



Universitatea
Transilvania
din Braşov

ŞCOALA DOCTORALĂ INTERDISCIPLINARĂ

Facultatea: Inginerie Electrică și Știința Calculatoarelor

Ing. Costin Florian CIUȘDEL

**Învățare profundă pentru analiza imaginilor
de diagnosticare în bolile cardiovasculare**

**Deep Learning for diagnostic image analysis
in cardiovascular disease**

REZUMAT / ABSTRACT

Conducător științific

Prof.dr.ing. Lucian Mihai ITU

BRAȘOV, 2022



Universitatea
Transilvania
din Braşov

D-lui (D-nei)

Componenta Comisiei de doctorat

Numită prin ordinul Rectorului Universităţii Transilvania din Braşov

Nr. din

PREŞEDINTE: - Conf. univ. dr. ing. UNGUREANU Delia
Prodecan
Universitatea Transilvania din Braşov

CONDUCĂTOR ŞTIINŢIFIC: - Prof. univ. dr. ing. ITU Lucian Mihai
Universitatea Transilvania din Braşov

REFERENŢI: - Prof. univ. dr. ing. POPESCU Dan
Universitatea Politehnica din Bucureşti
- Prof. univ. dr. ing. SALOMIE Ioan
Universitatea Tehnică din Cluj Napoca
- Prof. univ. dr. ing. MOLDOVEANU Florin Dumitru
Universitatea Transilvania din Braşov

Data, ora şi locul susţinerii publice a tezei de doctorat:, ora, sala

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

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

Vă mulţumim.

Cuprins

1	Introducere	1
1.1	Învățare Profundă pentru Seturi Mari de Date Medicale: Prezentare Generală, Provo- cări și Viitor	1
1.2	Rolul Imagisticii Cardiace	3
1.3	Structura și Conținutul Tezei	4
2	Soluții bazate pe Deep Learning pentru Imagistica Cardiacă 2D cu Ultrasunete	7
2.1	Introducere	7
2.2	Metode de Pre-antrenare	8
2.2.1	Introducere	8
2.2.2	Metode și Rezultate	8
2.2.3	Concluzii	10
2.3	Modelarea Generativă Condiționată a Ecocardiografiilor	10
2.4	Detectarea Fazelor Cardiace pe Ecocardiografii folosind Rețele Neuronale Recurente	13
2.4.1	Introducere	13
2.4.2	Metode	14
2.4.3	Rezultate	15
2.5	Segmentarea Semantică și Detectarea Punctelor de Reper	15
2.5.1	Introducere	15
2.5.2	Metode	16
2.5.2.1	Arhitectura Encoder-Decoder de bază	16
2.5.2.2	Utilizarea Conexiunilor de Salt și a Învățării Multi-sarcină	17
2.5.2.3	Predicția Simultană pe Cadrele ED și ES	19
2.5.3	Rezultate	20
2.6	Incertitudinea Rețelelor Neurale Adânci	20
2.6.1	Introducere	20
2.6.2	Modele bazate pe Energie	21
2.6.3	Procese Gaussiene Rare	23
2.6.3.1	Formularea Matematică	23
2.6.3.2	Estimarea Deterministă a Incertitudinii în Învățarea Profundă	24
2.6.4	Concluzii	25
3	Detectarea Punctelor de Reper în Ecocardiografiile 3D utilizând Învățarea prin Recompensă	27
3.1	Introducere	27
3.2	Metode	27
3.2.1	Învățarea prin Recompensă	27
3.2.2	Agenți Multi-scală pentru Ecocardiografii 3D	28
3.3	Rezultate	30
3.4	Concluzii	31

4	Detectarea Segmentărilor Incorecte ale Lumenului în Angiografiile Coronariene prin Tomografie Computerizată	33
4.1	Introducere	33
4.2	Metode	34
4.2.1	Arhitecturi NF	34
4.2.2	Perturbații Sintetice ale Măștii	37
4.3	Rezultate și Discuție	37
4.3.1	Evaluarea pe Perturbații Sintetice ale Măștii	37
4.3.2	Evaluare pe baza Adnotărilor Experților	39
4.3.3	Eșantionarea folosind Modelele Antrenate	39
4.4	Concluzii	40
5	Detectarea Fazelor Cardiacă în Angiografiile Coronariene Invazive	43
5.1	Introducere	43
5.2	Metode	44
5.2.1	Procesul Offline de Antrenare	45
5.2.1.1	Preprocesarea Semnalelor ECG și a Cadrelor Angiografice	45
5.2.1.2	Arhitectura Modelului de Segmentare a Vaselor și Procesul de Antrenare	46
5.2.1.3	Arhitectura Modelului de Detecție a Fazelor Cardiacă și Procesul de Antrenare	46
5.2.2	Procesul Online al Aplicației	47
5.3	Rezultate	48
5.3.1	Detecția Vaselor	48
5.3.2	Detecția Fazelor Cardiacă	48
5.3.2.1	Performanța de Detectare a Fazelor Cardiacă pe Setul Principal de Date de Evaluare	48
5.4	Discuție	50
5.5	Concluzii	50
6	Concluzii Finale	51
6.1	Concluzii	51
6.2	Contribuții Originale	52
6.2.1	Ecocardiografie	52
6.2.2	Angiografie Coronariană cu Tomografie Computerizată	53
6.2.3	Angiografie Coronariană Invazivă	54
6.3	Diseminarea Rezultatelor Cercetării	54
	Bibliografie	57
	Rezumat	61

1. Introducere

1.1 Învățare Profundă pentru Seturi Mari de Date Medicale: Prezentare Generală, Provocări și Viitor

O definiție formală a învățării profunde (Deep Learning - DL) ar descrie-o ca fiind o ramură a învățării automate (Machine Learning - ML), care conține tehnici de instruire a rețelelor neuronale multi-stratificate pentru diverse sarcini de predicție, în care caracteristicile relevante sunt deduse automat din datele de instruire, spre deosebire de metodele clasice de ML care necesită adesea o proiectare manuală a caracteristicilor.

O descriere informală ar putea caracteriza DL ca un fenomen ce capătă amploare pe multiple planuri:

În **ingineria**:

- **hardware**: tot mai multe segmente hardware sunt dezvoltate pentru accelerarea inferenței rețelelor neuronale adânci: procesoare cu instrucțiuni vectorizate pentru operații în virgulă mobilă de mare viteză; procesoare grafice (GPU-uri) cu nuclee de procesare dedicate tensorilor; unități de procesare a tensorilor (TPU) implementate în infrastructuri de tip "cloud"; hardware încorporat de mică putere capabil să ruleze modele DNN de uz general.
- **software**: au fost dezvoltate mai multe platforme și ecosisteme de învățare profundă pentru a sprijini antrenarea și implementarea modelelor. Noile arhitecturi software utilizează motoare puternice de calcul la distanță prin intermediul paradigmei "Infrastructura-ca-Serviciu".

În **logica de afaceri** a aplicațiilor utilizate în crearea de conținut, imagistică medicală, tranzacții financiare și multe altele, modelele DNN (eng. Deep Neural Networks) au fost integrate cu succes pe scară largă. Aplicațiile comerciale dispun de caracteristici și capacități de ultimă generație care sunt propulsate de soluții DNN. În **literatură și cercetare**, subiectul DL este unul actual și atrage în mod constant atenție. Cele mai recente descoperiri în domeniul fundamentale, precum vederea artificială și prelucrarea limbajului natural, s-au propagat rapid la produsele disponibile pentru utilizatori, cum ar fi serviciile de traducere în timp real, mașinile care se conduc singure etc.

Implementarea algoritmilor bazați pe inteligență artificială (IA) a produs un spor de productivitate umană similar cu efectul adopției pe scară largă a soluțiilor software de acum câteva decenii. Numeroase grupuri de cercetare depun eforturi pentru a atinge perspectivele inteligenței generale artificiale, care poate fi apoi utilizată pentru a rezolva probleme noi și dificile în beneficiul umanității.

Deși rețelele neuronale nu sunt un concept nou (deoarece cercetarea lor a început în a doua jumătate a secolului trecut), abia în ultimul deceniu adoptarea și creșterea lor au început să atingă rata ridicată pe care o înregistrează în prezent practicienii. Perioada 2011-2012 a reprezentat un punct de cotitură pentru învățarea profundă. [1] și [2] au fost două publicații ale căror performanțe de vârf au stârnit interesul comunității. De exemplu, în competiția din 2012 "ImageNet Large Scale Visual Recognition Challenge", modelul "AlexNet" (introdus în [2]) a obținut o eroare top-5 de 15.3%, cu peste 10.8 puncte procentuale mai mică decât cea a celui de-al doilea clasat (o soluție non-DL).

