putea înțelege mai bine și ar putea corela tiparele învățate de rețelele neuronale în timpul antrenării cu situațiile clinice pe care le-au experimentat și, prin urmare, ar putea crește încrederea lor în astfel de soluții. Cu toate acestea, analiza Shap poate fi, de asemenea, efectuată la nivel de eșantion, oferind astfel un instrument pentru a face interpretabile predicțiile modelului în timp real.

#### 4.2.4 Concluzii

După cum se menționează în secțiunea 4.2.3, rețeaua neuronală de corecție clinică îmbunătățește semnificativ scorul de calificare, făcând ca întregul rezultat al sistemului CAD să fie comparabil cu scorurile PI-RADS atribuite prin consens de către radiologi experimentați. În plus, atunci când volumul glandei este luat în considerare, sistemul CAD propus depășește PI-RADS clinic cu o marjă mică, fiind complet autonom.

Urmând o abordare similară, am demonstrat că nu numai QS poate beneficia de informații clinice suplimentare, ci și de scorul PI-RADS: de la un AUC median de 0,904, analiza noastră multivariată a arătat o îmbunătățire de 0,012 puncte a AUC atunci când vârsta și PSA au fost luate în considerare, în timp ce adăugarea volumului glandei a îmbunătățit și mai mult AUC median la peste 0,93.

Mai mult, am sporit transparența modelului prin efectuarea unei analize Shap care atribuie un scor de importanță relativă fiecărei caracteristici de intrare în raport cu riscul prezis de cancer de prostată clinic semnificativ.

#### 4.2.5 Discuții

Această secțiune a oferit un exemplu al modului în care informațiile suplimentare pot fi încorporate, atunci când sunt disponibile, într-o soluție complet autonomă pentru a-i crește performanța și stabilitatea generală. În mod specific, am demonstrat modul în care caracteristicile clinice și de-

mografice ar putea fi utilizate pentru a îmbunătăți precizia unui sistem DL-CAD care analizează imaginile RMN bi-parametrice și atribuie un scor de malignitate și, de asemenea, puterea de discriminare a scorului PI-RADS atribuit de radiologi. În plus, un algoritm de deducere a importanței caracteristicilor a fost utilizat în elucidarea raționamentului rețelei de asociere clinică propusă, sporind transparența și, astfel, fiabilitatea acestei soluții.

De remarcat, această lucrare reprezintă mai degrabă o confirmare a conceptului decât o soluție stabilită și, prin urmare, trebuie să fie șlefuită cu atenție pentru a-și atinge potențialul maxim. În primul rând, în lucrarea actuală am investigat modul în care analiza Shap poate oferi predicții explicabile ale rețelelor neuronale profunde luând în considerare doar modulul final al fluxului de lucru, și anume rețeaua de asociere clinică propusă în această secțiune. Prin urmare, modul exact în care sistemul DL-CAD inițial a produs scorul de calificare rămâne încă neclar, împiedicând interpretabilitatea și explicabilitatea generală a acestuia. Extinderea acestei analize la întreaga soluție reprezintă o direcție viitoare a acestei lucrări.

În al doilea rând, întregul flux de lucru este acoperit în prezent de mai multe module independente. Avantajul acestei abordări este că rezultatele intermediare pot fi utilizate pentru a face o predicție în cazul în care datele necesare modelelor ulterioare nu sunt disponibile pentru un anumit pacient. Pe de altă parte, această abordare poate facilita propagarea erorilor în cazul în care unul dintre modulele inițiale eșuează. O soluție alternativă ar fi un cadru complet care să acopere toate sarcinile descrise mai sus, profitând în același timp de un semnal de eroare mai bogat în timpul antrenării. Cu toate acestea, o astfel de abordare ar putea împiedica transparența generală a sistemului din cauza unei complexități sporite.

Dincolo de sistemul DL-CAD utilizat aici, constatările noastre sugerează că și sistemul de punctare PI-RADS poate beneficia de informații suplimentare disponibile clinic. În timp ce ghidurile curente sugerează că cititorii ar trebui să atribuie un scor PI-RADS bazat exclusiv pe imagini de rezonanță magnetică multi-parametrică, rezultatele noastre demonstrează că tipare semnificative pot fi învățate pentru a îmbunătăți performanța PI-RADS în discriminarea leziunilor semnificative clinic de celelalte.

În general, am demonstrat că se poate obține o performanță sporită atunci când imagistica este combinată cu alte tipuri de date și modelată la comun de algoritmi de învățare automată care pot identifica tipare semnificative și le pot utiliza în predicția anumitor rezultate. Pe lângă o evaluare riguroasă, sistemele DL-CAD ar trebui să posede proprietăți de transparență în procesul decizional pentru a deveni sigure, ceea ce ar putea debloca beneficii enorme pentru clinicieni și pacienții acestora.

## 4.3 Clasificarea subtipurilor de cancer pulmonar fără celule mici <sup>1</sup>

### 4.3.1 Introducerea scenariului clinic

Tehnologiile de genotipare bazate pe secvențierea de ultimă generație (NGS) sunt din ce în ce mai utilizate pentru a sprijini procesul decizional clinic. Abordarea standard pentru dezvoltarea modelelor de diagnostic, prognostic sau predicție bazate pe date înalt-dimensionale generate de NGS constă în preselecția unui număr mic de biomarkeri, de exemplu, mutații genetice care sunt identificate ca fiind asociate independent cu fenotipul de interes, și limitarea antrenării unui model statistic de bază acestui set restrâns de biomarkeri, de exemplu, pentru a prezice beneficiile imunoterapiilor [122, 123].

De fapt, etiologia bolii și răspunsul la tratament sunt adesea complexe, afectate de o multitudine de modificări genomice, fiecare exercitând un efect mic asupra fenotipului. Prin urmare, este de așteptat ca luarea în considerare a tuturor alterărilor identificate să extindă gama de riscuri fenotipice. Cu toate acestea, din cauza dimensionalității ridicate a datelor genomice, a eterogenității genetice

---

<sup>1</sup> Această secțiune descrie experimentele efectuate în [121], care reprezintă lucrări publicate anterior de autor, în cadrul programului de cercetare doctorală.

inerente, precum și a cohortelor mici de pacienți, formarea robustă a modelelor predictive reprezintă o provocare [124].

Se preconizează că citirile funcționale ale activității celulare și ale stării fiziologice furnizate de tehnologiile "omice", cum ar fi secvențierea ARN, care măsoară modificările expresiei ARNm la nivelul întregului genom, vor avea o capacitate mai mare de a fundamenta deciziile de gestionare clinică, inclusiv diagnosticul, prognosticul, selectarea tratamentului și monitorizarea. În mod ideal, ar trebui combinate tehnologiile "omice" care surprind interacțiunea moleculară complexă în cadrul diferitelor niveluri biologice. Cu toate acestea, din motive practice și financiare, genomica este modalitatea cea mai adoptată clinic până în prezent [126, 127, 128, 125].

În cazul cancerului, clasificarea tumorilor bazată pe histologie reflectă diferite prezentări clinice și evoluții ale bolii. De exemplu, cancerul pulmonar este clasificat ca fiind cu celule mici (SCLC) sau fără celule mici (NSCLC).

În mod important, au fost identificate determinante prognostice diferite în cazul adenocarcinomului pulmonar (LUAD) comparativ cu carcinomul cu celule scuamoase (LUSC), care în parte au impact opus asupra rezultatului clinic [131]. În plus, s-a constatat că histologia NSCLC poate prezice răspunsul la tratament. De exemplu, chimioterapia adjuvantă pe bază de platină pentru pacienții cu NSCLC în stadiu incipient a conferit un beneficiu de supraviețuire în LUSC-uri, dar nu și în LUAD [132]. În schimb, chimioterapia cu Pemetrexed a arătat doar o eficacitate îmbunătățită în comparație cu alte opțiuni de tratament standard la pacienții cu NSCLC non-scuamos avansat [133]. Histologia NSCLC este, de asemenea, considerată predictivă a răspunsului la terapiile țintite și la imunoterapii în cele mai recente orientări ASCO pentru pacienții cu NSCLC în stadiu avansat cu [134] și fără [135] alterări acționabile ale determinațiilor

Analog răspunsului diferențiat la medicamentele anticanceroase, efectul dozei de radioterapie corporală stereotactică (SBRT) a fost, de asemenea, descris ca fiind diferit între subtipurile de NSCLC [137, 138], cu o rată crescută de eșec local la pacienții cu LUSC, indicând necesitatea ajustării tratamentului în funcție de histologie. Prin urmare, cunoașterea histologiei NSCLC este esențială în selectarea optimizată din opțiunile terapeutice disponibile.

În această lucrare, am urmărit îmbunătățirea clasificării subtipului NSCLC din date mutaționale și am dezvoltat un model profund de profilare genomică care, pe lângă probele LUAD și LUSC, învață simultan din probele de adenocarcinom (AD) și carcinom cu celule scuamoase (SCC) ale altor tipuri de țesuturi. În special, performanța de clasificare poate fi îmbunătățită pe eșantioane cu predicții încrezătoare, identificate cu o abordare de modele ansamblu care captează incertitudinile de predicție. În plus, estimările incertitudinii probelor clasificate greșit indică limitările sistemului actual de clasificare NSCLC în ceea ce privește reprezentarea eterogenității mutaționale în cadrul subtipurilor, ceea ce ar putea împiedica predicția rezultatelor tratamentului.

### 4.3.2 Rezultate

Pentru a stabili o linie de bază, am instruit un model de profilare genomică constând într-o rețea neuronală densă (MLP) pentru a clasifica probele NSCLC în subtipurile LUAD și LUSC utilizând date mutaționale. Acest model de bază a obținut o suprafață sub curba caracteristică de funcționare a receptorului (ROC) (AUC) de 0,82 ca măsură a performanței de clasificare pe eșantioane de testare utilizând o schemă de validare încrucișată de 10 ori.

#### **Regularizarea bazată pe expresii îmbunătățește spațiul latent, fără a îmbunătăți performanța de clasificare**

Performanța de clasificare a modelului de profilare genomică de bază nu a putut fi depășită, indiferent de diferitele configurații de antrenament și parametrii ai modelului, ceea ce demonstrează provocarea prezentată de acest set de date genomic de înaltă dimensionalitate. Pentru a depăși această provocare, am creat un set de date extins prin creșterea setului de date NSCLC cu probe AD și SCC suplimentare de histologie non-NSCLC. Acest set de date extins a fost utilizat pentru a regu-

lariza antrenarea unui model de profilare genomică, prin clasificarea simultană a probelor AD și SCC din plămâni, precum și a altor tipuri de țesuturi, utilizând două capete de predicție. Cu toate acestea, performanța de clasificare a probelor NSCLC nu s-a îmbunătățit semnificativ față de cea a modelului de bază.

Am urmărit în continuare să îmbunătățim precizia clasificării NSCLC prin regularizarea antrenării modelului de profilare genomică cu reprezentarea latentă învățată de un model de profilare bazat pe expresia genelor, obținând astfel un model de profilare genomică sensibil la expresie. Ca și în rapoartele anterioare, performanța modelului de profilare bazat pe expresie în clasificarea subtipurilor NSCLC atinge un AUC de 0,98 ( $\pm 0,01$ ). Cu toate acestea, antrenarea regularizată nu a reușit să valorifice capacitatea de predicție a modelului de profilare bazat pe expresie pentru a îmbunătăți performanța de clasificare a modelului de bază.

### Estimările incertitudinii predicției permit creșterea performanței pe eșantioane sigure

Rezultatele indică o complexitate și o ambiguitate inerentă în ceea ce privește modul în care profilul genomic se traduce în activitatea celulară și starea fiziologică. Captarea acestei incertitudini în cadrul modelului ar putea permite identificarea și respingerea probelor ambigue, îmbunătățind în același timp performanța probelor rămase. În acest scop, am utilizat o abordare de agregare (numită și bagging) pentru a antrena un ansamblu de o sută de modele de profilare genomică și am calculat un scor de predicție agregat prin medierea predicțiilor tuturor modelelor care indică subtipul NSCLC prezis de majoritate.

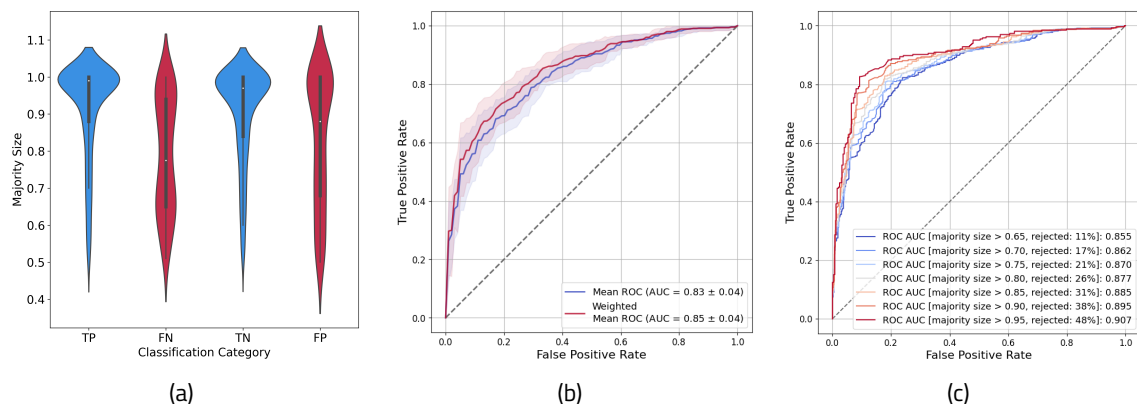


Figura 4.3: Performanța unui ansamblu de modele de profilare genomică sensibilă la expresie antrenate pentru a clasifica subtipurile NSCLC utilizând date mutaționale din setul de date extins. (a) Distribuția mărimii majorității ca estimare a incertitudinii pentru diferite categorii de clasificare, TP, FN, TN și FP corespunzând predicțiilor pozitive adevărate, negative false, negative adevărate și pozitive false. (b) Performanța modelului calculată cu (roșu) și fără (albastru) ponderarea bazată pe incertitudine a eșantioanelor. (c) Performanța modelului atunci când se aplică diferite praguri de estimare a incertitudinii și se resping eșantioanele de testare respective. În plus față de pragul aplicat, legenda enumeră fracțiunea de eșantioane respinse din toate eșantioanele.

Deși performanța de clasificare a ansamblului a rămas neschimbată (Fig. 4.3(b), curba ROC albastră), ansamblul a permis identificarea eșantioanelor neconcludente prin calcularea fracțiunii de modele care prezic subtipul NSCLC prezis majoritar (denumit dimensiunea majorității) ca măsură a incertitudinii predicției.

Figura 4.3(a) arată acordul dintre modelele ansamblului separat pentru fiecare categorie de clasificare. Deoarece acordul este mai mare pentru eșantioanele clasificate corect (TP și TN) decât pentru cele clasificate greșit (FP și FN), dimensiunea majorității s-a dovedit a fi o măsură rezonabilă a incertitudinii predicției.



Într-adevăr, AUC a crescut cu 0,02 atunci când se utilizează estimările incertitudinii ca ponderi în calcularea curbei ROC (Fig. 4.3(b), curba ROC roșie). În plus, se poate aplica o valoare prag pentru a respinge probele cu incertitudine ridicată de predicție și, ca urmare, pentru a crește performanța de clasificare a probelor rămase, după cum se demonstrează în Fig. 4.3(c). În mod esențial, AUC a crescut monoton cu cât pragul aplicat este mai restrictiv. Cu toate acestea, există un compromis între performanța de clasificare și fracțiunea de probe respinse. De exemplu, impunerea unei mărimi minime a majorității de 0,75 a crescut AUC la 0,87, dar a implicat, de asemenea, respingerea a 21% din probele de testare. De asemenea, AUC a depășit 0,9 atunci când s-a aplicat un prag mai restrictiv de 0,95, cu dezavantajul că aproape jumătate din eșantioanele de testare au fost respinse.

### Tiparele mutaționale contradictorii limitează performanța de clasificare

Pentru a estima contribuția fiecărei caracteristici, adică statutul mutațional al fiecărei gene, la predicțiile modelului, am efectuat o analiză SHAP (SHapley Additive exPlanations) utilizând Kernel SHAP[27]. Am estimat, de asemenea, contribuția statutului mutațional al genelor individuale la incertitudinea predicției. În general, 19 dintre cele mai importante 20 de gene, sunt identice. Cu toate acestea, mutațiile genelor pot prezenta efecte opuse asupra incertitudinii predicției, în funcție de mutațiile co-ocurente.

Acest efect nu poate fi detectat atunci când se compară analizele SHAP ale eșantioanelor LUAD clasificate corect și incorect în raport cu predicțiile modelului (figurile 4.4(a) și 4.4(b)). Cu toate acestea, o analiză SHAP cu privire la incertitudinile de predicție (figurile 4.4(e) și 4.4(f)) arată că mutațiile în genele care indică LUAD, de ex, KRAS, pot crește incertitudinea predicției din cauza mutațiilor conflictuale în alte gene (Fig. 4.4(f)). Efecte similare pot fi observate în eșantioanele LUSC clasificate corect vs. incorect (figurile 4.4(c), 4.4(d), 4.4(g) și 4.4(h)).

Pentru a explora astfel de modele mutaționale la nivel de instanță, graficele în cascadă din figurile 4.5(a) și 4.5(b) ilustrează contribuția statutului mutațional al genelor individuale la o predicție corectă și, respectiv, incorectă a două probe LUAD selectate. În timp ce mutația în KRAS are un impact mare asupra clasificării corecte a TCGA-05-4390-01 ca eșantion LUAD, modelul mutațional neconcludent din eșantionul LUAD TCGA-93-A4JN-01, care cuprinde ATM mutant, dar și KRAS nemutat, crește incertitudinea predicției (Fig. 4.5(f)) datorită efectelor opuse asupra modelului (Fig. 4.5(b)). În schimb, în eșantionul LUAD TCGA-05-4390-01, mutația ATM scade incertitudinea predicției datorită unei mutații concomitente în KRAS (Fig. 4.5(e)).

Din aceste observații, deducem că probele cu o incertitudine mai mare a predicției prezintă tipare mutaționale care sunt ambigue în ceea ce privește histologia NSCLC, ceea ce ar putea indica histologiei de tip mixt. Pentru a investiga acest aspect mai detaliat și datorită lipsei unei etichete corepunzătoare (probele sunt adnotate fie ca LUAD, fie ca LUSC), am selectat cele mai încrezătoare gene recurent mutate pentru LUAD (BRAF, EGFR, KRAS și STK11) și LUSC (CDKN2A, NFE2L2, PIK3CA și PTEN) și am realizat o analiză comparativă între probele care poartă o mutație specifică LUAD, o o mutație specifică LUSC, sau ambele (mutații conducătoare mixte) ca un substitut pentru histologia de tip mixt. În mod remarcabil, performanța clasificării modelului nostru de profilare genomică este comparabilă în toate subgrupurile (Fig. 4.6(a)). Mai mult decât atât, predicțiile probelor cu mutații mixte prezintă o tendință spre estimări intermediare ale incertitudinii predicției (Fig. 4.6(b)). În cele din urmă, separarea curbelor de supraviețuire fără progresie (PFS) subliniază relevanța clinică a subgrupului de conducători mixt (figura 4.6(c)).

### 4.3.3 Discuții

Capacitatea modelului nostru profund de profilare genomică de a evalua certitudinea de predicție permite identificarea probelor neconcludente sau în afara distribuției, aspect indispensabil pentru ca abordările de învățare profundă să fie acceptate și implementate în practica clinică. În această lucrare, am instruit un ansamblu de o sută de modele de profilare genomică. În timp ce dimensiunea ansamblului poate fi încă optimizată pentru a reduce resursele de calcul, ne putem imagina, de asemenea, o

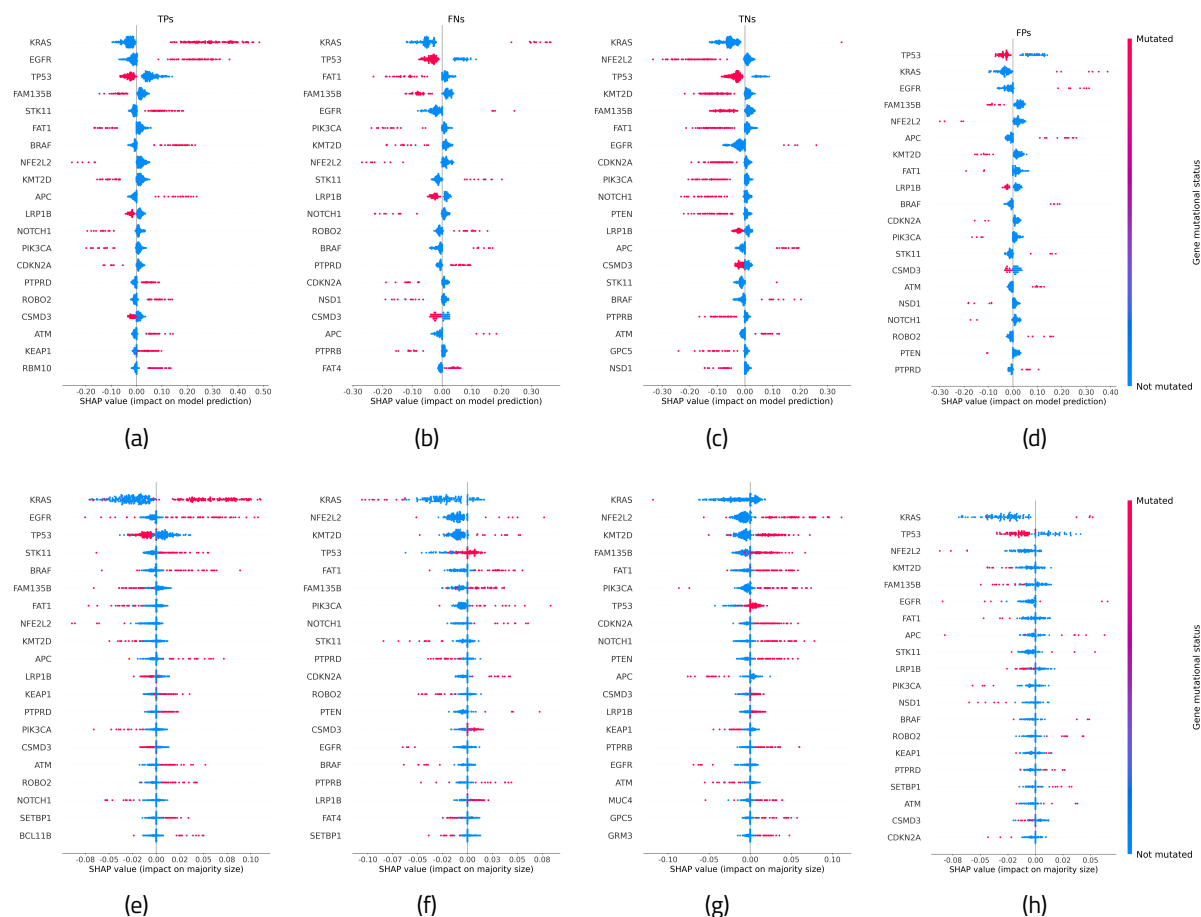


Figura 4.4: Rezumate SHAP la nivel de cohortă arată contribuția stării mutaționale a celor mai influente 20 de gene asupra (a-d) predicției modelului și (e-h) incertitudinii predicției pentru diferite categorii de predicție: (a, e) adevărat pozitiv (TP), (b, f) fals negativ (FN), (c, g) adevărat negativ (TN) și (d, h) predicții fals pozitive (FP).

configurație de învățare profesor-elev, oferind un model distilat care este mai practic pentru aplicarea clinică[139].

În general, estimarea incertitudinilor permite o investigație mai detaliată a predicțiilor modelului, deoarece contribuțiile caracteristicilor pot fi evaluate atât în raport cu rezultatele modelului, cât și cu incertitudinile acestora. De fapt, o astfel de analiză a identificat mutații concomitente care indică ambele subtipuri NSCLC în eșantioanele clasificate greșit, ceea ce explică performanța limitată observată în clasificarea subtipurilor NSCLC utilizând date mutaționale. Această observație este deosebit de importantă, deoarece modelele mutaționale specifice în NSCLC au fost, de asemenea, legate de eterogenitatea clinică. De exemplu, în NSCLC non-scuamos, s-a demonstrat că mutația KRAS interacționează cu mutațiile concomitente din TP53, STK11, PTPRD, RBM10 și ATM în ceea ce privește eficacitatea inhibitorului punctului de control imunitar[141]. Cele mai multe astfel de interacțiuni provin de la mutațiile de inițiere a tumorii în KRAS, TP53 și EGFR [142], rezultând modele de exclusivitate care s-au dovedit a fi asociate cu răspunsul atât la terapia țintită [143], cât și la imunoterapie [144].

Ca intrare în modelul nostru, am utilizat o codificare binară a statutului mutațional al fiecărei gene. Cu toate acestea, ar putea fi utilizate mai multe detalii privind impactul funcțional, de exemplu, prin luarea în considerare separată a mutațiilor cu impact scăzut, mediu și ridicat, obținute din „Variant Effect Predictor” (VEP)[145]. În plus, statisticile rezumative ar putea fi limitate la blocurile mutaționale sau la regiunile genetice care codifică domenii structurale. Cu toate acestea, ambele abordări pot

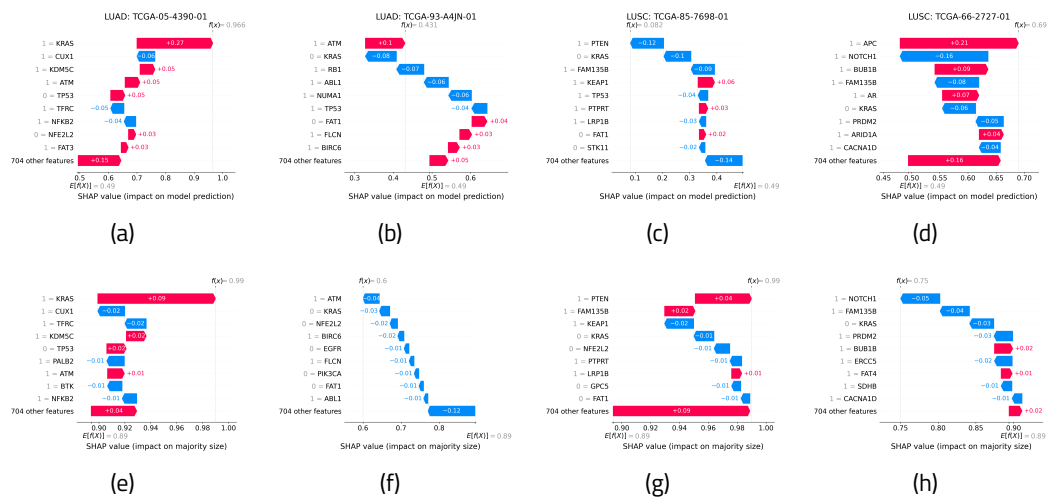


Figura 4.5: Diagrame în cascadă ale atribuțiilor caracteristicilor care arată contribuția stării mutaționale al celor nouă gene cu cel mai mare impact asupra (a-d) predicției modelului și (e-h) incertitudinii predicției pentru eșantioane selectate, reprezentând (a, e) un eșantion LUAD clasificat corect și (b, f) un eșantion LUAD clasificat incorect cu incertitudine de predicție scăzută și ridicată, precum și (c, g) un eșantion LUSC clasificat corect și (d, f) un eșantion LUSC clasificat incorect cu incertitudine de predicție scăzută și ridicată.  $f(x)$  corespunde (a-d) valorii de predicție (1 = LUAD, 0 = LUSC) și (b) incertitudinii de predicție asociate (1 = cea mai mică, 0,5 = cea mai mare incertitudine). Valorile caracteristicilor de 0 și 1 corespund genelor nemutate și, respectiv, mutante.

spori dispersia. În mod alternativ, modelele de limbaj proteic au fost descrise ca predictorii eficienți ai efectelor funcționale ale mutațiilor și pot fi utilizate pentru a codifica mutațiile fără sens ca o măsură continuă. În plus, ar putea fi necesar să se ia în considerare frecvențele alelelor pentru a îmbunătăți clasificarea probelor cu subtip mixt.

Pentru a aborda problema dimensionalității ridicate, unde varianța dintre eșantioane devine mare și dispersă [125], integrarea cunoștințelor anterioare despre interacțiunile proteină-proteină directe [146] sau indirecte [147] ca prejudecată inductivă relațională poate contribui la reducerea efectivă a spațiului parametrilor în raport cu abordarea naivă de modelare a tuturor termenilor de interacțiune și, astfel, poate permite formarea robustă a modelelor complexe pe cohorte mici.