Apariția soluțiilor bazate pe DL nu ar fi fost posibilă fără două ingrediente cheie: calculul rapid și cantitățile mari de date. Un model DNN are, de obicei, o structură paralelizabilă, permițând utilizarea eficientă a procesoarelor multi-nucleu. Din punct de vedere istoric, GPU-urile au avut un debit mult

mai mare de operații în virgulă mobilă (FLOPs) decât procesoarele obișnuite CPU, prin urmare, au fost candidați potriviți pentru calculele foarte paralelizabile. Cu toate acestea, GPU-urile au fost concepute în primul rând pentru sarcini grafice, iar pentru a le accesa abilitățile de procesare a fost necesară reformularea problemei inițiale ca fiind una care implică doar primitive grafice. Utilizarea API-urilor specifice graficii pentru GPU de uz general (GPGPU) era complicată și restrictivă. În 2007, nVIDIA a lansat prima versiune a CUDA ("Compute Unified Device Architecture", Arhitectură unificată pentru dispozitive de calcul), care a expus un API pentru programarea GPU de uz general. În prezent, CUDA a ajuns la cea de-a 11-a versiune și este foarte utilizată ca backend în toate platformele de DL. Existența unui API GPGPU ușor de utilizat, împreună cu creșterea exponențială a performanței de tip FLOPs, a împlinit cerința de calcul necesară pentru progresul DL.

Celălalt ingredient cheie, volumul de date, a fost, de asemenea, pus la dispoziție cu ușurință de alte inovații tehnologice: internetul și dispozitivele personale, cum ar fi camerele digitale, smartphone-urile etc. Utilizatorii creau conținut și îl partajau pe internet. De-a lungul timpului, au fost colectate și filtrate seturi mari de date, utilizate drept seturi de referință, un exemplu fiind setul de date din competiția ImageNet.

Un alt aspect care a favorizat adoptarea DL a fost capacitatea sa de generalizare. Nu mai era relevantă nevoia de a proiecta sau de a implementa caracteristici numerice specifice pentru fiecare sarcină și pereche de date individuale, deoarece un model DL își învață propriile caracteristici direct din date, dacă dispune de suficient timp de calcul și de date de instruire. În lucrarea AlexNet, autorii au afirmat că "Toate experimentele noastre sugerează că rezultatele pot fi îmbunătățite prin simpla așteptare a apariției unor GPU mai rapide și a unor seturi de date mai mari". Potențialul DL a fost descoperit ca fiind foarte mare, deoarece simpla valorificare a mai multor resurse a dus la creșteri constante ale performanței. În următorii ani, modelele mai mari au putut fi antrenate mai rapid pe GPU-uri mai avansate, pe seturi de antrenare din ce în ce mai mari, iar precizia obținută pe setul de testare s-a apropiat sau chiar a depășit performanța la nivel uman în multe sarcini de predicție.

De asemenea, au fost urmărite noi direcții de cercetare, cum ar fi modelarea generativă. Învățarea nesupravegheată a oferit modalități de a îmbunătăți și mai mult performanța modelării discriminative în fața unor seturi de date de mari dimensiuni, din care doar o mică parte era asociată cu adnotări. Chiar și fără adnotări, modelele de DL au putut învăța din datele brute, deducând caracteristici de bază dar și de nivel înalt despre structura datelor.

Imagistica medicală este un domeniu în care modelele DL au obținut performanțe remarcabile. Existența unor colecții mari de date care conțin numeroși pacienți și subgrupuri de populație a alimentat cercetarea și dezvoltarea modelelor neuronale adânci, care au îndeplinit numeroase sarcini în vederea sprijinirii diagnosticului pacienților, printre care [3]:

- **clasificare:** de exemplu, a subtipurilor de adenocarcinom pulmonar; a tumorilor pe imaginile ecografice ale sânelui; a nodulilor nesolizi; a cancerului de piele; a anomaliilor pulmonare; a leziunilor benigne și maligne; a limfomului evolutiv și a maselor reziduale; a cancerului de col uterin; a mutației EGFR; a progresului cancerului pulmonar; a tumorii cerebrale; etc.
- **localizare:** de exemplu, a tumorilor pulmonare; a punctelor de reper ale organelor; a leziunilor mamare; etc.
- **segmentare:** de exemplu, a nodulilor pulmonari; a tumorii hepatice; a plăcii aterosclerotice; a cancerului uterin; a tumorii în retinoblastom; a cancerului pancreatic; etc.
- **îmbunătățirea imaginii:** de exemplu, pentru eliminarea zgomotului; pentru translatarea domeniului (de la o modalitate de imagistică la alta); pentru accelerarea transformărilor de imagine (reconstrucție CT); etc.

Dezvoltarea unui model de DL pentru o sarcină de imagistică medicală necesită, de obicei, obținerea unui set de date etichetate, verificate de cititori experți și/sau medici practicieni. Procesul de adnotare manuală este de obicei lent și costisitor. Prin urmare, se urmărește proiectarea unor metode care să fie foarte eficiente din punct de vedere al datelor, adică obținerea unor modele de înaltă

performanță din seturi de antrenare de dimensiuni moderate. În producție, modelele sunt incluse, de obicei, în aplicații cu constrângeri privind timpul de execuție. Prin urmare, sunt preferate modelele eficiente din punct de vedere al parametrilor și al FLOP. Cu toate acestea, rezultatele de ultimă oră pe sarcini de predicție standardizate au arătat că modelele mari, antrenate pe seturi de date de mari dimensiuni, au cele mai bune performanțe. Acest lucru indică un compromis care trebuie adesea realizat între, pe de o parte, costul de adnotare a datelor de instruire și al hardware-ului de pe platforma țintă și, pe de altă parte, performanța de predicție, robustețea și generalizarea așteptate de la modelul DNN.

Din fericire, există metode pentru a spori performanța modelului. Pre-antrenarea este o tehnică puternică care poate valorifica cantități mari de date neetichetate pentru a produce proceduri de antrenare eficiente (din punct de vedere al datelor necesare) pentru sarcinile țintă finale. Robustețea modelelor poate fi legată de incertitudinea predicției și de detectarea valorilor aberante. Datele de intrare necunoscute pentru model (în ceea ce privește distribuția datelor observată în timpul instruirii) pot fi marcate, iar predicția pe astfel de date de intrare poate fi tratată separat în cadrul etapelor de postprocesare. Într-un proces de diagnosticare, datele imagistice care sunt prea zgomotoase sau care nu au calitatea necesară ar fi mai bine să fie eliminate decât să fie utilizate ca dovezi pentru stabilirea unui diagnostic.

Explicabilitatea este un alt aspect important în DL, în special în aplicațiile medicale. Faptul de a putea argumenta de ce un model a ajuns la o anumită predicție permite utilizatorului să aibă încredere în model în sine. Analiza cazurilor de eșec folosind modele explicabile permite, de asemenea, practicienilor DL să ia decizii mai bune în cadrul proceselor iterative de dezvoltare a modelului. De cealaltă parte a spectrului, tratarea modelelor de DL ca niște cutii negre sugerează că nu se cunosc elementele lor intrinseci, iar această abordare nu aduce aproape nicio valoare adăugată unei aplicații de DL în fața posibilelor artefacte și moduri de eșec ale modelului.

O idee futuristă în medicina modernă este conceptul de "geamă digital", în care sunt construite modele personalizate pentru pacienți, care se ajustează continuu pe baza istoricului de sănătate al fiecărui pacient. Deoarece datele unui singur pacient sunt insuficiente pentru a construi un model DL de sine stătător, un model meta-pacient poate fi în schimb condiționat pe baza parametrilor fiecărui pacient pentru a produce predicții personalizate. Având în vedere ritmul actual de dezvoltare, nu este nerealist să ne așteptăm ca acuratețea predicției algoritmilor bazați pe DL să atingă sau chiar să depășească radiologii sau ecografiștii experți în majoritatea sarcinilor.

Echipamentul de imagistică medicală va fi actualizat în mod continuu cu cei mai noi algoritmi bazați pe DL pentru analiza achizițiilor în timp real. În viitor ar putea fi atins auto-diagnosticul de precizie, în care scannerul de imagistică poate ajusta automat protocolul de scanare pe baza a ceea ce a fost observat până acum la un anumit pacient. În prezent, achiziția medicală este inspectată de un cititor expert, adesea la câteva ore sau zile după ce procedura de imagistică a fost finalizată și pacientul a părăsit incinta. În cazul în care constatările justifică investigații mai aprofundate, pacientul trebuie să se întoarcă și trebuie efectuată o altă procedură de imagistică. O soluție de analiză automată a imagisticii, alimentată de inteligența artificială, ar putea reduce drastic timpul dintre primul consult și diagnosticul final.

1.2 Rolul Imagisticii Cardiace

Structura acestei teze urmează fluxul clinic pentru un pacient suspectat de boli cardiovasculare. După o evaluare inițială, medicul poate solicita imagistică cardiacă pentru un diagnostic sigur [4]. Caracteristicile individuale ale pacientului și accesibilitatea locală influențează selectarea modalității de imagistică cardiacă. Ecocardiografia (inclusiv ecografia de stres) este capabilă să furnizeze informațiile clinice necesare, pentru un procent mare de cazuri; de asemenea, are avantajul de a evita expunerea la radiații. Pe de altă parte, coronarografia prin tomografie computerizată (CT) este din ce în ce mai mult utilizată pentru a detecta boala coronariană la pacienții cu rezultate ambigue la testul de efort și la cei cu un risc intermediar [4].

Evaluarea dispneei și investigarea bolii coronariene sunt două dintre cele mai frecvente scenarii clinice care pot necesita imagistică cardiacă [4]. Pentru dispnee, se folosesc adesea două modalități de imagistică:

- Radiografia toracică: în evaluarea preliminară a bolilor cardiovasculare. Se poate observa o creștere a dimensiunii inimii și indicatori de congestie pulmonară. Nu poate exclude în mod sigur etiologia cardiacă.
- Ecocardiografia: extrem de utilă pentru evaluarea dispneei, deoarece oferă informații structurale și funcționale care indică un anumit diagnostic.

Atunci când imagistica pentru dispnee nu este concludentă și au fost excluse afecțiunile respiratorii, se ia în considerare boala coronariană (CAD). Diagnosticul precis al bolii coronariene este crucial în planificarea tratamentului pentru pacienții care prezintă dureri toracice. Printre altele, pot fi efectuate următoarele investigații imagistice [4]:

- ECG de efort: test cu costuri reduse și riscuri minime, dar cu sensibilitate și specificitate scăzute pentru CAD.
- Ecocardiografia de stres: eficientă și neinvazivă. Informațiile furnizate sunt adesea foarte utile. Este fiabilă din punct de vedere al diagnosticului și în general accesibilă.
- Angiografia coronariană prin CT: vizualizează direct arterele coronare atât pentru CAD obstructivă, cât și pentru CAD non-obstructivă. Utilă pentru pacienții cu rezultate ambigue la testul de efort.
- Scorul de calciu coronarian bazat pe CT: poate prezice riscul de evenimente cardiovasculare viitoare la persoanele cu risc intermediar de CAD.

Pentru diagnosticarea cardiomiopatiei, a bolilor cardiace structurale și a bolilor cardiace congenitale, ecocardiografia transoesofagiană și RMN-ul cardiac pot fi utilizate atunci când nu există suficiente informații de diagnosticare [4].

Prin urmare, imagistica medicală cardiacă este un ingredient cheie în diagnosticarea și monitorizarea bolilor cardiovasculare. Angiografia coronariană CT și IRM-ul cardiac sunt relativ recente în utilizarea lor clinică în comparație cu ecocardiografia [4]. Având o importanță atât de mare în practica clinică, îmbunătățirea continuă pe baza algoritmilor DL a procedurilor de diagnosticare bazate pe imagistică medicală nu poate decât să sporească impactul acestora asupra îngrijirii pacienților, prin automatizarea măsurătorilor și creșterea preciziei acestora.

1.3 Structura și Conținutul Tezei

Această teză prezintă metode de învățare profundă și rezultatele unor aplicații în domeniul imagisticii medicale. Teza este structurată în 6 capitole. Acest **prim capitol** a prezentat o imagine de ansamblu asupra impactului și potențialului algoritmilor DL pentru imagistica medicală.

Al doilea capitol prezintă modalitatea de imagistică prin ultrasunete. Sunt urmărite sarcinile/etapele tipice de DL cuprinse în ciclul de dezvoltare a unei soluții bazate pe inteligență artificială: mai întâi, sunt prezentate metodele de pre-antrenare. Astfel de tehnici sunt utilizate pentru a învăța reprezentări generale din date neetichetate, pentru a asigura performanța sarcinilor supravegheate aferente din aval. O altă opțiune este generarea sintetică a mai multor date prin modelare generativă. Se studiază un cadru pentru generarea de ecocardiografii apicale condiționate de măștile de segmentare a camerelor cardiace. O măsură tipică de interes în evaluarea funcției cardiace este fracția de ejeție (FE). O soluție auto-EF necesită automatizarea a 2 etape cheie: detectarea fazei cardiace în achizițiile video ecografice, selectând în mod specific cadrele final-diastolice (ED) și final-sistolice (ES), și segmentarea semantică a camerelor inimii pentru obținerea conturilor corespunzătoare în vederea estimării volumelor camerelor. Pentru a realiza automatizarea cu succes a acestor sarcini, este

dezvoltat un model bazat pe rețele recurente (RNN) pentru detectarea fazei cardiace, în timp ce mai multe arhitecturi de model sunt investigate și îmbunătățite pentru segmentarea semantică și detectarea punctelor de reper. În cele din urmă, cel de-al doilea capitol investighează utilizarea metodelor de estimare a incertitudinii; acestea sunt adaptate pentru a rula modelele de segmentare, obținându-se astfel o hartă de incertitudine atașată măștilor prezise.

Capitolul al treilea prezintă detectarea reperelor în ecocardiografiile 3D ca o problemă de învățare prin recompensă. Șase repere trebuie detectate pentru ventriculul stâng în interiorul volumelor ecografice la anumite momente de timp ED/ES. În loc să existe o singură rețea convoluțională mare care să opereze pe întregul volum de intrare, o serie de agenți care observă regiuni locale întreprind acțiuni succesive pentru a minimiza distanța dintre poziția lor și locațiile țintă ale punctelor de reper. Această abordare conduce la o metodă rapidă ce necesită puține resurse pentru detectarea reperelor 3D. Este descris protocolul de antrenare, împreună cu cadrul matematic care îl guvernează. Sunt luate în considerare aspectele legate de implementare pentru a minimiza timpul de execuție.

În **capitolul al patrulea**, sunt explorate metode nesupervizate pentru detectarea intrărilor aberante (de ex. în afara distribuției de antrenare, eng. OoD). Sunt dezvoltate modele de fluxuri normalizatoare pentru a detecta segmentările incorecte ale lumenului 3D în angiografiile coronariene prin tomografie computerizată. Acestea sunt plasate după o etapă de segmentare, analizând intrările furnizate fie de către utilizator, fie de către un segmentor DNN, acționând astfel ca un model de audit. O nouă arhitectură NF este comparată cu una de referință pentru capacitățile lor de analiză semantică. Investigațiile arată că, prin respectarea anumitor constrângeri, se obține un model NF capabil de o mai bună interpretare semantică și, prin urmare, de o mai bună capacitate de detectare OoD. Procedurile de eșantionare arată, de asemenea, că modelul de bază se concentrează mai mult pe textură în loc de semantică, așa cum s-a raportat anterior în literatura de specialitate, în timp ce modelul propus reușește să producă eșantioane sintetice realiste.

Capitolul 5 prezintă o problemă de clasificare video, și anume detectarea fazei cardiace în angiografiile coronariene invazive. Modelul propus este prezentat împreună cu rezultate și analize detaliate pe seturi de test alese pentru evaluarea întregii aplicații. Similar cu detectarea fazei cardiace pe ecocardiografiile apicale, faza cardiacă în angiografiile coronariene invazive este dedusă exclusiv pe baza cadrelor de imagistică, evitând necesitatea de a procesa semnalele ECG zgomotoase (care pot fi, de altfel, absente pentru unele achiziții).

Capitolul final prezintă concluziile, evidențiază contribuțiile originale și diseminarea rezultatelor cercetării și indică posibile direcții de dezvoltare viitoare.

2. Soluții bazate pe Deep Learning pentru Imagistica Cardiacă 2D cu Ultrasunete

2.1 Introducere

Ecocardiografia este esențială în cardiologie, deoarece ajută la evaluarea funcției valvelor cardiace, cum ar fi stenoza sau insuficiența, puterea de contracție a mușchiului cardiac și hipertrofia sau dilatarea camerelor principale. De obicei, în timpul unui studiu ecocardiografic, se fac mai multe achiziții din diferite poziții pentru a capta suficiente informații pentru determinarea mărimilor de interes precum: dimensiunile și volumele camerelor, fracția de ejeție, îngroșarea peretelui, tensiunea longitudinală globală etc.