Deoarece modelul de profilare genomică conștientă de expresie captează modelele mutaționale legate de histologie și de eficacitatea tratamentului, anticipăm că acesta va fi potrivit ca model de bază al datelor genomice pentru sarcini de prognostic sau predicție în aval, mai relevante din punct de vedere clinic. Pentru a spori generalizabilitatea sa, este posibil să fie necesară integrarea regularizării bazate pe modalități și sarcini suplimentare pentru a permite învățarea unei reprezentări mai holistice a fenotipului [148].

## 4.3.4 Materiale și metode

### 4.3.4.1 Date

Rezultatele prezente se bazează pe datele generate de rețeaua de cercetare TCGA: <https://www.cancer.gov/tcga>. Seturile de date „Pan-Cancer Atlas” derivate [130] au fost descărcate din cBioPortal [149]. Am inclus numai probe de tumori primare cu histologie AD și SCC pentru care erau disponibile atât date de mutație, cât și date de expresie. În plus, am limitat analiza noastră la genele adnotate în Recensământul Genei Cancerului (CGC) din Catalogul mutațiilor somatice în cancer (COSMIC, versiunea 95) [150], excluzând genele pentru care nu erau disponibile date de mutație sau de expresie, rezultând astfel 713 gene. În cele din urmă, am eliminat probele cu date incomplete pentru genele selectate.

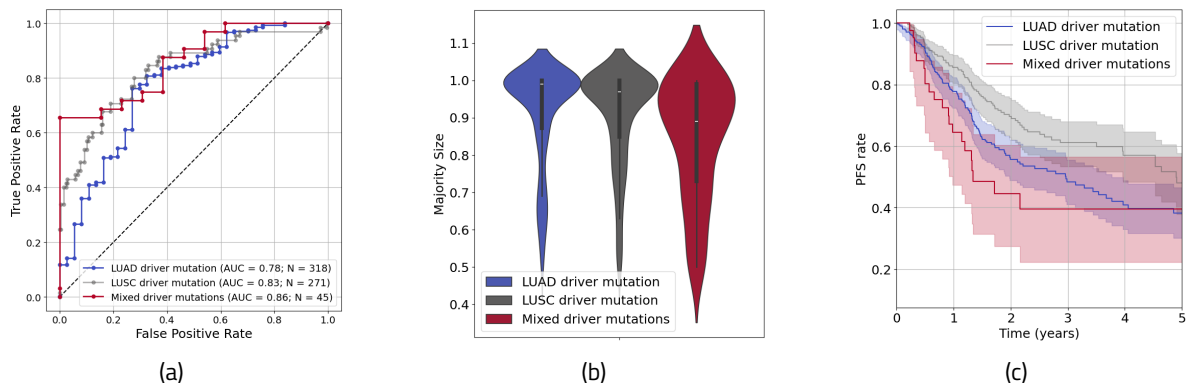


Figura 4.6: Analiza comparativă a subgrupurilor NSCLC pe baza prezenței unei mutații specifice LUAD (BRAF, EGFR, KRAS sau STK11), a unei mutații specifice LUSC (CDKN2A, NFE2L2, PIK3CA sau PTEN) sau a ambelor (mutații mixte ale driverului). (a) Performanța ansamblului de modele de profilare genomică pe subgrupuri. Linia punctată reprezintă curba ROC a unui clasificator aleatoriu. (b) Distribuția estimării incertitudinii predicției (dimensiunea majorității) în cadrul subgrupurilor. (c) Curbele Kaplan-Meier ale supraviețuirii fără progresie (PFS) în cadrul subgrupurilor.

#### 4.3.4.2 Metodologie

Antrenarea modelului a cuprins două etape. În primul rând, un model bazat pe expresia genelor (denumit model de profilare bazat pe expresie) a fost instruit pentru a clasifica probele în subtipuri NSCLC prin minimizarea sumei a doi termeni de pierdere a entropiei încrucișate binare (BCE) (unul pentru fiecare cap de predicție). Ulterior, un model bazat pe mutații genetice (denumit model de profilare genomică conștient de expresie) a fost antrenat pe aceeași sarcină de clasificare prin adăugarea la pierdere a unui termen de regularizare L2 care este calculat între reprezentarea latentă a modelului bazat pe mutații și reprezentarea latentă a modelului bazat pe expresie.

Din cauza numărului mic de eșantioane disponibile, toate experimentele au fost efectuate utilizând o schemă de validare încrucișată stratificată (de 10 ori). În timp ce, în fiecare iterație, un grup a fost reținut ca set de testare, eșantioanele celor nouă grupuri rămase au fost împărțite în continuare într-un set de antrenare (80%) și un set de validare (20%), stratificate din nou în funcție de subtipul NSCLC. Pentru a antrena un ansamblu de o sută de modele, setul de antrenare a fost eșantionat de o sută de ori (fiecare eșantion servind ca set de antrenament pentru un model). Eșantioanele au fost de aceeași dimensiune ca și setul de formare prin permiterea duplicatelor. Scorul de predicție al ansamblului de modele a fost obținut prin medierea valorilor prezise de toate modelele care indică subtipul NSCLC prezis majoritar. Pentru a agrega rezultatele peste grupurile de testare, am calculat media ratelor pozitive adevărate și false la fiecare punct de operare posibil din curba ROC și am calculat AUC-ul ulterior. Incertitudinile predicțiilor au fost calculate ca fracția modelelor care prezic clasa majoritară (numită dimensiunea majorității).

Modelele au fost antrenate utilizând optimizatorul Adam cu o rată inițială de învățare de  $10^{-4}$  și o dimensiune a lotului de 64. Pentru a contracara potențialele efecte de dezechilibru al claselor, am eșantionat fiecare lot de formare utilizând metoda propusă în [151]. În cadrul fiecărei iterații de validare încrucișată, cel mai bun model a fost selectat pe baza erorii minime pe setul de validare. Pentru a permite transferul optim de cunoștințe de la modelul de profilare bazat pe expresie, modelul de profilare genomică conștient de expresie a fost antrenat utilizând numai termenul de regularizare L2 timp de 30 de epoci (determinat empiric din curbele de învățare), înainte de a încorpora și termenii funcției de cost specifice sarcinii.

Pentru a crea subgrupuri NSCLC bazate pe prezența unei mutații specifice LUAD, specifice LUSC, sau ambelor (profile mixte), am selectat cele mai recurente gene mutate în LUAD (BRAF, EGFR, KRAS, STK11) sau LUSC (CDKN2A, NFE2L2, PIK3CA, PTEN) pe baza unei analize Pan-Cancer Atlas[152],

folosind un scor de consens mai mare de patru. TP53, care este o genă conducătoare cu mutație recurentă în ambele subtipuri, a fost exclusă.

## 4.4 Concluzii

În acest capitol am abordat limitările actuale ale fiabilității soluțiilor bazate pe învățarea profundă, explorând eficiența a două tehnici care abordează (1) lipsa inerentă de transparență și (2) încrederea excesivă în predicția probelor incerte. În timp ce aceste preocupări au reprezentat un obstacol în adoptarea generală a sistemelor DL-CAD în rutina clinică, constatările noastre sugerează că inteligența artificială sigură poate fi creată prin luarea în considerare a acestor limitări la proiectarea soluțiilor specifice.

Din punct de vedere al explicabilității și interpretabilității, am folosit aici o analiză Shap pentru a atribui scoruri de importanță relativă caracteristicilor de intrare, cuantificând impactul acestora asupra predicției finale. În mod specific, am investigat modul în care această abordare elucidează raționamentul a două rețele neuronale profunde în (1) ajustarea rezultatelor unui sistem DL-CAD existent pentru a identifica mai bine cancerul de prostată semnificativ din punct de vedere clinic, prin adăugarea de informații clinice și demografice suplimentare, și (2) clasificarea cancerului pulmonar fără celule mici în subtipuri – și anume adenocarcinom pulmonar (LUAD) și carcinom pulmonar cu celule scuamoase (LUSC) – pe baza profilurilor de mutații ale pacienților. Pentru fiecare scenariu, am efectuat două tipuri de analiză, una care oferă o imagine de ansamblu a importanței relative a caracteristicilor la nivel de cohortă (analiză generală) și una care descrie influența relativă a fiecărui element de intrare asupra predicției la nivel de caz.

Analiza generală oferă o imagine clară a modului în care predictorii selectați sunt corelați cu rezultatul, fiind potrivită pentru evaluarea inițială și caracterizarea modelului. Prin furnizarea unor astfel de rezultate, experții în domeniu ar putea obține o perspectivă mai bună asupra tiparelor și corelațiilor care sunt valorificate de model pentru a ajunge la anumite concluzii, sporind astfel semnificativ transparența acestuia. Prin urmare, raționamentul ar putea fi validat sau ignorat încă din faza incipientă, îmbunătățind semnificativ aspectele legate de siguranță în încorporarea unor astfel de soluții DL-CAD în rutinele existente. Pe de altă parte, explicațiile la nivel de instanță ar putea fi utilizate în timp real pentru a asigura în continuare o funcționalitate sigură a modelului implementat. În cazul în care raportarea este concepută pentru a descrie, de asemenea, modul în care sistemul a ajuns la o anumită concluzie, clinicienii ar putea confirma sau infirma rapid rezultatul acesteia, îmbunătățind astfel în mod semnificativ fiabilitatea și menținând în același timp volumul lor de muncă la niveluri rezonabile.

În plus, estimările de incertitudine pot fi folosite pentru a reduce și mai mult sarcina de lucru a personalului medical, menținând în același timp un cadru de funcționare sigur. Acest capitol prezintă modul în care ansamblurile de modele pot fi valorificate pentru a produce astfel de scoruri de incertitudine, permițând modelului să identifice cazurile limită care ar necesita o atenție specială din partea clinicienilor. Exemple de astfel de situații ar putea reprezenta cazuri neobișnuite care nu au fost bine reprezentate în datele de antrenare, cazuri ambigue în care apar tipare contradictorii etc. Prin urmare, tehnicile de cuantificare a incertitudinii permit modelului să spună "Nu știu", decât să încerce să furnizeze predicții predispușe la erori, ceea ce reprezintă o proprietate obligatorie a unui sistem autonom sigur. În plus, estimările de certitudine care însoțesc predicția reduc semnificativ volumul de muncă al practicienilor din domeniul sănătății, care își pot îndrepta atenția doar asupra fracțiunii de cazuri în care sistemul DL-CAD este predispus la erori.

În general, concluzionăm că fiecare soluție bazată pe inteligența artificială ar trebui să posede proprietățile descrise mai sus, și anume explicabilitatea, interpretabilitatea și estimarea incertitudinii. Am demonstrat aici beneficiile remarcabile ale furnizării unui sistem demn de încredere, care asigură o funcționalitate sigură în practică, fără a crește drastic complexitatea generală a procesului de dezvoltare. Prin urmare, merită să furnizăm inteligență artificială sigură!

## 4.5 Discuții

Rezultatele raportate în acest capitol arată că modelele de învățare profundă, considerate anterior cutii negre, pot fi îmbunătățite pentru a reflecta mai bine procesul de raționament care stă la baza predicțiilor lor. Am demonstrat că inteligența artificială explicabilă poate fi obținută prin utilizarea unei analize post-hoc care oferă estimări ale ponderii fiecărei caracteristici de intrare în construirea unei anumite decizii specifice, indiferent de complexitatea arhitecturală sau de strategia de optimizare a modelului. Mai mult, pe lângă beneficiile inițiale care derivă dintr-o mai bună caracterizare a modelului, tehnicile de interpretabilitate pot susține funcționarea sigură a sistemului în timpul exploatării.

Deoarece algoritmi de învățare automată se bazează în mare măsură pe cantitatea, calitatea și completitudinea datelor de antrenare, estimările incertitudinii asigură o utilizare adecvată în practică. Acest capitol a furnizat dovezi cu privire la modul în care astfel de estimări identifică cazurile ambigue care ar trebui să fie tratate mai degrabă de experți în domeniu, pentru a asigura servicii medicale adecvate. Având în vedere potențialul de vătămare a pacienților, fiecare sistem autonom ar trebui să dispună de astfel de mecanisme de securitate pentru a minimiza riscul de a comite erori.

De remarcat, rezultatele prezentate aici reprezintă doar o validare conceptuală și, prin urmare, nu un set stabilit de bune practici. Alte tehnici sau abordări pentru a face sistemele DL-CAD explicabile, interpretabile și conștiente de incertitudine ar putea fi mai potrivite de la un caz de utilizare la altul. De exemplu, estimările incertitudinii bazate pe ansambluri pot să nu fie adecvate în aplicațiile în timp real, unde timpul maxim de inferență ar putea fi limitat. Deși există un compromis între numărul de modele din ansamblu și timpul de inferență, estimările incertitudinii pot fi mai puțin precise în cazul utilizării unui număr insuficient de instanțe. Cu toate acestea, tehnicile de distilare a cunoștințelor ar adresa limitarea menționată mai sus, însă evaluarea fezabilității acestora nu a făcut obiectul prezentei lucrări.

Din punctul de vedere al importanței caracteristicilor, multe studii au încercat să rezolve problemele de transparență ale modelelor de învățare profundă [24, 25, 26, 27, 28]. În timp ce analiza Shap reprezintă o soluție promițătoare, metoda „KernelShap”, care este utilizată pentru a explica raționamentul rețelelor neuronale profunde, folosește un set de procese stocastice care simulează diverse coaliții ale caracteristicilor de intrare și impactul acestora asupra rezultatului. Datorită potențialelor prejudecăți care provin din strategia de eșantionare, aceste procese ar putea reprezenta o sursă de eroare. Cu toate acestea, explicațiile algoritmilor de învățare automată relativ simpli (de exemplu, clasificatorii RandomForest) sunt în mod natural mai fiabile în comparație cu soluțiile sofisticate. Prin urmare, pentru a obține cea mai bună transparență a sistemelor autonome, un principiu de ghidare ar trebui să fie utilizarea soluțiilor cu complexitate minimă pentru o anumită sarcină, atunci când metricele de performanță nu justifică opusul.

În plus, pe lângă faptul că permite clinicienilor să evalueze fiabilitatea predicțiilor, inteligența artificială explicabilă are potențialul de a identifica posibile modele necunoscute în date, care ar putea duce la o mai bună înțelegere a patologiilor. Prin urmare, astfel de explicații ar putea declanșa studii clinice concepute pentru a explora în continuare impactul diferitelor caracteristici asupra anumitor afecțiuni medicale, posibil conducând la descoperiri revoluționare în medicină.

Cu toate acestea, merită menționat faptul că beneficiile utilizării sistemelor DL-CAD în medii adecvate nu se limitează doar la reducerea volumului de muncă a personalului medical, ci și la precizia generală a diagnosticului. Deoarece s-a demonstrat că precizia diagnosticului depinde de nivelul de experiență al radiologului [11, 10, 9], adoptarea unor astfel de soluții autonome în rutinele clinice ar putea reprezenta un avantaj extraordinar în sprijinirea clinicienilor, în special la începutul carierei lor.