Fracția de ejeție (EF) este o măsură importantă în evaluarea funcției cardiace [8]. De obicei, se măsoară doar în ventriculul stâng (LV), principala cameră de pompare a inimii. Valorile reduse ale EF pot fi cauzate de cardiomiopatie, de leziuni ale mușchiului cardiac în urma unui atac de cord, de probleme ale valvelor cardiace sau de hipertensiune arterială pe termen lung. EF este definită ca fiind:

$$EF = \frac{EDV - ESV}{ESV} \quad (2.1)$$

unde EDV și ESV sunt volumele LV la două momente de timp distincte: sfârșitul diastolei (ED) și, respectiv, sfârșitul sistolei (ES). Pentru a estima cele 2 volume, se pot utiliza vizualizările apicale cu 2 camere și cu 4 camere.

Având segmentările LV pentru cele două vederi, metoda Simpson poate fi utilizată pentru a obține o estimare bi-planară a volumului ventricular (estimare care este aproape de valoarea fizică reală). Această metodă ia în considerare împărțirea LV (de-a lungul axei sale de la valva mitrală spre vârf) în elipse subțiri. Cele două diametre ale elipsei primesc dimensiunile diametrelor corespunzătoare ale LV în vederile A2C și A4C în locații corespunzătoare de-a lungul axei ventriculului. Un flux de lucru pentru estimarea EF ar cuprinde următoarele etape:

1. se obține ecocardiografii A2C și A4C (fiecare dintre ele conținând cel puțin un ciclu cardiac complet)
2. se detectează și se selectează cadrele ED și ES în interiorul celor două vederi
3. se segmentează LV în interiorul celor 4 cadre selectate
4. se aplică regula lui Simpson pentru a estima EDV și ESV și se calculează EF

Etapele 2 și 3 pot fi automatizate folosind tehnici de învățare profundă (DL), în timp ce etapa 4 poate fi implementată în cadrul etapelor de post-procesare ale unei soluții auto-EF. Alte metodologii [8] antrenează modele pentru a prezice direct EDV, ESV și EF ca sarcini de regresie, însă astfel de abordări sunt deficitare în ceea ce privește explicabilitatea.

2.2 Metode de Pre-antrenare¹

2.2.1 Introducere

Dezvoltarea unui model DNN pentru a efectua predicții pentru o sarcină specifică necesită un set de antrenare care să conțină perechi de eșantioane de intrare și ieșirile țintă corespunzătoare. În cazul în care se face o instruire de la zero și dimensiunea setului de antrenare este relativ mică, performanța modelului pe un set separat de test poate fi sub-optimală. Procedura de antrenare poate eșua cu totul, deoarece supra-adaptarea pe un set de antrenare mic este un fenomen bine cunoscut.

Pre-antrenarea este o tehnică care a ajutat scenariile cu volum redus de date să obțină o performanță mai mare la testare decât antrenarea de la zero. Această tehnică implică, de obicei, schimbarea ponderilor inițiale (adică înainte de a începe procedura de instruire finală) ale modelului în curs de dezvoltare cu cele preluate de la un alt model care a fost instruit pentru o altă sarcină similară. Privind modelul donator, nu este necesar să aibă o arhitectură identică cu cea a modelului final, deoarece acesta poate fi antrenat pe diferite tipuri de probleme de predicție, de exemplu, modelul donator a fost antrenat pe un set mare de date de clasificare, în timp ce modelul final urmează să fie antrenat pe un set mai mic de date de segmentare.

Obținerea de adnotări pentru seturile de date medicale implică, de obicei, ca experți să adnoteze manual fiecare caz în parte, ceea ce reprezintă un proces costisitor și de durată. În schimb, există colecții mari de achiziții medicale disponibile pentru utilizare, o oportunitate care poate fi ratată dacă se iau în considerare doar procentele mici de date care au adnotări de la experți. Prin urmare, o metodă cu un potențial ridicat este aceea de a învăța structura inerentă a datelor din cadrul acestor colecții de mari dimensiuni, fără a utiliza adnotări. Caracteristicile generale pot fi învățate mai întâi pe o sarcină pretext care nu are nicio importanță clinică. Aceste caracteristici generale sunt apoi modificate în mod specific pentru o sarcină finală cu relevanță clinică.

Motivația principală este că, atunci când se face o instruire de la zero, modelul trebuie să descopere și aceste caracteristici generale care descriu semantic datele de intrare. Având un model pre-antrenat ca punct de plecare inițial, procedura de antrenare abordează o problemă mult mai ușoară, deoarece modelul are deja o înțelegere decentă a conținutului semantic tipic. O problemă mai ușoară are nevoie de mai puține date de instruire pentru a obține performanțe ridicate de detectare. Prin urmare, este foarte important să se găsească metode bune de pre-antrenare care să ofere un avantaj considerabil pentru sarcinile finale.

2.2.2 Metode și Rezultate

Învățarea autosupravegheată este o serie de metode care vizează antrenarea modelelor DNN pentru a produce reprezentări învățate relevante. În general se utilizează augmentarea datelor și sarcini de tip pretext. Acestea formulează obiective de instruire simple, care nu au prea multă utilitate practică în sine, dar pentru ca modelul să le rezolve trebuie să învețe reprezentări utile, care pot fi ulterior reutilizate la instruirea pentru sarcina finală.

Această secțiune tratează sarcinile pretext euristice, în care o transformare T_i este aleasă aleatoriu dintr-un set predefinit $\{T_1, \dots, T_n\}$ și este aplicată imaginii de intrare. Modelul este antrenat pe o sarcină de clasificare, încercând să deducă care transformare anume a fost aplicată. În [9], au fost testate două metode de pre-antrenare autosupravegheată pe ecocardiografiile BMode apicale: oglindiri orizontale aleatorii și ordonarea cadrelor.

Pentru prima metodă, au fost luate în considerare cadrele de ecocardiografii individuale și au fost inversate de-a lungul axei verticale centrale cu o probabilitate de 50%. A fost formulată o sarcină de clasificare binară în care modelul trebuie să prezică "fals"/0 pe imaginile originale și "adevărat"/1 pe cele inversate. La o antrenare inițială, modelul a obținut o acuratețe de 95.8% pe setul de test (cu o acuratețe de 100% în timpul instruirii). În literatura de specialitate s-a raportat anterior că, atunci

¹Această secțiune descrie experimente realizate în [9], care reprezintă o lucrare publicată anterior de autor, în cadrul programului de cercetare doctorală.

când se utilizează o sarcină de pretext atât de simplă, modelul poate găsi mijloace de a trișa, adică în loc să învețe reprezentări utile ale conținutului semantic, modelul poate căuta anumite artefacte de imagine care se corelează bine cu etichetele țintă.

Pentru a testa acest aspect, a fost efectuat un alt experiment în care fiecare imagine de antrenare a fost injectată cu un artefact în colțul din dreapta jos, și anume un mic pătrat alb aflat întotdeauna în aceeași poziție pe fundalul negru. Inversarea unei imagini ar modifica, de asemenea, locația artefactului în partea de jos spre stânga. Prin urmare, a existat un indiciu ușor care poate fi exploatat în loc să se încerce interpretarea conținutului imaginii ecocardiografice. Comparativ cu cazul datelor neperturbate, modelul converge rapid la valoarea 0 a funcției criteriu pe ambele seturi (de antrenare și validare) și la o acuratețe maximă de clasificare. Atunci când a fost testat pe setul de test original (fără artefacte), performanța modelului a scăzut până la nivelul de alegere aleatorie: 51%, ceea ce sugerează că modelul nu a învățat nimic util.

Analiza salienței a fost efectuată pentru a confirma că modelul se concentrează asupra artefactelor. O imagine inversată a fost utilizată ca intrare de test. Artefactul a fost aplicat cu o intensitate graduală în colțul din stânga jos (locația pentru imaginile inversate), pornind de la o intensitate a pixelilor de 0 (adică niciun artefact) până la intensitatea maximă (adică culoarea albă, așa cum a fost folosită la antrenare). Inițial, fără aplicarea niciunui artefact, modelul emite o etichetă greșită aproape de 0. Când intensitatea artefactului este suficient de mare, ieșirea modelului crește brusc până la valoarea 1, eticheta corectă pentru intrările inversate. Se poate observa că, atunci când artefactul este suficient de vizibil, modelul se concentrează exclusiv pe acesta, fără a ține cont de orice conținut semantic prezent în interiorul ecocardiografiei.

Aceste experimente evidențiază un neajuns al unei astfel de sarcini simple pretext ce operează pe imagini întregi: datele de formare trebuie să fie pre-procesate cu atenție pentru a exclude orice artefacte posibile care dezvăluie transformarea aplicată în timpul antrenării și, astfel, anulează orice efect benefic al procedurii de preformare.

O altă metodă de preînvățare investigată în [9] a folosit sarcina pretext de a ordona corect o secvență de cadre sistolice. Au fost luate în considerare trei cadre: începutul (1), mijlocul (2) și sfârșitul sistolei (3), din același ciclu cardiac. Există șase permutări posibile în ordonarea celor 3 cadre: (1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2) și (3,2,1). În timpul antrenării, unui tuplu de cadre de intrare i s-a aplicat aleatoriu o permutare, iar eticheta țintă a fost o codificare one-hot a indicelui de permutare. Ca funcție criteriu a fost utilizată entropia categorială.

Au fost luate în considerare patru scenarii diferite, combinații între:

- inițializarea modelului: fie se antrenează de la zero, fie se utilizează modelul preformat din sarcina de detectare a inversării (fără artefacte) ca inițializare pentru encoder. Etajul final de ieșire a avut întotdeauna o inițializare aleatorie.
- encoder-ul responsabil cu calculul caracteristicilor imaginilor a fost fie înghețat, fie antrenabil. Etajul final de ieșire a fost întotdeauna antrenabil.

Tab. 2.1 afișează indicatorii de performanță pe setul de test. Ori de câte ori se pornește de la sarcina de detectare a inversării, acuratețea crește. La înghețarea encoder-ului, acuratețea este puțin mai bună cu învățarea prin transfer (+5%), dar rămâne masiv în urma scenariilor în care modelul este complet antrenabil. Acest lucru sugerează că prima sarcină a reușit să extragă reprezentări ușor mai informative decât caracteristicile aleatorii, dar nu pe deplin relevante pentru o sarcină mult mai dificilă, cum ar fi ordonarea cadrelor. Acest lucru este intuitiv, deoarece indicii relevanți pentru detectarea inversărilor (de exemplu, plasarea camerelor inimii) ar putea să nu fie la fel de relevanți ca indicii pentru ordonarea corectă a cadrelor (de exemplu, deschiderea valvei mitrale și dimensiunea LV).

Prin urmare, atunci când se utilizează sarcini euristice pretext, trebuie să se acorde o atenție deosebită asocierii sarcinii finale cu o sarcină pretext care depinde, cel puțin în parte, de aceleași regiuni și structuri ale imaginii.

Tabela 2.1: Performanțele de ordonare a cadrelor în diferite scenarii de inițializare și antrenare.

Mod de inițializare	Acuratețe
Inițializare aleatoare	92.38%
Inițializare cu parametri învățați din sarcina de clasificare binară (învățare prin transfer)	95.43%
Inițializare aleatoare și straturi înghețate	34.51%
Inițializarea din sarcina de detectare a inversării (învățare prin transfer) și straturi înghețate	39.59%

2.2.3 Concluzii

În cazul modalităților de imagistică medicală, cum ar fi cardiografiile cu ultrasunete, trebuie să se acorde o atenție deosebită atunci când se aleg operațiunile de augmentare a datelor. Spre deosebire de imaginile naturale, o achiziție obișnuită de ecocardiografie BMode are o formă constantă (de exemplu, conul de ultrasunet din mijlocul imaginii), iar conținutul de culoare nu este, de obicei, relevant. În loc de perturbația culorilor, ar trebui utilizate cele de contrast și luminozitate.

În cazul în care se utilizează decuparea aleatorie din imagine, aceasta ar trebui să fie restricționată pentru a capta structurile cardiace relevante și ar trebui să se asigure o suprapunere suficientă între patch-uri. În caz contrar, modelul poate fi forțat să producă codificări similare din regiuni disjuncte ale unui cadru, de exemplu, dacă se impune ca reprezentarea unui fragment din jurul camerei RA să fie similară cu cea din jurul LV, nu se pot învăța reprezentări utile. În schimb, se pot utiliza patch-uri incluzive pentru a impune ca modelul să poată localiza structurile locale ale camerelor cardiace în cadrul unor vizualizări mai mari ale întregii camere, iar patch-urile adiacente și suprapuse pot fi utilizate pentru a se asigura că modelul poate detecta dacă două patch-uri captează aceeași cameră cardiacă din aceeași achiziție.

2.3 Modelarea Generativă Condiționată a Ecocardiografiilor

Modelarea generativă implică învățarea distribuției datelor de instruire, fie în mod explicit (de exemplu, ca într-un auto-encoder variațional, un model de normalizare a fluxului etc.), fie în mod implicit (ca în rețelele generative adversative - GAN). În modelarea generativă condiționată, un nou eșantion este generat din distribuția învățată pe baza unor semnale de condiționare externe. Astfel de tehnici de modelare au o mare utilitate, deoarece permit eșantionarea din regiuni slab reprezentate ale spațiului de eșantionare original și pot ajuta la dezvoltarea modelelor supravegheate prin pre-antrenarea acestora pe date generate sintetic.

Această secțiune tratează modelarea generativă condiționată a ecocardiografiilor apicale BMode, și anume generarea cadrelor ecografice foto-realiste din măștile de segmentare a camerelor inimii. Acest obiectiv este opusul unei sarcini obișnuite de segmentare:

- în segmentare, modelul emite o mască corelată cu conținutul spațial al unui cadru ecocardiografic de intrare; conținutul semantic este dedus din datele de intrare, iar detaliile spațiale sunt utilizate pentru a poziționa și a regla masca prezisă.
- la generare, modelul produce un cadru ecocardiografic care este plauzibil față de masca de condiționare furnizată, adică amplasarea, dimensiunea și forma camerelor cardiace se potrivesc cu segmentările prescrise.

Această secțiune urmează metoda generativă condițională introdusă în [15]. Este dezvoltat un GAN în care generatorul utilizează un nou tip de strat de normalizare numit "Spatially Adaptive (De)Normalization" (SPADE). Măștile de condiționare sunt procesate de straturi convoluționale pentru a cal-

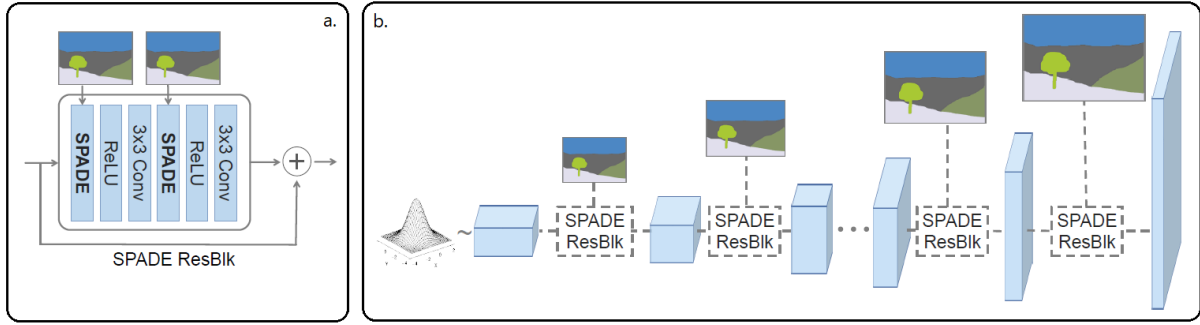


Figura 2.1: a: Bloc rezidual care utilizează modulele SPADE; b: arhitectura generatorului. În fiecare etapă, măștile de condiționare având aceeași scală spațială sunt utilizate pentru condiționarea hărții de caracteristici prin intermediul modulelor SPADE. Figura este o versiune modificată a [15].

cula tensori de demodulare spațială:

$$h_{n,c,y,x}^{i+1} = \gamma_{c,y,x}^i(m) \frac{h_{n,c,y,x}^i - \mu_c^i}{\sigma_c^i} + \beta_{c,y,x}^i(m) \quad (2.2)$$

unde $h_{n,c,y,x}^i$ și $h_{n,c,y,x}^{i+1}$ sunt hărți de caracteristici înainte de (la nivelul i) și, respectiv, după (la nivelul $i + 1$) normalizarea SPADE. Indicii n , c , y și x sunt de-a lungul axelor de grup, canal, înălțime și, respectiv, lățime. μ_c^i și σ_c^i sunt coeficienți de normalizare per canal, similari funcționării unui strat BatchNorm obișnuit fără transformarea afină, și anume: media și abaterea standard.

γ^i și β^i sunt tensori de demodulare care au aceleași dimensiuni spațiale și același număr de canale ca și h^i , prin urmare fiecare locație spațială a tensorului h^{i+1} este actualizată pe baza regiunii corespunzătoare a măștilor de condiționare. Tensorii de (de)modulație sunt calculați cu ajutorul unui mic modul DNN format din mai multe straturi convoluționale și de activare, fără a utiliza nicio normalizare între ele. În [15] s-a demonstrat empiric că acest modul de condiționare previne pierderea de informații semantice din măștile de condiționare, spre deosebire de alte tipuri de straturi de normalizare.