În general, explicabilitatea, interpretabilitatea și conștientizarea incertitudinii reprezintă elemente de bază pentru o inteligență artificială sigură, care ar trebui să devină în prezent o practică standard pentru furnizarea de soluții autonome în diverse industrii, în special atunci când eventualele erori ar putea avea consecințe dezastruoase pentru beneficiarul final.

## 5. Concluzii finale

---

Concluzii  
Contribuții originale  
Diseminarea rezultatelor cercetării  
Discuții

---

### 5.1 Concluzii

Această teză de doctorat demonstrează modul în care provocările majore în dezvoltarea de soluții bazate pe învățarea profundă pentru diverse nevoi clinice pot fi abordate prin utilizarea unor paradigme precum auto-supervizarea, generarea de date sintetice, cuantificarea incertitudinii și estimarea importanței caracteristicilor. În ciuda potențialului remarcabil de îmbunătățire a practicilor clinice actuale și, în cele din urmă, a rezultatelor pentru pacienți, adoptarea imediată a IA în activitățile medicale de rutină este împiedicată de o serie de preocupări, incluzând confidențialitatea pacienților, cerințele de etichetare a datelor și aspectele legate de siguranță în funcționare.

Progresele tehnologice recente din industria semiconductorilor, și ulterior puterea de calcul au plasat industria medicală într-o fază de transformare, schimbând în mod substanțial practicile actuale în direcția adoptării tehnicilor de diagnostic minim invazive, a planificării terapiei ghidate de imagistică, a furnizării și a monitorizării sănătății. Deși prezintă beneficii extraordinare pentru pacienți, orientările actuale au crescut semnificativ volumul de muncă al practicienilor clinici, cum ar fi radiologii, fiind astfel necesară adaptarea soluțiilor software pentru a sprijini o furnizare sustenabilă de servicii medicale, menținând în același timp volumul de muncă a clinicienilor la niveluri rezonabile. Abordările bazate pe învățarea profundă s-au dovedit prima opțiune actuală în abordarea diferitelor probleme clinice. În prezent, singura paradigmă de antrenare cu adevărat fezabilă este învățarea supervizată. În această măsură, o etapă adecvată de etichetare a datelor reprezintă o condiție prealabilă pentru formarea oricărui model de învățare profundă, etapă care ar putea crește și mai mult volumul de muncă al personalului medical. În acest scop, capitolul 2 prezintă o abordare auto-supervizată pentru formarea rețelelor neuronale în extrapolarea imaginilor medicale, în care perechile de intrare-ieșire sunt simulate aleatoriu la momentul antrenării, nefiind astfel necesară nicio etapă de adnotare manuală.

În plus, dintre toate tipurile de informații personale, datele medicale fac parte dintre cele mai restricționate, ca urmare a faptului că încorporează informații sensibile ce pot periclita confidențialitatea pacienților. Deși corecte și eficiente din punct de vedere fundamental, constrângerile actuale privind circulația datelor medicale au un impact negativ asupra posibilității de exploatare extinsă a inteligenței artificiale în diverse scenarii clinice, în care practicile actuale ar putea fi îmbunătățite. În timp ce o antrenare robustă a algoritmilor de învățare automată depinde în mare măsură de disponibilitatea unor seturi de date la scară largă, complete și calitative, colectarea unor astfel de baze de date în practică este adesea imposibilă. Capitolul 3 abordează acest obstacol prin descrierea unei metodologii de creare de date longitudinale sintetice ale pacienților, fără riscuri legate de confidențialitate, care pot fi utilizate în antrenarea rețelelor neuronale pentru a răspunde diverselor nevoi clinice.

Exemplificăm această idee prin abordarea unui caz de utilizare clinică extrem de relevant, și anume stratificarea pacienților cu cancer de prostată.

Cu toate acestea, modelele AI sunt susceptibile de a furniza predicții incorecte atunci când li se prezintă eșantioane insuficient reprezentate în baza lor de date de antrenare. Lipsa inerentă de transparență a algoritmilor de învățare profundă îi împiedică pe utilizatorii finali să înțeleagă cu adevărat procesele de raționament din spatele anumitor predicții, sporind astfel scepticismul și reticența în adoptarea unor astfel de soluții în rutina lor clinică, în special având în vedere impactul negativ enorm pe care o eroare l-ar putea avea asupra bunăstării pacienților. Capitolul 4 abordează această preocupare prin îmbogățirea modelelor de învățare profundă cu proprietăți de explicabilitate, interpretabilitate și conștientizare a incertitudinii. S-a demonstrat că toate aceste capacități adiționale joacă un rol semnificativ în sporirea robusteții și stabilității sistemelor bazate pe DL, acționând ca mecanisme de securitate pentru a asigura o utilizare sigură în rutinele clinice. Am demonstrat aici că ansamblurile de modele și analiza Shap pot fi utilizate cu succes pentru a furniza estimări ale incertitudinii și explicații cantitative ale modului în care un model a dedus anumite rezultate, urmărind câteva cazuri de utilizare relevante, și anume predicția cancerului de prostată clinic semnificativ și clasificarea cancerului pulmonar fără celule mici în subtipuri.

### **5.1.1 Învățare auto-supervizată pentru extrapolarea și coregistrarea imaginilor medicale**

Capitolul 2 prezintă o abordare de învățare auto-supervizată pentru a spori robustețea sistemelor de ghidare bazate pe imagistica CT intraoperatorie. Din pricina câmpului vizual relativ redus al achizițiilor în timp real și, prin urmare, a informațiilor contextuale reduse, etapa de coregistrare necesară pentru alinierea acestor imagini cu scanări preoperatorii de înaltă calitate este predispusă la erori. Prin urmare, am arătat cum algoritmii de învățare profundă pot fi utilizați cu succes în extinderea câmpului vizual al imaginilor CT subțiri, prin utilizarea unui cadru generativ adversarial. Rezultatele prezentate în această teză demonstrează impactul pozitiv pe care acești algoritmi îl pot avea asupra stabilității generale a sistemului, reducând erorile de coregistrare mediane cu un ordin de magnitudine, ceea ce, în cele din urmă, duce la o ghidare mai robustă a intervențiilor chirurgicale.

Pe de altă parte, în contextul acestui studiu, rolul paradigmei de învățare autosupravegheată este esențial. Având în vedere natura acestui caz de utilizare, colectarea perechilor de date intrare-ieșire pentru antrenarea algoritmilor într-o manieră supravegheată nu este doar costisitoare și epuizantă din punct de vedere al resurselor, ci și nefezabilă din cauza necesității de a alinia perfect intrările (imaginile intraoperatorii) cu ieșirile (achizițiile preoperatorii). Prin urmare, experimentele prezentate în cadrul acestei teze demonstrează modul în care paradigmele de învățare auto-supervizată ar putea acționa ca facilitatori pentru unele scenarii clinice considerate anterior nerealiste.

### **5.1.2 Generarea de date sintetice pentru stratificarea cancerului de prostată**

Ca răspuns la provocările actuale legate de colectarea de seturi de date longitudinale la scară largă, complete și calitative, capitolul 3 prezintă o abordare sistematică pentru a crea fișe medicale electronice sintetice care să reflecte parcursul pacienților cu cancer de prostată de la diagnostic la monitorizarea tratamentului. Prin proiectarea atentă a hărților de îngrijire clinică, combinate cu informațiile de incidență și statisticile relevante ale bolii, această teză a demonstrat că pot fi create fișe pur sintetice coerente și realiste ale pacienților, și că acestea pot fi utilizate în cele din urmă pentru a răspunde diferitelor nevoi clinice.

Dintr-o perspectivă calitativă, am demonstrat că cohortele noastre generate sintetic prezintă proprietăți distribuționale similare cu diferite seturi de date reale, dar mai puțin complete, prezentate în literatura de specialitate. De remarcat, concluzia menționată mai sus nu se referă numai la măsurătorile individuale de laborator sau la rezultatele diferitelor investigații, ci și la natura longitudinală a datelor: am demonstrat aici coerența între diferitele teste și proceduri de diagnosticare, opțiuni de tratament și efectele secundare ulterioare.



Ca rezultat al acestei activități, generatorul de date sintetice poate produce, în esență, un număr nelimitat de eșantioane de înaltă fidelitate care ar putea fi utilizate într-o gamă largă de activități, inclusiv modelare, analiză de date, dezvoltare de prototipuri, testare de sisteme etc. Am exemplificat acest lucru prin antrenarea unui model de stratificare a cancerului de prostată pentru a deduce stadiul TNM al pacienților pe baza întregului lor dosar electronic. Precizia de clasificare aproape perfectă obținută pe seturile de testare servește drept dovadă suplimentară a coerenței datelor, oferind o verificare a calității acestora. În plus, componente ale acestor predictor rezultați ar putea acționa ca modele fundamentale care pot fi ajustate în continuare pe seturi de date relativ mai mici pentru a îndeplini o mare varietate de sarcini relevante din punct de vedere clinic, valorificând astfel reprezentările abstracte ale caracteristicilor învățate din datele noastre sintetice.

### 5.1.3 Inteligența artificială sigură

Capitolul 4 prezintă câteva tehnici care ar putea fi atașate oricărui model neuronal profund pentru a crește stabilitatea și transparența generală a sistemului. Am oferit aici exemple concrete ale modului în care estimările importanței caracteristicilor pot ameliora în mod semnificativ natura de „cutie neagră” a acestor modele, în timp ce cuantificarea incertitudinii poate fi utilizată pentru a asigura funcționarea sigură în cadrul rutinelor clinice.

Concret, valorile Shapely au fost considerate un indicator fiabil al modului în care diferite caracteristici au influențat modelul în furnizarea unei anumite predicții, oferind estimări cantitative ale importanței relative pe care fiecare caracteristică de intrare a avut-o asupra predicției finale. Acest tip de analiză poate fi efectuat fie la nivel de cohortă, fie la nivel de instanță, fiecare având un rol important distinct în furnizarea de inteligență artificială sigură: în timp ce cea din urmă vizează îmbunătățirea aspectelor legate de siguranță ale utilizării unor astfel de sisteme în practică prin furnizarea unei explicații împreună cu predicția, analizele globale permit o caracterizare și o evaluare a modelului într-un stadiu incipient prin furnizarea de informații privind proprietățile învățate din date. Această teză demonstrează fezabilitatea și utilitatea acestei analize în (1) îmbunătățirea predicției cancerului de prostată semnificativ din punct de vedere clinic prin utilizarea unor caracteristici clinice și demografice suplimentare și (2) diferențierea leziunilor de cancer pulmonar cu celule mari în subtipuri pe baza unor profiluri genomice eterogene: în ambele scenarii, analiza Shap a furnizat informații clare cu privire la modul în care diferite caracteristici au orientat inferența către anumite predicții, punând astfel în lumină pe procesul de raționament.

În plus, a fost explorată o abordare de cuantificare a incertitudinii bazată pe modele ansamblu pentru a identifica eșantioanele neconcludente care pot conduce, de asemenea, la predicții nesigure ale subtipului de cancer pulmonar. Deoarece rețelele neuronale tind să fie prea încrezătoare în inferențele lor, astfel de tehnici contribuie în mod semnificativ, la furnizarea soluțiilor sigure, oferind modelelor o modalitate de a identifica cazurile în care predicția lor ar putea fi predispusă la erori și, prin urmare, de a oferi un răspuns „Nu știu”, mai degrabă decât de a încerca să furnizeze un anumit răspuns. Secțiunea 4.3 oferă o demonstrație a modului în care o măsură de incertitudine bazată pe modele ansamblu ar putea identifica eșantioanele neconcludente care nu pot fi diferențiate pe baza setului de caracteristici considerate ca date de intrare. În general, această capacitate suplimentară poate fi interpretată ca o măsură de siguranță, care îmbunătățește semnificativ stabilitatea sistemului și aspectele legate de fiabilitate, menținând în același timp volumul de muncă al clinicienilor la niveluri ușor de gestionat: de exemplu, cazurile cu incertitudine scăzută pot fi tratate în mod autonom, în timp ce predicțiile cu incertitudine ridicată ar putea declanșa un semnal de alarmă, necesitând astfel asistență suplimentară din partea personalului medical.

## 5.2 Contribuții originale

În primul rând, toate contribuțiile făcute în această teză se bazează pe o fază exhaustivă de documentare, în care **a fost compilată o listă cuprinzătoare de provocări specifice dezvoltării soluțiilor**

**bazate pe învățarea profundă în industria medicală.** În special, (1) cererile substanțiale de date de antrenare sunt adesea imposibile din cauza constrângerilor privind aspectele legate de confidențialitate, a completitudinii și/sau a lipsei de adnotări și (2) reticența față de lipsa de transparență inerentă a rețelelor neuronale profunde reprezintă adesea un blocaj în adoptarea generală a unor astfel de soluții în rutina clinică. **În plus, un set de soluții pentru aceste limitări au fost definite și exploatate printr-o serie de cazuri de utilizare relevante din punct de vedere clinic, aducând contribuții semnificative în domeniu, după cum este descris în continuare.**

### **Învățare auto-supervizată pentru extrapolarea și coregistrarea imaginilor medicale**

**Ideea de a utiliza o soluție de extrapolare bazată pe rețele neuronale ca o etapă premergătoare coregistrării imaginilor cu câmp vizual mic reprezintă o abordare originală care, în cele din urmă, a îmbunătățit performanța sistemelor de ghidare a imaginilor CT intervenționale.** Extinderea câmpului vizual redus al imaginii intraoperatorii a furnizat mai multe informații contextuale pentru etapa ulterioară de aliniere, care a devenit mai robustă, producând o stabilitate îmbunătățită a sistemului de ghidare, demonstrată printr-o reducere a erorilor mediane cu un ordin de magnitudine.