Arhitectura rețelei Generator este similară unui decodor de imagini (a se vedea Fig. 2.1). Zgomotul aleatoriu z poate fi utilizat ca intrare în rețea pentru a modula stilul și texturile (în timp ce conținutul semantic este dictat de măștile de condiționare). După fiecare etapă a generatorului, se utilizează operații de supraeșantionare pentru a crește succesiv rezoluția hărții de caracteristici. Fiecare etapă constă în blocuri reziduale SPADE. Măștile de condiționare sunt redimensionate pentru a corespunde rezoluțiilor fiecărei etape de generare și sunt utilizate pentru condiționare în interiorul blocurilor reziduale. Rezoluția de ieșire țintă a fost de 384x512 (în alb-negru). Au fost utilizate șase etape de supraeșantionare pentru a crește rezoluția de la o dimensiune inițială de 6x8.

Trei discriminatori multi-scală au fost utilizați pentru a realiza clasificarea (la nivel de regiune) între imaginile reale și cele sintetice la mai multe scări: scara originală, 0.5x și 0.25x. Intrarea a fost formată din concatenarea unei măști de condiționare cu o imagine ecocardiografică reală/sintetică. Primul discriminator la scară are un FoV receptiv $\sim 18\%$ din dimensiunea înălțimii și, prin urmare, se concentrează mai mult pe textură, în timp ce discriminatorul la cea mai mică scară are un FoV $\sim 73\%$ din dimensiunea înălțimii și, prin urmare, se concentrează mai mult pe forme/structuri și pe conținutul semantic.

Varianta Wasserstein GAN [16] a fost utilizată pentru instruire. Pornind de la distanța "Earth-Mover", aplicarea dualității Kantorovich-Rubinstein conduce la următoarea funcție criteriu adversativă pentru discriminator:

$$\max_w \mathbb{E}_{x \sim \mathbb{P}_r} [f_w(x)] - \mathbb{E}_{z \sim p(z)} [f_w(g_\theta(z))] \quad (2.3)$$

și pentru generator:

$$\max_\theta \mathbb{E}_{z \sim p(z)} [f_w(g_\theta(z))] \quad (2.4)$$

unde w și θ sunt parametrii discriminatorului (multi-scală) f_w și, respectiv, ai generatorului g_θ ; \mathbb{P}_r este distribuția eșantioanelor de date reale și $z \sim p(z)$ este distribuția variabilelor latente.

Discriminatorul nu are o funcție de activare la ieșire. Pentru a aplica dualitatea Kantorovich-Rubinstein, discriminatorul trebuie să fie o funcție continuă 1-Lipschitz [17]. Pentru a asigura această constrângere, în literatura de specialitate au fost propuse mai multe metode, cum ar fi limitarea ponderilor [16]; penalizarea gradientului [18]; normalizarea spectrală [19]; etc. În acest caz, normalizarea spectrală este utilizată în interiorul discriminatorului.

Pentru a obține o valoare scalară a funcției criteriu pentru toate etapele de discriminare, erorile sunt calculate și mediate pe toate patch-urile spațiale multi-scală. Componente suplimentare de antrenare, cum ar fi calibrarea caracteristicilor perceptuale, au fost utilizate [15] pentru a spori calitatea eșantioanelor sintetizate.

Normalizarea spectrală a fost utilizată și în arhitectura generatorului. Dimensiunea spațiului latent a fost $z \in \mathbb{R}^{256}$. Un strat inițial complet conectat a proiectat z pe o hartă de caracteristici cu rezoluție redusă de 6x8, utilizată ca intrare în prima etapă a generatorului. S-a utilizat optimizatorul Adam cu o regulă de actualizare în care rata de învățare a discriminatorului a fost de 4 ori mai mare decât cea a generatorului.

Într-un prim experiment, un model a fost antrenat pe un set de date care cuprinde vizualizări ecocardiografice A2C. Măștile de condiționare au avut 3 canale: LV, LA și fundal. Se poate observa că modelul reușește să producă imagini fotorealiste respectând măștile de segmentare, însă fiecare cadru ecocardiografic este modelat independent și, prin urmare, perechile de cadre ED/ES pot fi sintetizate cu aspect diferit (de exemplu, texturi, mușchi papilari LV etc.) între ED/ES, în funcție de cât de mult s-au modificat măștile de segmentare între cele două faze cardiace. De asemenea, deschiderea valvei mitrale nu poate fi condiționată, deoarece generatorul nu poate deduce în mod fiabil faza cardiacă din măștile de segmentare de condiționare. O soluție la această problemă este modelarea în comun a perechilor de cadre ED/ES, în loc de cadre disjuncte. Atât generatorul, cât și discriminatorii multi-scală trebuie să fie modificați:

- Generator: același z latent este utilizat pentru ambele cadre ED/ES. Fiecare bloc rezidual este aplicat de două ori, împărțind toți parametrii convoluționali, cu excepția straturilor din interiorul blocurilor de (de)normalizare SPADE; există straturi separate care calculează coeficienții de modulare ($\gamma_{ED}^i, \beta_{ED}^i$) și ($\gamma_{ES}^i, \beta_{ES}^i$). BatchNorm-ul clasic aplicat înainte de modularea SPADE a procesat atât cadrele ED, cât și ES în cadrul aceluiași grup.
- Discriminator: stratul final din fiecare discriminator multi-scală este eliminat și fiecare discriminator este aplicat de două ori (o dată pentru cadrele ED și o dată pentru cadrele ES); rezultatul lor este concatenat de-a lungul axei canalelor și introdus într-un etaj de ieșire complet convoluțional, care realizează clasificarea spațială în funcție de patch-uri ca mai înainte, cu excepția faptului că fiecare patch este calculat pe baza patch-urilor corespunzătoare atât din tensorii ED, cât și ES.

Discriminatorii multi-scală vor încerca să distingă dacă perechile de cadre sunt reale sau sintetice, impunând ca generatorul să producă cadre ED/ES care au un aspect coerent, adică imaginile au o poziționare similară între ED/ES, valva mitrală ar trebui să aibă poziții diferite în funcție de fazele cardiace corespunzătoare și texturile sunt similare. Două astfel de modele au fost antrenate pe seturi de date compuse din vederi A2C și respectiv A4C. Fig. 2.2 prezintă exemple de eșantionare din ambele modele. Se poate observa că cadrele sintetizate sunt foto-realiste și că structurile anatomice relevante sunt generate în mod consistent între cadrele ED și ES.

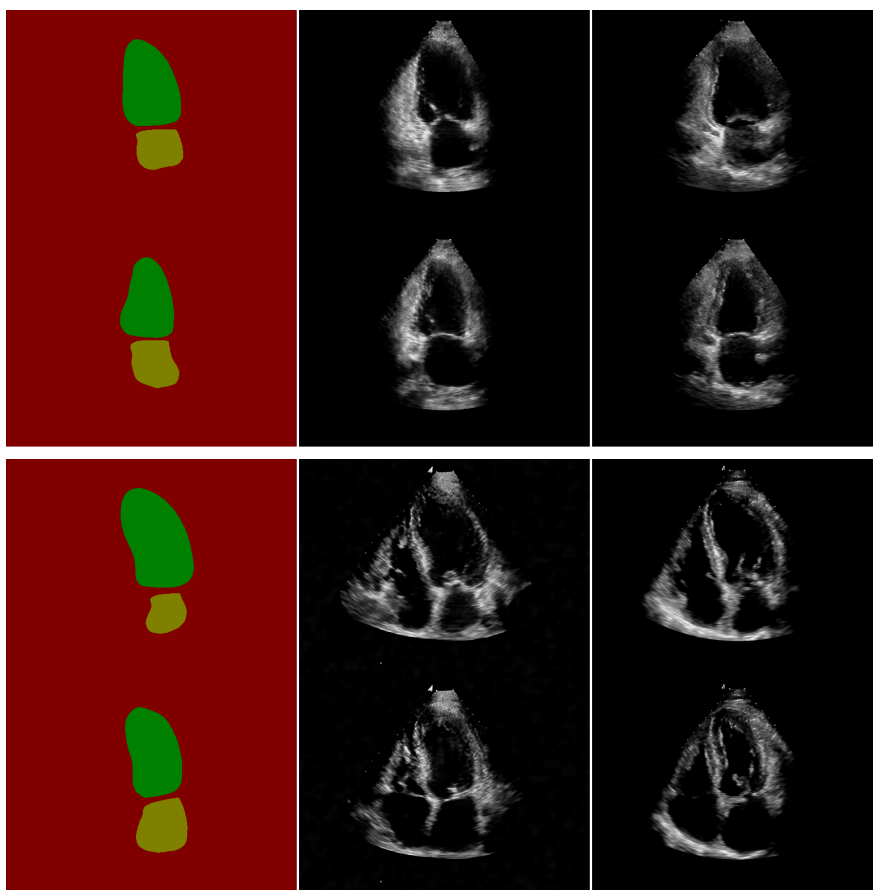


Figura 2.2: Exemple de perechi ED/ES A2C (grupul de sus) și A4C (grupul de jos) generate sintetic (coloana din mijloc) pe baza perechilor de măști de condiționare (coloana din stânga: verde este LV, galben este LA, roșu este fundalul). Coloana din dreapta prezintă eşantioane de date reale echivalente. În fiecare subgrup de figuri, rândurile de sus reprezintă ED, iar cele de jos ES.