Paradigma învățării autosupervizate a fost utilizată pentru a crea perechi intrare-ieșire care au ghidat optimizarea modelului generativ de extrapolare printr-un proces formare adversarială. Au fost explorate două abordări diferite: În primul rând, a fost investigată o abordare de extrapolare asimetrică, **unde un pas suplimentar de aliniere a fost propus pentru a deriva informațiile spațiale ale imaginilor extrapolate. Concomitent, a fost evaluată o strategie relativ mai simplă de extrapolare simetrică, subliniind avantajele și dezavantajele fiecărei metode în contextul fezabilității clinice generale.**

### **Generarea de date sintetice pentru stratificarea cancerului de prostată**

**Au fost concepute module de înaltă calitate pentru cancerul de prostată în scopul generării de date longitudinale coerente care reflectă diferitele faze ale parcursului pacientului, de la diagnostic la stadializare, tratament și, în cele din urmă, la monitorizare. Prin valorificarea observațiilor experților medicali, au fost definite hărți de îngrijire clinică cuprinzătoare pentru a integra toate etapele relevante prin care trece un pacient pentru diagnosticarea, tratarea și monitorizarea cancerului de prostată. În plus, pentru a asigura plauzibilitatea și coerența datelor generate, a fost compilată o listă cuprinzătoare de publicații științifice relevante pentru a extrage proprietățile statistice ale cohortelor reale și pentru a constrânge, în consecință, toate stările stohastice ale mapelor de îngrijire.** Mai departe, calitatea datelor pur sintetice produse în această lucrare a fost evaluată pentru a asigura echivalența statistică cu cohortele reale. În plus, în timp ce datele reale prezentate în diverse studii clinice respectă un set clar de criterii de includere/excludere, fiind de obicei orientate către o anumită fază a bolii sau agresivitate, generatorul de date sintetice prezentat în această teză nu duce lipsă de diversitate, fiind capabil să producă întregul spectru de cazuri (de exemplu, de la risc scăzut la risc foarte ridicat, de la pacienți localizați la pacienți regionali și/sau metastatici, de la cazuri tratate printr-un regim de supraveghere activă la cazuri foarte avansate tratate cu îngrijiri paliative etc.).

Mai mult, această lucrare a evaluat fezabilitatea utilizării seturilor de date longitudinale generate sintetic în dezvoltarea modelelor predictive. **O abordare de procesare a limbajului natural a fost utilizată, în mod original, în stratificarea pacienților cu cancer de prostată prin sistemul de stadializare TNM, pe baza descrierilor codurilor clinice prezentate în dosarele medicale electronice, având rezultate preliminare foarte promițătoare.**

### **Inteligența artificială sigură**

**Pentru a obține o discriminare superioară a cancerului de prostată semnificativ din punct de vedere clinic, a fost propusă o etapă de rectificare, în care informațiile clinice și demografice au**

**fost asociate rezultatului unui sistem de diagnostic asistat de calculator de ultimă generație, pentru a îmbunătăți precizia predicției.** Rezultatele prezentate în această teză demonstrează în mod clar beneficiile acestei abordări, nu numai prin evaluarea performanței pe seturi de testare, ci și prin utilizarea estimărilor importanței caracteristicilor pentru a spori transparența generală a sistemului. În plus, perspectivele derivate din această analiză sunt în concordanță cu cercetările publicate în literatura de specialitate, indicând astfel fezabilitatea generală a adoptării unor astfel de soluții în practicile clinice.

În timp ce informațiile genomice sunt recunoscute ca fiind asociate cu predispoziția la anumite boli și cu răspunsul la tratament, abordarea standard pentru dezvoltarea modelelor prognostice sau predictive este de a preselecciona un mic subset de gene asociate cu fenotipul țintă. Dimpotrivă, **această lucrare a propus un cadru de modelare care vizează găsirea de interacțiuni între un număr substanțial mai mare de gene, care ar putea fi valorificate pentru a crea o amprentă a fenotipului pacientului (de exemplu, modelul fundațional).** În ciuda clasificării subtipului de cancer pulmonar utilizată aici ca sarcină de pretext, astfel de modele de profilare genomică pot fi ulterior ajustate și specializate pentru scenarii mai relevante din punct de vedere clinic, cum ar fi predicția răspunsului la tratament, sau prognosticul.

Pentru a adresa eterogenitatea inerentă a profilelor mutaționale, **a fost explorată o tehnică de transfer de cunoștințe, în care reprezentările abstracte ale fenotipurilor create pe baza datelor de expresie genică au fost utilizate pentru a regulariza antrenarea modelelor de profilare genomică, demonstrând fezabilitatea generală prin vizualizări condensate ale spațiului latent.** În plus, având în vedere că ambiguitatea profilelor mutaționale nu este o situație neobișnuită, **această lucrare prezintă o abordare bazată pe modele ansamblu pentru a cuantifica incertitudinea predicției și, astfel, pentru a identifica eșantioanele neconcludente care ar putea necesita investigații suplimentare.** Rezultatele prezentate în această teză relevă o bună asociere a metricii de incertitudine propuse cu rezultate extrem de relevante, cum ar fi progresia cancerului. În același timp, analiza contribuției caracteristicilor a expus modelele mutaționale contradictorii care reprezintă factori limitativi ai performanței de clasificare, ameliorând în mod semnificativ natura netransparentă a acestor modele neuronale profunde, și astfel, sporind fiabilitatea generală a sistemului.

### 5.2.1 Sinteza contribuțiilor

Tabelul 5.1: Sinteza contribuțiilor și a diseminării rezultatelor cercetării

Nr. criteriu	Contribuție	Capitol / Secțiune	Diseminare
1	A fost efectuată o recenzie exhaustivă a literaturii de specialitate pentru a defini provocările confruntate în prezent în dezvoltarea de soluții bazate pe inteligența artificială pentru diverse nevoi clinice, provocări care provin din insuficiența datelor și lipsa transparenței. Soluțiile la aceste limitări au fost ulterior explorate prin diverse scenarii clinice.	1	Puiu, A., Vizitiu, A., Nita, C., Itu, L., Sharma, P., & Comaniciu, D. (2021). Privacy-Preserving and Explainable AI for Cardiovascular Imaging. <i>Studies in Informatics and Control</i> , 30(2), 21–32. <a href="https://doi.org/10.24846/v30i2y202102">https://doi.org/10.24846/v30i2y202102</a> .

Continuat pe pagina următoare

**Tabelul 5.1 – Continuare a paginii anterioare**

2	În contextul intervențiilor ghidate imagistic, o contribuție originală provine din alinierea îmbunătățită a imaginilor intraoperatorii și pre-operatorii, care este alimentată de o fază de extrapolare pentru creșterea câmpului vizual al achizițiilor CT subțiri.	2	Puiu, A., Reangamornrat, S., Pheiffer, T., Itu, L. M., Suci, C., Ghesu, F. C., & Mansi, T. (2022). Generative Adversarial CT Volume Extrapolation for Robust Small-to-Large Field of View Registration. Applied Sciences (Switzerland), 12(6). <a href="https://doi.org/10.3390/app12062944">https://doi.org/10.3390/app12062944</a>
3	O altă contribuție originală constă în proiectarea unei etape suplimentare de aliniere pentru obținerea informațiilor spațiale ale imaginilor extrapolate, permițând utilizarea unei abordări asimetrice.	2.3	
4	Au fost concepute module Synthea în scopul de a genera fișe medicale electronice longitudinale coerente pentru pacienții cu cancer de prostată. Prin constrângerea core-spunzătoare a tuturor stărilor stohastice din aceste module, datele generate sintetic prezintă proprietăți statistice similare cohortelor reale prezentate în literatura de specialitate.	3, 3.3.1	Aceste rezultate nu au fost publicate
5	Utilitate datelor sintetice rezultate în dezvoltarea modelelor predictive a fost explorată prin intermediul unui caz clinic relevant, și anume stratificarea pacienților cu cancer de prostată utilizând sistemul de stadializare TNM. În mod original, a fost propusă o abordare de prelucrare a limbajului natural prin valorificarea descrierilor codurilor clinice randomizate în dosarele electronice.	3.3.2	Aceste rezultate nu au fost publicate
6	În ceea ce privește identificarea cancerului de prostată clinic semnificativ, adăugarea de caracteristici clinice suplimentare la o soluție DL-CAD de ultimă generație s-a dovedit a juca un rol important în îmbunătățirea performanței generale. În plus, cuantificarea impactului relativ al fiecărei caracteristici de intrare asupra predicției finale a sporit suplimentar fiabilitatea și transparența metodei propuse.	4.2	Aceste rezultate nu au fost publicate

Continuat pe pagina următoare

**Tabelul 5.1 – Continuare a paginii anterioare**

7	O altă contribuție originală a acestei teze constă în dezvoltarea unui predictor al subtipului de cancer pulmonar bazat pe un set vast de gene, cu scopul final de a crea un model neuronal profund de proliferare genomică, care poate fi reutilizat ulterior pentru a aborda diverse probleme clinice nesatisfăcute, cum ar fi prognosticul sau predicția răspunsului la tratament.	4.3	
8	Având în vedere natura datelor genomice în care eterogenitatea inerentă reprezintă provocări semnificative pentru abordările de învățare automată, a fost utilizată o schemă de transfer de cunoștințe pentru îmbunătățirea reprezentării latente prin utilizarea expresiilor genetice mai puțin frecvent analizate în rutinele clinice, dar mai informative. În plus, utilizarea estimărilor importanței caracteristicilor împreună cu estimările incertitudinii predicției a îmbunătățit fiabilitatea metodei, permițând identificarea probelor neconcludente, care par a fi asociate cu prognosticul.	4.3	Puiu, A., Gómez Tapia, C., Weiss, M. E. R., Singh, V., Kamen, A., & Siebert, M. (2024). Prediction uncertainty estimates elucidate the limitation of current NSCLC subtype classification in representing mutational heterogeneity. <i>Scientific Reports</i> , 14(1), 6779. <a href="https://doi.org/10.1038/s41598-024-57057-3">https://doi.org/10.1038/s41598-024-57057-3</a>

### 5.3 Diseminarea rezultatelor cercetării

Ca urmare a cercetărilor efectuate pe parcursul acestui program de doctorat, au fost obținute 14 publicații în calitate de autor sau coautor în diverse reviste științifice.

Trei articole au fost publicate în calitate de prim autor în jurnale indexate ISI, în timpul programului de doctorat:

- Puiu, A., Gómez Tapia, C., Weiss, M. E. R., Singh, V., Kamen, A., & Siebert, M. (2024). Prediction uncertainty estimates elucidate the limitation of current NSCLC subtype classification in representing mutational heterogeneity. *Scientific Reports*, 14(1), 6779. <https://doi.org/10.1038/s41598-024-57057-3>. (impact factor: 4.6, Q2)
- Puiu, A., Reaungamornrat, S., Pfeiffer, T., Itu, L. M., Suci, C., Ghesu, F. C., & Mansi, T. (2022). Generative Adversarial CT Volume Extrapolation for Robust Small-to-Large Field of View Registration. *Applied Sciences*, 12(6), 2944. <https://doi.org/10.3390/app12062944>. (impact factor: 2.7, Q3)
- Puiu, A., Vizitiu, A., Nita, C., Itu, L., Sharma, P., & Comaniciu, D. (2021). Privacy-Preserving and Explainable AI for Cardiovascular Imaging. *Studies in Informatics and Control*, 30(2), 21–32. <https://doi.org/10.24846/v30i2y202102>. (impact factor: 1.6, Q4)

Nouă articole au fost publicate în calitate de coautor în jurnale indexate ISI, în timpul programului de doctorat:

- Benedek, T., Ferent, I., Benedek, A., Cernica, D., Nita, C., Puiu, A., Itu, L., Rapaka, S., Puneet, S., & Benedek, I. S. (2020). P1434 Evolution of coronary wall shear stress following implantation of bioabsorbable vascular scaffolds—First results of a 1-year follow-up pilot study. *European Heart Journal - Cardiovascular Imaging*, 21(Supplement 1), <https://doi.org/10.1093/ehjci/jez319.863>. (impact factor: 6.20, Q1)
- Ciusdel, C., Turcea, A., Puiu, A., Itu, L., Calmac, L., Weiss, E., Margineanu, C., Badila, E., Berger, M., Redel, T., Passerini, T., Gulsun, M., & Sharma, P. (2020). Deep neural networks for ECG-free cardiac phase and end-diastolic frame detection on coronary angiographies. *Computerized Medical Imaging and Graphics*, 84, 101749. <https://doi.org/10.1016/j.compmedimag.2020.101749>. (impact factor: 5.70, Q1)
- Vizitiu, A., Nita, C. I., Puiu, A., Suciu, C., & Itu, L. M. (2020). Applying Deep Neural Networks over Homomorphic Encrypted Medical Data. *Computational and Mathematical Methods in Medicine*, 2020, 1–26. <https://doi.org/10.1155/2020/3910250>. (impact factor: 0.94, Q3)
- Nita, C.-I., Puiu, A., Bunescu, D., Mihai Itu, L., Mihalef, V., Chintalapani, G., Armstrong, A., Zampi, J., Benson, L., Sharma, P., & Rapaka, S. (2022). Personalized Pre- and Post-Operative Hemodynamic Assessment of Aortic Coarctation from 3D Rotational Angiography. *Cardiovascular Engineering and Technology*, 13(1), 14–40. <https://doi.org/10.1007/s13239-021-00552-9>. (impact factor: 1.80, Q3)
- Ploscaru, V., Popa-Fotea, N.-M., Calmac, L., Itu, L. M., Mihai, C., Bataila, V., Dragoescu, B., Puiu, A., Cojocaru, C., Costin, M. A., & Scafa-Udriste, A. (2022). Artificial intelligence and cloud based platform for fully automated PCI guidance from coronary angiography—study protocol. *PLOS ONE*, 17(9), e0274296. <https://doi.org/10.1371/journal.pone.0274296>. (impact factor: 3.70, Q2)
- OGREZEANU, I., VIZITIU, A., CIUSDDEL, C., PUIU, A., COMAN, S., BOLDIȘOR, C., ITU, A., DEMETER, R., MOLDOVEANU, F., SUCIU, C., & ITU, L. (2022). Privacy Preserving and Explainable AI in Industrial Applications. In *Applied Sciences (Switzerland)* (Vol. 12, Issue 13). MDPI. <https://doi.org/10.3390/app12136395>. (impact factor: 2.70, Q3)
- Hatfaludi, C. A., Tache, I. A., Ciușdel, C. F., Puiu, A., Stoian, D., Itu, L. M., Calmac, L., Popa-Fotea, N. M., Bataila, V., & Scafa-Udriste, A. (2022). Towards a Deep-Learning Approach for Prediction of Fractional Flow Reserve from Optical Coherence Tomography. *Applied Sciences (Switzerland)*, 12(14). <https://doi.org/10.3390/app12146964>. (impact factor: 2.70, Q3)
- Tache, I. A., Hatfaludi, C. A., Puiu, A., Itu, L. M., Popa-Fotea, N. M., Calmac, L., & Scafa-Udriste, A. (2023). Assessment of the functional severity of coronary lesions from optical coherence tomography based on ensembled learning. *BioMedical Engineering Online*, 22(1). <https://doi.org/10.1186/s12938-023-01192-x>. (impact factor: 3.90, Q3)
- Scafa-Udris̄te, A., Itu, L., Puiu, A., Stoian, A., Moldovan, H., & Popa-Fotea, N.-M. (2023). In-stent restenosis in acute coronary syndrome—a classic and a machine learning approach. *Frontiers in Cardiovascular Medicine*, 10. <https://doi.org/10.3389/fcvm.2023.1270986>. (impact factor: 3.60, Q2)

Un capitol de carte a fost publicat în calitate de coautor:

- Meister, F., Houle, H., Nita, C., Puiu, A., Itu, L. M., & Rapaka, S. (2020). Additional clinical applications. In *Artificial Intelligence for Computational Modeling of the Heart* (pp. 183–210). Elsevier. <https://doi.org/10.1016/B978-0-12-817594-1.00017-6>.

## 5.4 Discuții

În timp ce punctele forte și limitările individuale au fost discutate pentru fiecare caz de utilizare și abordare prezentată în cadrul acestei teze în capitolele corespunzătoare, această secțiune își propune să sublinieze activitatea de cercetare desfășurată, la un nivel holistic.

Această teză de doctorat exemplifică modul în care ar putea fi depășite mai multe bariere în utilizarea soluțiilor bazate pe inteligența artificială în rezolvarea diverselor nevoi clinice prin considerarea a unui set de paradigme precum învățarea autosupervizată, generarea de date sintetice, estimarea importanței caracteristicilor și cuantificarea incertitudinii modelului. Toate aceste abordări au fost explorate prin intermediul unui set de cazuri de utilizare clinică legate de diferite etape ale gestionării cancerului, inclusiv diagnosticarea, stadializarea și tratamentul. Cu toate acestea, activitatea actuală are mai multe limitări, descrise în cele ce urmează.

Pentru început, deși secțiunea introductivă oferă o listă mai cuprinzătoare a soluțiilor de ultimă oră pentru provocările identificate, unele dintre acestea fiind concurente într-o anumită măsură în abordarea anumitor subiecte, am dori să menționăm că scopul acestei teze de doctorat nu a fost de a efectua un studiu comparativ între acestea, ci mai degrabă de a selecta empiric cea mai potrivită opțiune pentru rezolvarea unei probleme specifice. Prin urmare, o limitare a lucrării actuale constă în lipsa unei comparații între mai multe soluții de ultimă oră aplicabile unui anumit obstacol. Deși aceasta reprezintă o direcție viitoare de dezvoltare a acestei cercetări, de obicei, în practică, specificul problemelor abordate ar oferi indicii privind adoptarea unei metodologii în detrimentul celorlalte. De exemplu, alegerea între paradigmele de învățare semi-supervizată și auto-supervizată ar depinde în principal de disponibilitatea și calitatea datelor slab etichetate, sau de scopul general al modelului (de exemplu, contrastarea unui comportament general cu un comportament specific dorit).

În al doilea rând, deși sunt foarte promițătoare, majoritatea rezultatelor generate în cadrul acestui program de doctorat necesită o validare suplimentară pe seturi de date de testare reprezentative, sau pe scenarii clinice reale. Cu toate acestea, din cauza indisponibilității unor astfel de seturi de date, acest pas reprezintă în prezent o direcție viitoare a acestei lucrări.

În cele din urmă, având în vedere cercetările extinse care sunt direcționate în prezent către obținerea unor algoritmi de inteligență artificială din ce în ce mai buni, o serie de tehnologii recente inovatoare nu au fost explorate în timpul programului de doctorat. De exemplu, îmbunătățirile remarcabile obținute recent în domeniul modelării limbajului natural ar putea fi valorificate pentru a îmbunătăți credibilitatea datelor noastre sintetice privind cancerul de prostată, sau pentru a depăși provocările de modelare datorate eterogenității genomice inerente, în timp ce derivatele lor în domeniul procesării imaginilor ar putea îmbunătăți în continuare performanța extrapolării imaginilor CT.

Cu toate acestea, credem cu tărie că rezultatele obținute pe parcursul acestui program de doctorat ar putea avea un impact pozitiv major asupra transformării sistemului medical, aducând dovezi semnificative că provocările și temerile majore legate de adoptarea soluțiilor AI în rutinele clinice ar putea fi abordate eficient în vederea îmbunătățirii îngrijirii pacienților și, în cele din urmă, a rezultatelor actului medical.





# Bibliografie

- [1] World Health Organization. Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. Tech. rep. 2020. URL: <https://www.who.int/data/gho/data/themes/topics/indicator-groups/indicator-group-details/GHO/gho-gho-global-health-estimates-life-tables> (visited on 06/18/2021).
- [2] Khanh Bao Tran et al. "The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019". In: *The Lancet* 400 (10352 Aug. 2022), pp. 563–591. ISSN: 01406736. DOI: 10.1016/S0140-6736(22)01438-6. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0140673622014386>.
- [3] Robert E. Schoen et al. "Colorectal-Cancer Incidence and Mortality with Screening Flexible Sigmoidoscopy". In: *New England Journal of Medicine* 366 (25 June 2012), pp. 2345–2357. ISSN: 0028-4793. DOI: 10.1056/nejmoa1114635.
- [4] Andrei Puiu et al. "Privacy-Preserving and Explainable AI for Cardiovascular Imaging". In: *Studies in Informatics and Control* 30 (2 June 2021), pp. 21–32. ISSN: 1841429X. DOI: 10.24846/v30i2y202102.
- [5] R. J.M. Bruls and R. M. Kwee. "Workload for radiologists during on-call hours: dramatic increase in the past 15 years". In: *Insights into Imaging* 11 (1 Dec. 2020). How imaging studies increased and therefore clinicians workload. ISSN: 18694101. DOI: 10.1186/s13244-020-00925-z.
- [6] Christine Dan Lantsman et al. "Trend in radiologist workload compared to number of admissions in the emergency department". In: *European Journal of Radiology* 149 (Apr. 2022), p. 110195. ISSN: 0720048X. DOI: 10.1016/j.ejrad.2022.110195.
- [7] Tommaso Mansi, Tiziano Passerini, and Dorin Comaniciu, eds. *Artificial intelligence for computational modeling of the heart*. 1st ed. San Diego: Academic press in an imprint of Elsevier, 2019. ISBN: 978-0-12-817594-1.
- [8] A. Fischer et al. "Deep Learning Based Automated Coronary Labeling For Structured Reporting Of Coronary CT Angiography In Accordance With SCCT Guidelines". en. In: *Journal of Cardiovascular Computed Tomography* 14.3 (July 2020), S21–S22. ISSN: 19345925. DOI: 10.1016/j.jcct.2020.06.019. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1934592520301921> (visited on 12/20/2021).
- [9] Moritz Kasel-Seibert et al. "Assessment of PI-RADS v2 for the Detection of Prostate Cancer". In: *European Journal of Radiology* 85 (4 Apr. 2016), pp. 726–731. ISSN: 0720048X. DOI: 10.1016/j.ejrad.2016.01.011.
- [10] Antonio C. Westphalen et al. "Variability of the Positive Predictive Value of PI-RADS for Prostate MRI across 26 Centers: Experience of the Society of Abdominal Radiology Prostate Cancer Disease-focused Panel". In: *Radiology* 296 (1 July 2020). Interuser variability in PI-RADS. pp. 76–84. ISSN: 0033-8419. DOI: 10.1148/radiol.2020190646.

- [11] David J. Winkel et al. "A Novel Deep Learning Based Computer-Aided Diagnosis System Improves the Accuracy and Efficiency of Radiologists in Reading Biparametric Magnetic Resonance Images of the Prostate: Results of a Multireader, Multicase Study". In: *Investigative Radiology* 56 (10 Oct. 2021), pp. 605–613. ISSN: 15360210. DOI: 10.1097/RLI.0000000000000780.
- [12] Anamaria Vizitiu et al. "Applying Deep Neural Networks over Homomorphic Encrypted Medical Data". In: *Computational and Mathematical Methods in Medicine 2020* (2020). ISSN: 17486718. DOI: 10.1155/2020/3910250.
- [13] Shih Cheng Huang et al. "Self-supervised learning for medical image classification: a systematic review and implementation guidelines". In: *npj Digital Medicine* 6 (1 Dec. 2023). ISSN: 23986352. DOI: 10.1038/s41746-023-00811-0.
- [14] Jakub Konečný et al. "Federated Learning: Strategies for Improving Communication Efficiency". In: (Oct. 2016). URL: <http://arxiv.org/abs/1610.05492>.
- [15] Aviad Kipnis and Eliphaz Hibshoosh. "Efficient Methods for Practical Fully Homomorphic Symmetric-key Encrypton, Randomization and Verification". In: *IACR Cryptol. ePrint Arch. 2012* (2012), p. 637. URL: <https://api.semanticscholar.org/CorpusID:14250402>.
- [16] Ehsan Hesamifard, Hassan Takabi, and Mehdi Ghasemi. "CryptoDL: Deep Neural Networks over Encrypted Data". In: (Nov. 2017).
- [17] Hervé Chabanne et al. Privacy-Preserving Classification on Deep Neural Network. *Cryptology ePrint Archive, Paper 2017/035*. <https://eprint.iacr.org/2017/035>. 2017. URL: <https://eprint.iacr.org/2017/035>.
- [18] Charu C. Aggarwal and Philip S. Yu, eds. *Privacy-Preserving Data Mining*. Vol. 34. Springer US, 2008. ISBN: 978-0-387-70991-8. DOI: 10.1007/978-0-387-70992-5.
- [19] Henry Surendra and S MohanH. "A Review Of Synthetic Data Generation Methods For Privacy Preserving Data Publishing". In: *International Journal of Scientific & Technology Research* 6 (2017), pp. 95–101. URL: <https://api.semanticscholar.org/CorpusID:67051890>.
- [20] Aldren Gonzales, Guruprabha Guruswamy, and Scott R. Smith. "Synthetic data in health care: A narrative review". In: *PLOS Digital Health* 2 (1 Jan. 2023), e0000082. DOI: 10.1371/journal.pdig.0000082.
- [21] Jacob Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019, pp. 4171–4186. URL: <https://github.com/tensorflow/tensor2tensor>.
- [22] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. "Globally and Locally Consistent Image Completion". In: *ACM Transactions on Graphics (Proc. of SIGGRAPH)* 36.4 (2017), p. 107.
- [23] Pengpeng Liu et al. Semantically Consistent Image Completion with Fine-grained Details. 2017. arXiv: 1711.09345 [cs.CV].
- [24] Sebastian Bach et al. "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation". In: *PLOS ONE* 10 (7 July 2015), e0130140. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0130140.
- [25] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier". In: (Feb. 2016).
- [26] Avanti Shrikumar et al. "Not Just a Black Box: Learning Important Features Through Propagating Activation Differences". In: (May 2016).
- [27] Scott Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: (May 2017). URL: <http://arxiv.org/abs/1705.07874>.
- [28] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. "Axiomatic Attribution for Deep Networks". In: (Mar. 2017).

- [29] Florin C. Ghesu et al. "Quantifying and leveraging predictive uncertainty for medical image assessment". In: *Medical Image Analysis* 68 (2021), p. 101855. ISSN: 1361-8415. DOI: <https://doi.org/10.1016/j.media.2020.101855>. URL: <https://www.sciencedirect.com/science/article/pii/S136184152030219X>.
- [30] Costin Florian Ciuşdel et al. "Normalizing Flows for Out-of-Distribution Detection: Application to Coronary Artery Segmentation". In: *Applied Sciences* 12.8 (2022). ISSN: 2076-3417. DOI: 10.3390/app12083839. URL: <https://www.mdpi.com/2076-3417/12/8/3839>.
- [31] Natália Alves et al. "Prediction Variability to Identify Reduced AI Performance in Cancer Diagnosis at MRI and CT". In: *Radiology* 308 (3 Sept. 2023), e230275. ISSN: 15271315. DOI: 10.1148/radiol.230275.
- [32] Joeran S. Bosma et al. "Semisupervised Learning with Report-guided Pseudo Labels for Deep Learning-based Prostate Cancer Detection Using Biparametric MRI". In: *Radiology: Artificial Intelligence* 5 (5 Sept. 2023). ISSN: 26386100. DOI: 10.1148/ryai.230031.
- [33] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18 (2 Feb. 2021), pp. 203–211. ISSN: 15487105. DOI: 10.1038/s41592-020-01008-z.
- [34] Kamal M Ali and Michael J Pazzani. *Error Reduction through Learning Multiple Descriptions*. 1996, pp. 173–202.
- [35] Janneke M T Hendriksen et al. "Clinical characteristics associated with diagnostic delay of pulmonary embolism in primary care: a retrospective observational study". In: *BMJ Open* 7 (3 Mar. 2017), e012789. ISSN: 2044-6055. DOI: 10.1136/bmjopen-2016-012789.
- [36] Andrei Puiu et al. "Generative Adversarial CT Volume Extrapolation for Robust Small-to-Large Field of View Registration". In: *Applied Sciences* 12.6 (2022). ISSN: 2076-3417. DOI: 10.3390/app12062944. URL: <https://www.mdpi.com/2076-3417/12/6/2944>.
- [37] Matthew A. Mauro et al. *Mauro: image-guided interventions: expert radiology series*. Third. Philadelphia: Elsevier, Inc, 2020. ISBN: 978-0-323-61204-3.
- [38] Kevin Cleary and Terry M. Peters. "Image-Guided Interventions: Technology Review and Clinical Applications". In: *Annual Review of Biomedical Engineering* 12.1 (2010). PMID: 20415592, pp. 119–142. DOI: 10.1146/annurev-bioeng-070909-105249. eprint: <https://doi.org/10.1146/annurev-bioeng-070909-105249>. URL: <https://doi.org/10.1146/annurev-bioeng-070909-105249>.
- [39] J. Modersitzki. *Numerical Methods for Image Registration*. Jan. 2004, pp. 27–44.
- [40] R. Liao et al. "A Review of Recent Advances in Registration Techniques Applied to Minimally Invasive Therapy". In: *IEEE Transactions on Multimedia* 15.5 (2013), pp. 983–1000. ISSN: 1941-0077. DOI: 10.1109/TMM.2013.2244869.
- [41] Barbara Zitová and Jan Flusser. "Image Registration Methods: A Survey". In: *Image and Vision Computing* 21 (Oct. 2003), pp. 977–1000. DOI: 10.1016/S0262-8856(03)00137-9.
- [42] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. "Mutual-information-based registration of medical images: a survey". In: *IEEE Transactions on Medical Imaging* 22.8 (2003), pp. 986–1004. ISSN: 1558-254X. DOI: 10.1109/TMI.2003.815867.
- [43] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].
- [44] Tero Karras et al. *Progressive Growing of GANs for Improved Quality, Stability, and Variation*. 2017. arXiv: 1710.10196 [cs.NE].
- [45] Tero Karras, Samuli Laine, and Timo Aila. *A Style-Based Generator Architecture for Generative Adversarial Networks*. 2018. arXiv: 1812.04948 [cs.NE].