2.4 Detectarea Fazelor Cardiace pe Ecocardiografii folosind Rețele Neuronale Recurente

2.4.1 Introducere

Detectarea fazei cardiace este un pas important în soluțiile auto-EF. În mod obișnuit, o înregistrare ecocardiografică conține mai multe cicluri cardiace. Din fiecare dintre acestea, sunt de interes cadrele ED și ES, care vor fi utilizate în modelele DNN din aval.

În comparație cu alte modalități de imagistică medicală, cum ar fi angiografiile coronariene, ecocardiografiile tind să aibă o variație mult mai mare a ratei cadrelor pe secundă (fps) ale achiziției video. De asemenea, mișcările inimii, cum ar fi valvele care se închid și se deschid rapid, pot face ca sub-eșantionarea la o rată de fps mică și constantă să nu ia în considerare modelele importante de mișcare (din cauza efectului de aliasing al eșantionării) și, prin urmare, să determine performanțe sub-optime ale DNN în detectarea fazei cardiace. Utilizarea ecocardiografiilor la valorile fps originale ridică probleme adiționale:

- rețeaua trebuie să țină cont de intervalele mari de fps. În cazul videoclipurilor cu rate fps ridicate, mișcarea structurilor cardiace se întinde pe mai multe cadre în comparație cu cele cu fps scăzut (unde există schimbări mult mai pronunțate de la un cadru la altul).
- în mod ideal, predicția pentru un cadru ar trebui să fie calculată pe baza conținutului semantic din toate celelalte cadre. Abordările cu ferestre glisante captează doar conținutul temporal

local. Deoarece numărul de cadre din cadrul unei achiziții este puternic variabil, rețeaua trebuie să se adapteze la ecocardiografii de lungime arbitrară.

2.4.2 Metode

Rețelele neuronale recurente reprezintă o soluție naturală la problemele de mai sus, deoarece pot opera pe secvențe de lungime arbitrară. Valoarea fps poate fi injectată ca o intrare suplimentară a modelului, permițându-i să facă față intervalelor largi de fps întâlnite de obicei în seturile de date. Se propune o arhitectură nouă, care permite determinarea cadrelor final-diastolice și final-sistolice numai din imagini medicale, fără a necesita semnale de intrare suplimentare (de exemplu, ECG etc.).

Semnalele țintă pot fi generate fie manual, de exemplu de către experții adnotatori, fie automat din semnale achiziționate simultan, cum ar fi ECG. În acest caz, trebuie să se acorde o atenție deosebită pentru a se asigura că semnalul este lipsit de zgomot, bine sincronizat și că algoritmi de detectare a ED / ES pe baza ECG sunt suficient de preciși.

În următoarele experimente, semnalul țintă a constat dintr-un semnal binar alături de un semnal 1D de ponderare a erorilor (generat prin scăderea unor Gaussiene centrate pe indicii cadrelor ED și ES dintr-un semnal constant; deviația standard a Gaussienelor a fost definită ca funcție a lungimii sistolei/diastolei măsurată în numărul de cadre). Pentru o gamă largă de valori ale pulsului, diastolele au o durată mai mare decât sistolele. Pentru a avea o contribuție echilibrată sistole-diastrale la valoarea agregată a funcției criteriu, durata sistolelor a fost ponderată invers proporțional cu raportul mediu sistole/diastrale observat pe întregul set de antrenare. Fig. 2.3 prezintă un exemplu de semnal țintă alături de semnalul de ponderare al erorilor per cadru. În timpul antrenării, entropia binară a fost calculată la fiecare index de cadru, iar suma ponderată a tuturor erorilor pe cadre individuale a produs eroarea totală a achiziției curente.

Arhitectura modelului combină atât sub-rețele convoluționale, cât și sub-rețele recurente și este capabilă, prin proiectare, să genereze predicția atât a posteriori (adică după ce întreaga achiziție devine disponibilă), cât și în timp real (de îndată ce sunt achiziționate noi cadre, modelul poate fi aplicat pentru a genera predicția pentru etapa de timp curentă și, eventual, pentru a actualiza predicțiile anterioare).

În prima etapă a rețelei, fiecare cadru al achiziției este rulat independent printr-o rețea neuronală convoluțională. Acest codificator operează pe imagini de intrare alb-negru de dimensiuni fixe și emite vectori unidimensionali de caracteristici de lungime 64. Codificatorul utilizează structuri simple de straturi care constau în convoluții 2D urmate de decimare prin maximizare, fiecare dintre acestea reducând dimensiunea de intrare la jumătate și creând caracteristici de nivel superior. Un strat final, complet conectat, responsabil pentru generarea ieșirii, primește ca intrare volume de caracteristici de la două scări spațiale diferite. Scopul este de a îmbunătăți performanța codificatorului în fața variabilității dimensiunilor ventriculare (datorată zoom-ului, tipului de vedere etc.). În etapa următoare,

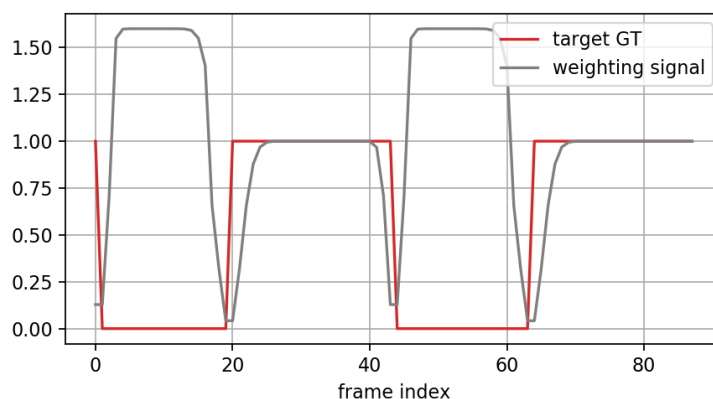


Figura 2.3: Exemplu de semnal țintă al unui eșantion utilizat la antrenarea rețelei recurente.

grupuri de vectori de caracteristici adiacente (de exemplu, tupluri de 3) sunt rulate printr-un CNN temporal. Acest CNN este format din două straturi: un strat convoluțional separabil și un strat convoluțional 1D. Scopul acestui CNN este de a produce caracteristici care să surprindă variația locală în timp între cadrele adiacente. Pentru fiecare grup de vectori de caracteristici spațiale se generează un vector de caracteristici temporale de lungime mai mare (de exemplu, 128).

Apoi, acești vectori de caracteristici temporale sunt trecuți printr-o rețea neuronală recurentă bidirecțională. Aceiași parametri RNN sunt utilizați pentru ambele treceri, înainte și înapoi. Ideea din spatele acestui RNN este de a realiza o mediere neliniară în timp a variației locale dintre cadrele imaginii, permițând astfel rețelei să proceseze secvențe de imagini înregistrate la diferite frecvențe de cadre. Starea internă a RNN este emisă la fiecare moment de timp, pentru ambele direcții, alimentând un strat complet conectat distribuit în timp. Acest strat dens acționează ca un detector de caracteristici temporale, de exemplu, pentru a detecta modificări ale volumului ventricular sau ale poziției valvei mitrale.

Ieșirea acestui detector este introdusă într-un alt RNN bidirecțional, acum cu parametri separați pentru cele două treceri înainte/înapoi. Ideea din spatele acestui RNN este de a învăța alternanța normală între sistole și diastole. Ultimul strat al rețelei este un clasificator binar distribuit, complet conectat, cu o activare sigmoidă. Acesta emite probabilitatea ca fiecare cadru component să facă parte dintr-o diastolă. Rețeaua poate fi antrenată pe de-a întregul. Pot fi utilizate diferite neliniarități de activare, de exemplu "SeLU" (pentru straturile convoluționale), sigmoid și tanh (pentru straturile recurente). Se pot utiliza tehnici de augmentare a datelor și straturi de mascare a activărilor ca mijloace de evitare a supra-ajustării. Elementul de bază al straturilor RNN a fost unitatea LSTM, care a fost inițializată la începutul antrenării pentru a atenua problema dispariției gradientului (care apare de obicei în cazul secvențelor lungi) prin utilizarea unei polarizări mari în interiorul porții de atenuare.