- [46] Taesung Park et al. "Semantic Image Synthesis with Spatially-Adaptive Normalization". In: arXiv:1903.07291 [cs] (Nov. 2019). arXiv: 1903.07291. URL: <http://arxiv.org/abs/1903.07291> (visited on 12/20/2021).
- [47] Yue Zhang et al. Task Driven Generative Modeling for Unsupervised Domain Adaptation: Application to X-ray Image Segmentation. 2018. arXiv: 1806.07201 [cs.CV].
- [48] Chenyu You et al. "CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE)". In: IEEE Transactions on Medical Imaging (2019), pp. 1–1. ISSN: 1558-254X. DOI: 10.1109/tmi.2019.2922960. URL: <http://dx.doi.org/10.1109/TMI.2019.2922960>.
- [49] Q. Yang et al. "Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss". In: IEEE Transactions on Medical Imaging 37.6 (2018), pp. 1348–1357. ISSN: 1558-254X. DOI: 10.1109/TMI.2018.2827462.
- [50] A. Vizitiu et al. "Data-Driven Adversarial Learning for Sinogram-Based Iterative Low-Dose CT Image Reconstruction". In: 2019 23rd International Conference on System Theory, Control and Computing (ICSTCC). 2019, pp. 668–674. DOI: 10.1109/ICSTCC.2019.8885947.
- [51] Zhu Jun-Yan et al. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". In: Computer Vision (ICCV), 2017 IEEE International Conference on. 2017.
- [52] Armanious Karim et al. "MedGAN: Medical Image Translation using GANs". In: Computerized Medical Imaging and Graphics (2019), p. 101684. ISSN: 0895-6111. DOI: 10.1016/j.compmedimag.2019.101684. URL: <http://dx.doi.org/10.1016/j.compmedimag.2019.101684>.
- [53] Chao Yang et al. High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis. 2016. arXiv: 1611.09969 [cs.CV].
- [54] Mark Sabini and Gili Rusak. Painting Outside the Box: Image Outpainting with GANs. 2018. arXiv: 1808.08483 [cs.CV].
- [55] Yi Wang et al. "Wide-Context Semantic Image Extrapolation". In: June 2019.
- [56] Julius Surya Sumantri and In Kyu Park. 360 Panorama Synthesis from a Sparse Set of Images with Unknown Field of View. 2019. arXiv: 1904.03326 [cs.CV].
- [57] Donald Ervin Knuth. The art of computer programming. 3rd ed. Reading, Mass: Addison-Wesley, 1997, p. 232. ISBN: 978-0-201-89683-1 978-0-201-89684-8 978-0-201-89685-5.
- [58] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance Normalization: The Missing Ingredient for Fast Stylization. 2016. arXiv: 1607.08022 [cs.CV].
- [59] Phillip Isola et al. Image-to-Image Translation with Conditional Adversarial Networks. 2016. arXiv: 1611.07004 [cs.CV].
- [60] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. 2017. arXiv: 1701.07875 [stat.ML].
- [61] Ishaan Gulrajani et al. "Improved Training of Wasserstein GANs". In: CoRR abs/1704.00028 (2017). arXiv: 1704.00028. URL: <http://arxiv.org/abs/1704.00028>.
- [62] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. 2016. arXiv: 1603.08155 [cs.CV].
- [63] Fabian Isensee et al. "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge". In: ArXiv abs/1802.10508 (2017).
- [64] Florin C. Ghesu et al. "An Artificial Agent for Anatomical Landmark Detection in Medical Images". In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016. Ed. by Sebastien Ourselin et al. Cham: Springer International Publishing, 2016, pp. 229–237. ISBN: 978-3-319-46726-9.

- [65] Florin-Cristian Ghesu et al. "Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.1 (Jan. 2019), pp. 176–189. ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2017.2782687. URL: <https://ieeexplore.ieee.org/document/8187667/> (visited on 12/20/2021).
- [66] Yasmina Chenoune et al. "Rigid registration of Delayed-Enhancement and Cine Cardiac MR images using 3D Normalized Mutual Information". In: vol. 37. Oct. 2010, pp. 161–164.
- [67] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: (Oct. 2020).
- [68] M. I. Jordan and T. M. Mitchell. "Machine learning: Trends, perspectives, and prospects". In: *Science* 349 (6245 July 2015), pp. 255–260. ISSN: 0036-8075. DOI: 10.1126/science.aaa8415.
- [69] Jason Walonoski et al. "Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record". In: *Journal of the American Medical Informatics Association* 25 (3 Mar. 2018), pp. 230–238. ISSN: 1527974X. DOI: 10.1093/jamia/ocx079.
- [70] Jean-Francois Rajotte et al. "iScience Synthetic data as an enabler for machine learning applications in medicine". In: (). DOI: 10.1016/j.isci. URL: <https://doi.org/10.1016/j.isci..>
- [71] Bogdan A. Gheorghită et al. "Improving robustness of automatic cardiac function quantification from cine magnetic resonance imaging using synthetic image data". In: *Scientific Reports* 12 (1 Dec. 2022). ISSN: 20452322. DOI: 10.1038/s41598-022-06315-3.
- [72] Shu Hui Hsu et al. "Synthetic CT generation for MRI-guided adaptive radiotherapy in prostate cancer". In: *Frontiers in Oncology* 12 (Sept. 2022). ISSN: 2234943X. DOI: 10.3389/fonc.2022.969463.
- [73] Maram Mahmoud A. Monshi, Josiah Poon, and Vera Chung. "Deep learning in generating radiology reports: A survey". In: *Artificial Intelligence in Medicine* 106 (2020), p. 101878. ISSN: 0933-3657. DOI: <https://doi.org/10.1016/j.artmed.2020.101878>. URL: <https://www.sciencedirect.com/science/article/pii/S0933365719302635>.
- [74] Yuxiang Liao, Hantao Liu, and Irena Spasić. "Deep learning approaches to automatic radiology report generation: A systematic review". In: *Informatics in Medicine Unlocked* 39 (2023), p. 101273. ISSN: 2352-9148. DOI: <https://doi.org/10.1016/j.imu.2023.101273>. URL: <https://www.sciencedirect.com/science/article/pii/S235291482300117X>.
- [75] Edward Choi et al. "Generating Multi-label Discrete Patient Records using Generative Adversarial Networks". In: (Mar. 2017). URL: <http://arxiv.org/abs/1703.06490>.
- [76] Omid Sayadi, Mohammad B. Shamsollahi, and Gari D. Clifford. "Synthetic ECG generation and Bayesian filtering using a Gaussian wave-based dynamical model". In: *Physiological Measurement* 31 (10 2010), pp. 1309–1329. ISSN: 13616579. DOI: 10.1088/0967-3334/31/10/002.
- [77] Edmond Adib, Fatemeh Afghah, and John J. Prevost. "Synthetic ECG Signal Generation Using Generative Neural Networks". In: (Dec. 2021). URL: <http://arxiv.org/abs/2112.03268>.
- [78] Burak Yelmen et al. "Creating artificial human genomes using generative neural networks". In: *PLoS Genetics* 17 (2 Feb. 2021). ISSN: 15537404. DOI: 10.1371/JOURNAL.PGEN.1009303.
- [79] Sahel Shariati Samani et al. "Quantifying genomic privacy via inference attack with high-order SNV correlations". In: *Institute of Electrical and Electronics Engineers Inc.*, July 2015, pp. 32–40. ISBN: 9781479999330. DOI: 10.1109/SPW.2015.21.
- [80] Bristena Oprisanu, Georgi Ganev, and Emiliano De Cristofaro. "On Utility and Privacy in Synthetic Genomic Data". In: (Feb. 2021). URL: <http://arxiv.org/abs/2102.03314>.

- [81] Karim Armanious et al. "MedGAN: Medical Image Translation using GANs". In: (June 2018). DOI: 10.1016/j.compmedimag.2019.101684. URL: <http://arxiv.org/abs/1806.06397><http://dx.doi.org/10.1016/j.compmedimag.2019.101684>.
- [82] Jun-Yan Zhu et al. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". In: (Mar. 2017). URL: <http://arxiv.org/abs/1703.10593>.
- [83] Rebecca L. Siegel et al. "Cancer Statistics, 2021". In: *CA: A Cancer Journal for Clinicians* 71 (1 Jan. 2021), pp. 7–33. ISSN: 0007-9235. DOI: 10.3322/caac.21654.
- [84] Harold Evelyn Taitt. *Global Trends and Prostate Cancer: A Review of Incidence, Detection, and Mortality as Influenced by Race, Ethnicity, and Geographic Location*. Nov. 2018. DOI: 10.1177/1557988318798279.
- [85] Mary Beth B. Culp et al. "Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates". In: *European Urology* 77 (1 Jan. 2020), pp. 38–52. ISSN: 0302-2838. DOI: 10.1016/j.eururo.2019.08.005.
- [86] Stephen B. Edge and American Joint Committee on Cancer. *AJCC cancer staging manual*, p. 650. ISBN: 9780387884400.
- [87] James D Brierley, Mary K Gospodarowicz, and Christian Wittekind. *TNM classification of malignant tumours*. John Wiley & Sons, 2017. ISBN: 9781119263548.
- [88] Jeremy L. Warner, Mia A. Levy, and Michael N. Neuss. "Feasibility and accuracy of extracting cancer stage information from narrative electronic health record data". In: *Journal of Oncology Practice* 12 (2 Feb. 2016), e169–e179. ISSN: 1935469X. DOI: 10.1200/JOP.2015.004622.
- [89] Stacy Loeb et al. "Distribution of PSA velocity by total PSA levels: Data from the Baltimore Longitudinal study of aging". In: *Urology* 77 (1 Jan. 2011), pp. 143–147. ISSN: 00904295. DOI: 10.1016/j.urology.2010.04.068.
- [90] Thiago N. Valette et al. "Probability of extraprostatic disease according to the percentage of positive biopsy cores in clinically localized prostate cancer". In: *International Braz J Urol* 41 (3 2015), pp. 449–454. ISSN: 16776119. DOI: 10.1590/S1677-5538.IBJU.2014.0223.
- [91] Jure Murgic et al. "The role of the maximum involvement of biopsy core in predicting outcome for patients treated with dose-escalated radiation therapy for prostate cancer". In: *Radiation Oncology* 7 (1 Aug. 2012). ISSN: 1748717X. DOI: 10.1186/1748-717X-7-127.
- [92] John B. Eifler et al. "An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011". In: *BJU International* 111 (1 2013), pp. 22–29. ISSN: 1464410X. DOI: 10.1111/j.1464-410X.2012.11324.x.
- [93] Hirohiko Kamiyama et al. *Unusual False-Positive Mesenteric Lymph Nodes Detected by PET/CT in a Metastatic Survey of Lung Cancer*. 2016. DOI: 10.1159/000446579.
- [94] Mingxia Feng et al. "Retrospective analysis for the false positive diagnosis of PET-CT scan in lung cancer patients". In: *Medicine (United States)* 96 (42 Oct. 2017). ISSN: 15365964. DOI: 10.1097/MD.00000000000007415.
- [95] James L. Mohler et al. "Prostate cancer, version 2.2019". In: *JNCCN Journal of the National Comprehensive Cancer Network* 17 (5 2019), pp. 479–505. ISSN: 15401413. DOI: 10.6004/jnccn.2019.0023.
- [96] Aaron C. Spalding et al. "Percent Positive Biopsy Cores as a Prognostic Factor for Prostate Cancer Treated with External Beam Radiation". In: *Urology* 69.5 (2007), pp. 936–940. ISSN: 0090-4295. DOI: <https://doi.org/10.1016/j.urology.2007.01.066>. URL: <https://www.sciencedirect.com/science/article/pii/S009042950700132X>.
- [97] Ji Won Seo et al. "PI-RADS version 2: Detection of clinically significant cancer in patients with biopsy gleason score 6 prostate cancer". In: *American Journal of Roentgenology* 209 (1 July 2017), W1–W9. ISSN: 15463141. DOI: 10.2214/AJR.16.16981.

- [98] Kouji Izumi et al. "The Relationship Between Prostate-Specific Antigen and TNM Classification or Gleason Score in Prostate Cancer Patients With Low Prostate-Specific Antigen Levels". In: *The Prostate* 75 (Mar. 2015). DOI: 10.1002/pros.22985.
- [99] Daniel Jones et al. The diagnostic test accuracy of rectal examination for prostate cancer diagnosis in symptomatic patients: A systematic review. June 2018. DOI: 10.1186/s12875-018-0765-y.
- [100] Anthony V D et al. Preoperative PSA Velocity and the Risk of Death from Prostate Cancer after Radical Prostatectomy. 2004. URL: [www.nejm.org](http://www.nejm.org).
- [101] Stacy Loeb et al. "PSA Doubling Time Versus PSA Velocity to Predict High-Risk Prostate Cancer: Data from the Baltimore Longitudinal Study of Aging". In: *European Urology* 54 (5 Nov. 2008), pp. 1073–1080. ISSN: 03022838. DOI: 10.1016/j.eururo.2008.06.076.
- [102] David E. Neal et al. "Ten-year Mortality, Disease Progression, and Treatment-related Side Effects in Men with Localised Prostate Cancer from the ProtecT Randomised Controlled Trial According to Treatment Received". In: *European Urology* 77.3 (2020), pp. 320–330. ISSN: 0302-2838. DOI: <https://doi.org/10.1016/j.eururo.2019.10.030>. URL: <https://www.sciencedirect.com/science/article/pii/S0302283819308371>.
- [103] Jenny L. Donovan et al. "Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer". In: *New England Journal of Medicine* 375 (15 Oct. 2016), pp. 1425–1437. ISSN: 0028-4793. DOI: 10.1056/nejmoa1606221.
- [104] Anthony V. D'Amico et al. "Biochemical Outcome After Radical Prostatectomy, External Beam Radiation Therapy, or Interstitial Radiation Therapy for Clinically Localized Prostate Cancer". In: *JAMA* 280.11 (Sept. 1998), pp. 969–974. DOI: 10.1001/jama.280.11.969. eprint: <https://jamanetwork.com/journals/jama/articlepdf/187980/joc80111.pdf>. URL: <https://doi.org/10.1001/jama.280.11.969>.
- [105] Prasanna Sooriakumaran et al. "Comparative effectiveness of radical prostatectomy and radiotherapy in prostate cancer: Observational study of mortality outcomes". In: *BMJ (Online)* 348 (Feb. 2014). ISSN: 17561833. DOI: 10.1136/bmj.g1502.
- [106] John K. Gohagan et al. "Prostate Cancer Screening in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial of the National Cancer Institute". In: *The Journal of Urology* 152.5, Part 2 (1994), pp. 1905–1909. ISSN: 0022-5347. DOI: [https://doi.org/10.1016/S0022-5347\(17\)32412-6](https://doi.org/10.1016/S0022-5347(17)32412-6). URL: <https://www.sciencedirect.com/science/article/pii/S0022534717324126>.
- [107] Tim Hulsen. "An overview of publicly available patient-centered prostate cancer datasets". In: *Translational Andrology and Urology* 8 (S1 Mar. 2019), S64–S77. ISSN: 22234683. DOI: 10.21037/tau.2019.03.01.
- [108] Rupam Deori, Bijoyananda Das, and Mustafa Abdur Rahman. "A Study of Relationship of Prostate Volume, Prostate Specific Antigen and age in Benign Prostatic Hyperplasia". In: *International Journal of Contemporary Medical Research* 4 (2017). ISSN: 2454-7379. URL: [www.ijcmr.com](http://www.ijcmr.com).
- [109] Emily Alsentzer et al. Publicly Available Clinical BERT Embeddings. 2019, pp. 72–78. URL: <https://www.ncbi.nlm.nih.gov/pmc/>.
- [110] David J. Winkel et al. "Autonomous detection and classification of pi-rads lesions in an mri screening population incorporating multicenter-labeled deep learning and biparametric imaging: Proof of concept". In: *Diagnostics* 10 (11 Nov. 2020). ISSN: 20754418. DOI: 10.3390/diagnostics10110951.

- [111] Pritesh Mehta et al. "Computer-aided diagnosis of prostate cancer using multiparametric MRI and clinical features: A patient-level classification framework". In: *Medical Image Analysis* 73 (2021), p. 102153. ISSN: 1361-8415. DOI: <https://doi.org/10.1016/j.media.2021.102153>. URL: <https://www.sciencedirect.com/science/article/pii/S1361841521001997>.
- [112] Hashim U. Ahmed et al. "Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study". In: *The Lancet* 389 (10071 Feb. 2017), pp. 815–822. ISSN: 1474547X. DOI: 10.1016/S0140-6736(16)32401-1.
- [113] Baris Turkbey et al. "Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2". In: *European Urology* 76.3 (2019), pp. 340–351. ISSN: 0302-2838. DOI: <https://doi.org/10.1016/j.eururo.2019.02.033>. URL: <https://www.sciencedirect.com/science/article/pii/S0302283819301800>.
- [114] Shijun Wang et al. Computer aided-diagnosis of prostate cancer on multiparametric MRI: A technical review of current research. 2014. DOI: 10.1155/2014/789561.
- [115] Kimia Kohestani et al. "Performance and inter-observer variability of prostate MRI (PI-RADS version 2) outside high-volume centres". In: *Scandinavian Journal of Urology* 53 (5 Sept. 2019), pp. 304–311. ISSN: 21681813. DOI: 10.1080/21681805.2019.1675757.
- [116] Xin Yu et al. "False Positive Reduction Using Multiscale Contextual Features for Prostate Cancer Detection in Multi-Parametric MRI Scans". In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020, pp. 1355–1359. DOI: 10.1109/ISBI45749.2020.9098338.
- [117] Satoshi Washino et al. "Combination of prostate imaging reporting and data system (PI-RADS) score and prostate-specific antigen (PSA) density predicts biopsy outcome in prostate biopsy naïve patients". In: *BJU International* 119 (2 Feb. 2017), pp. 225–233. ISSN: 1464410X. DOI: 10.1111/bju.13465.
- [118] Tobias Nordström et al. "Prostate-specific antigen (PSA) density in the diagnostic algorithm of prostate cancer". In: *Prostate Cancer and Prostatic Diseases* 21.1 (2018), pp. 57–63. ISSN: 1476-5608. DOI: 10.1038/s41391-017-0024-7. URL: <https://doi.org/10.1038/s41391-017-0024-7>.
- [119] Yuan Fei Lu et al. "Optimizing prostate cancer accumulating model: Combined PI-RADS v2 with prostate specific antigen and its derivative data". In: *Cancer Imaging* 19 (1 May 2019). ISSN: 14707330. DOI: 10.1186/s40644-019-0208-6.
- [120] Dong Yang et al. "Automatic Liver Segmentation Using an Adversarial Image-to-Image Network". In: (July 2017). URL: <http://arxiv.org/abs/1707.08037>.
- [121] Andrei Puiu et al. "Prediction uncertainty estimates elucidate the limitation of current NSCLC subtype classification in representing mutational heterogeneity". In: *Scientific Reports* 14 (1 Mar. 2024), p. 6779. ISSN: 2045-2322. DOI: 10.1038/s41598-024-57057-3. URL: <https://www.nature.com/articles/s41598-024-57057-3>.
- [122] Ayal B Gussow, Eugene V Koonin, and Noam Auslander. "Identification of combinations of somatic mutations that predict cancer survival and immunotherapy benefit". en. In: *NAR Cancer* 3.2 (May 2021), zcab017.
- [123] Junyu Long et al. "A mutation-based gene set predicts survival benefit after immunotherapy across multiple cancers and reveals the immune response landscape". en. In: *Genome Med* 14.1 (Feb. 2022), p. 20.
- [124] Robert Clarke et al. "The properties of high-dimensional data spaces: implications for exploring gene and protein expression data". en. In: *Nat Rev Cancer* 8.1 (Jan. 2008), pp. 37–49.
- [125] Mohan Babu and Michael Snyder. "Multi-omics profiling for health". en. In: *Mol Cell Proteomics* 22.6 (Apr. 2023), p. 100561.



- [126] Kathryn A Phillips et al. "Availability and funding of clinical genomic sequencing globally". en. In: *BMJ Glob Health* 6.2 (Feb. 2021).
- [127] A Bayle et al. "ESMO study on the availability and accessibility of biomolecular technologies in oncology in Europe". en. In: *Ann Oncol* 34.10 (July 2023), pp. 934–945.
- [128] Gemma L D'Adamo, James T Widdop, and Edward M Giles. "The future is now? Clinical and translational aspects of "omics" technologies". en. In: *Immunol Cell Biol* 99.2 (Oct. 2020), pp. 168–176.
- [129] Valeria Relli et al. "Abandoning the notion of non-small cell lung cancer". en. In: *Trends Mol Med* 25.7 (May 2019), pp. 585–594.
- [130] Katherine A Hoadley et al. "Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer". en. In: *Cell* 173.2 (Apr. 2018), 291–304.e6.
- [131] Valeria Relli et al. "Distinct lung cancer subtypes associate to distinct drivers of tumor progression". en. In: *Oncotarget* 9.85 (Oct. 2018), pp. 35528–35540.
- [132] Shih-Hsin Hsiao et al. "Comparative survival analysis of platinum-based adjuvant chemotherapy for early-stage squamous cell carcinoma and adenocarcinoma of the lung". en. In: *Cancer Med* 11.10 (Mar. 2022), pp. 2067–2078.
- [133] Giorgio Scagliotti et al. "Treatment-by-histology interaction analyses in three phase III trials show superiority of pemetrexed in nonsquamous non-small cell lung cancer". en. In: *J Thorac Oncol* 6.1 (Jan. 2011), pp. 64–70.
- [134] Navneet Singh et al. "Therapy for stage IV non-small-cell lung cancer with driver alterations: ASCO living guideline". en. In: *J Clin Oncol* 40.28 (July 2022), pp. 3310–3322.
- [135] Navneet Singh et al. "Therapy for stage IV non-small-cell lung cancer without driver alterations: ASCO living guideline". en. In: *J Clin Oncol* 40.28 (July 2022), pp. 3323–3343.
- [136] Caicun Zhou et al. "Interim survival analysis of the randomized phase III GEMSTONE-302 trial: sugemalimab or placebo plus chemotherapy as first-line treatment for metastatic NSCLC". en. In: *Nat Cancer* 4.6 (June 2023), pp. 860–871.
- [137] Neil M Woody et al. "A histologic basis for the efficacy of SBRT to the lung". en. In: *J Thorac Oncol* 12.3 (Dec. 2016), pp. 510–519.
- [138] Nozomi Kita et al. "Comparison of recurrence patterns between adenocarcinoma and squamous cell carcinoma after stereotactic body radiotherapy for early-stage lung cancer". en. In: *Cancers (Basel)* 15.3 (Jan. 2023).
- [139] Jang Hyun Cho and Bharath Hariharan. "On the efficacy of knowledge distillation". In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019, pp. 4793–4801.
- [140] Ferdinandos Skoulidis and John V Heymach. "Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy". en. In: *Nat Rev Cancer* 19.9 (Aug. 2019), pp. 495–509.
- [141] Fan Zhang et al. "Co-occurring genomic alterations and immunotherapy efficacy in NSCLC". en. In: *NPJ Precis Oncol* 6.1 (Jan. 2022), p. 4.
- [142] Xiang Ge Luo, Jack Kuipers, and Niko Beerenwinkel. "Joint inference of exclusivity patterns and recurrent trajectories from tumor mutation trees". en. In: *Nat Commun* 14.1 (June 2023), p. 3676.
- [143] Catherine Labbé et al. "Prognostic and predictive effects of TP53 co-mutation in patients with EGFR-mutated non-small cell lung cancer (NSCLC)". en. In: *Lung Cancer* 111 (June 2017), pp. 23–29.
- [144] Zhong-Yi Dong et al. "Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma". en. In: *Clin Cancer Res* 23.12 (Dec. 2016), pp. 3012–3024.

- [145] William McLaren et al. "The Ensembl variant effect predictor". en. In: *Genome Biol* 17.1 (June 2016), p. 122.
- [146] Lei Zhang et al. "AutoGGN: A gene graph network AutoML tool for multi-omics research". In: *Artificial Intelligence in the Life Sciences* 1 (Dec. 2021), p. 100019.
- [147] Haitham A Elmarakeby et al. "Biologically informed deep neural network for prostate cancer discovery". en. In: *Nature* 598.7880 (Sept. 2021), pp. 348–352.
- [148] Marylyn D Ritchie et al. "Methods of integrating data to uncover genotype-phenotype interactions". en. In: *Nat Rev Genet* 16.2 (Jan. 2015), pp. 85–97.
- [149] Ethan Cerami et al. "The cBio cancer genomics portal: an open platform for exploring multi-dimensional cancer genomics data". en. In: *Cancer Discov* 2.5 (May 2012), pp. 401–404.
- [150] John G Tate et al. "COSMIC: the Catalogue Of Somatic Mutations In Cancer". en. In: *Nucleic Acids Res* 47.D1 (Jan. 2019), pp. D941–D947.
- [151] Pavlos S. Efrimidis and Paul G. Spirakis. "Weighted random sampling with a reservoir". In: *Information Processing Letters* 97.5 (2006), pp. 181–185.
- [152] Matthew H Bailey et al. "Comprehensive Characterization of Cancer Driver Genes and Mutations". en. In: *Cell* 173.2 (Apr. 2018), 371–385.e18.

# Rezumat

În contextul Inteligenței Artificiale (IA) care în prezent impactează transformările din domeniul sănătății, această teză de doctorat explorează provocările specifice cu care se confruntă dezvoltatorii în proiectarea tehnologiilor de vârf pentru îmbunătățirea practicilor clinice curente. Centrându-se pe diverse scenarii medicale legate de una dintre cele mai devastatoare boli a timpurilor noastre, cancerul, această lucrare investighează mai multe strategii pentru depășirea acestor provocări, având ca scop final eliminarea lacunelor actuale și deschiderea drumului pentru soluțiile bazate pe învățare profundă (din eng. „Deep Learning, DL”) spre adoptarea la scară largă. Prin urmare, scopul acestei teze de doctorat este de a debloca întregul potențial al IA în optimizarea rutinelor clinice și, în final, în îmbunătățirea îngrijirii pacienților.

Structurată în secțiuni specifice, această teză de doctorat evidențiază potențialul imens pe care inteligența artificială îl are în sprijinirea transformării domeniului medical, subliniază provocările tipice care în prezent încetinesc progresul său în abordarea diferitelor probleme clinice și identifică în final mai multe strategii pentru depășirea acestor obstacole.

Având în vedere că problemele legate de insuficiența și incompletitudinea datelor reprezintă o provocare comună în dezvoltarea soluțiilor bazate pe învățare automată, au fost create variante sintetice echivalente din punct de vedere statistic. În continuare, acestea pot fi utilizate pentru a dezvolta abordări inovatoare în rezolvarea diferitelor nevoi clinice, generând astfel un impact pozitiv enorm asupra sănătății pacienților. Pe de altă parte, atunci când complexitatea inerentă a domeniului medical împiedică acumularea seturilor suficiente de date cu adnotări pentru a alimenta algoritmi de învățare automată, paradigmele de antrenare precum învățarea auto-supervizată prezintă un mare potențial în facilitarea dezvoltării de astfel de aplicații, și, prin urmare, în construirea următoarei generații de sistem medical. Cu toate acestea, integrarea AI în rutinele clinice trebuie realizată cu rigurozitate, punând accent pe aspectele de siguranță. Exerțând un impact direct asupra bunăstării pacienților, asigurarea unei funcționalități sigure în practică este o cerință crucială a oricărui software medical. Prin urmare, această lucrare a luat în considerare progresele recente în inteligența artificială sigură pentru a îmbunătăți transparența și fiabilitatea modelelor de învățare profundă.

Pentru a demonstra fezabilitatea generală a acestor abordări, au fost urmate o serie de scenarii practice în jurul îngrijirii inovatoare a cancerului. În mod special, această teză demonstrează impactul IA asupra diferitelor etape de gestionare a bolii, incluzând diagnosticarea leziunilor prostatice clinic semnificative, stratificarea cancerului pulmonar fără celule mici și a celui de prostată pentru planificarea terapiei, și extrapolarea imaginilor medicale pentru ghidarea intervențiilor chirurgicale.

În final, toate rezultatele și perspectivele obținute din activitățile desfășurate în cadrul acestui program de doctorat sunt evidențiate pentru a oferi o imagine completă asupra tezei. În ansamblu, descoperirile noastre oferă dovezi convingătoare că preocupările și incertitudinile principale legate de integrarea soluțiilor AI în practicile clinice pot fi gestionate eficient, conducând la îmbunătățirea îngrijirii pacienților și, în cele din urmă, la prognoști îmbunătățite.