2.4.3 Rezultate

După o antrenare de câteva zeci de epoci pe ~1000 de cazuri, performanța a fost evaluată pe setul de test utilizat în secțiunea 2.5 (care a constatat din 50 de pacienți, fiecare având câte o achiziție pentru fiecare vedere apicală cardiacă: A2C, A3C și A4C). Un expert a adnotat câte un ciclu cardiac pentru fiecare achiziție. Tab. 2.2 prezintă performanța de detecție a modelelor recurente, pentru fiecare vedere și fază cardiacă. Au fost calculate distanțe ale indexurilor cadrelor între adnotările țintă și cadrele prezise, iar valorile au fost normalizate la 30 fps (pentru a permite ca măsurătorile pe achiziții cu FPS foarte diferit să fie direct comparabile).

Pentru orice tranziție țintă, dacă nu există exact o tranziție prezisă în interiorul unei ferestre temporale predefinite, atunci cazul respectiv este considerat fals negativ (FN). Se poate observa că modelul prezintă o performanță bună de detectare, deoarece în > 90% din cazuri cadrele de tranziție prezise se află la mai puțin de 2 (în distanța normalizată) cadre față de adnotare. Performanța este similară pentru toate vizualizările, iar valorile medii ale erorilor indică faptul că nu există niciun decalaj semnificativ de predicție, pentru niciuna dintre fazele cardiace.

2.5 Segmentarea Semantică și Detectarea Punctelor de Reper

2.5.1 Introducere

Conturarea camerelor inimii este o sarcină de bază în soluțiile auto-EF. Un contur poate fi aproximat printr-un șir de puncte 2D. Astfel de șiruri pot fi obținute prin interpolarea între reperele prezise, prin extragerea marginilor măștilor de segmentare prezise sau prin utilizarea unei euristici care combină ambele tipuri de ieșiri. În experimentele următoare, soluțiile DL investigate au trebuit să îndeplinească ambele obiective:

- **Segmentare:** se produc 3 hărți de segmentare pentru LV, LA și fundal.

Tabela 2.2: Performanța detectorului de fază cardiacă pe setul de test. TP și FN numără câte tranziții țintă au fost detectate sau, respectiv, ratate. Erorile de indexare a cadrelor sunt calculate pentru tranzițiile detectate, cu distanțe normalizate la o rată de referință de 30 fps. Ultimele două coloane măsoară procentul de tranziții detectate care sunt la ± 2 sau ± 1 cadre distanță de țintă.

Fază	Vedere	TP	FN	Recall [%]	diff. med. \pm std [cadre]	abs. \pm std [cadre]	diff. med. \pm std [cadre]	Acuratețe [%] (± 2 tol.)	Acuratețe [%] (± 1 tol.)
ED	A4C	50	0	100	0.363 \pm 0.410		0.158 \pm 0.524	100	96
	A3C	50	0	100	0.377 \pm 0.556		-0.129 \pm 0.659	98	92
	A2C	49	1	98	0.368 \pm 0.758		-0.117 \pm 0.834	97.96	95.92
ES	A4C	50	0	100	0.584 \pm 0.591		-0.099 \pm 0.825	100	84
	A3C	49	1	98	0.752 \pm 0.688		-0.173 \pm 1.004	93.88	83.67
	A2C	50	0	100	0.718 \pm 0.566		-0.547 \pm 0.733	100	82

- **Localizare repere:** se produc probabilități spațiale pentru 6 puncte de interes (LV și LA, fiecare având câte 2 anule și un vârf).

Rezoluția de intrare a fost setată la 320px lățime pe 256px înălțime, pe un singur canal (imagini alb-negru). Pornind de la o colecție de achiziții medicale adnotate, au fost utilizate tehnici de preprocesare pentru a construi seturi de date adecvate pentru dezvoltarea și evaluarea modelului DL. Conul ecografic a fost decupat, imaginea rezultată a fost convertită în alb-negru și redimensionată. Fiecare achiziție medicală a fost un buffer de pixeli 3D cu înălțimea și lățimea imaginii ca primele două axe și dimensiunea temporală ca ultimă axă. Din fiecare buffer, doar cadrele din jurul punctelor de timp ED și ES au fost extrase și adnotate de experți. Experimentele au fost efectuate utilizând seturi de date constând din cele 3 vederi apicale (A2C, A3C, A4C). Seturile de instruire și validare au avut peste 12000 de cadre din peste 1800 de dicomuri. Setul de test a fost format din 50 de pacienți, fiecare având câte un dicom pentru fiecare vedere. Adnotările au constat din 17 puncte în jurul pereților ventriculului și atrului stângi, unde primul și ultimul punct erau cele două anule, iar punctul din mijloc era vârful LV/LA. Hărțile de segmentare țintă au fost generate prin umplerea poligonului obținut printr-o interpolare spline a celor 17 puncte adnotate.

2.5.2 Metode

2.5.2.1 Arhitectura Encoder-Decoder de bază

Prima abordare este o arhitectură de bază în care sarcinile de segmentare și de detectare a reperelor au fost tratate separat. Arhitectura DNN este o cascadă simplă de straturi convoluționale care formează un encoder (în care rezoluția imaginii este redusă progresiv și sunt calculate caracteristici de nivel superior) și un decoder (în care rezoluția originală de intrare este recuperată prin operațiuni succesive de supra-eșantionare, pornind de la ieșirea encoder-ului). Cele două părți ale rețelei au câte 5 etape fiecare, constând în blocuri secvențiale de convoluție, normalizare pe grupuri, activare și decimare prin maximizare în encoder, și supraeșantionare biliniară, convoluție, normalizare pe grupuri și activare în decoder. Toate filtrele de convoluție au avut dimensiunea 3x3 cu pas 1x1. După fiecare etapă de codificare/decodificare, rezoluția este diminuată/crește cu un factor de 2x, în timp ce numărul de canale de caracteristici este dublat, respectiv înjumătățit.

Considerând cazul dezvoltării unor modele separate pentru fiecare obiectiv, activarea SoftMax poate fi utilizată ca funcție de activare a stratului de ieșire, pentru ambele modele. Pentru segmentare, numărul de canale de ieșire este egal cu numărul de elemente segmentate (în acest caz, 3: LV, LA și fundal). Operația Softmax se aplică pe canale, independent pentru fiecare pixel. Tensorul rezultat are aceeași dimensiune spațială ca și imaginea de intrare, în care fiecare pixel conține o distribuție discretă de probabilitate pentru cele 3 clase de segmentare. Funcția criteriu "DICE" a fost utilizată pentru antrenare:

$$DICE(P_i, T_i) = \frac{2 \sum_j P_{i,j} T_{i,j} + \epsilon}{\sum_j P_{i,j} + \sum_j T_{i,j} + \epsilon} \quad (2.5)$$

unde P_i și T_i sunt distribuțiile de probabilitate prezise și, respectiv, țintă (spațiale, nenormalizate) pentru obiectul modelat pe canalul i . ϵ este o constantă mică utilizată pentru stabilitate numerică. Sumele sunt calculate pentru toate locațiile spațiale j din interiorul canalului de ieșire/țintă i . Valoarea finală a funcției criteriu este suma tuturor coeficienților DICE pe toate canalele.

Pentru detectarea reperelor, numărul de canale de ieșire este egal cu numărul de repere. Activarea SoftMax se aplică spațial pe dimensiunile înălțime-lățime, independent pentru fiecare canal (reper). În mod intuitiv, tensorul rezultat este o distribuție spațială de probabilitate similară unei hărți de activare, în care locațiile plauzibile ale punctelor de reper au valori mari, în timp ce alte regiuni au valori scăzute ale probabilității. Funcția criteriu utilizată a fost "Adaloss" [6]. În această metodă, locațiile punctelor sunt aproximate prin distribuții gaussiene centrate pe locațiile adnotate, dar cu valori diferite ale deviației standard. La începutul antrenării, localizarea este (încă) o problemă nouă și dificilă pentru modelul DNN, prin urmare distribuția țintelor este relativ largă, adică precizia solicitată pentru localizarea reperelor este relativ scăzută. Pe măsură ce instruirea avansează, metoda monitorizează valoarea funcției criteriu pe o fereastră prestabilită a ultimelor T epoci. În cazul în care variația erorii scade, înseamnă că antrenarea converge pentru nivelul de dificultate actual și, prin urmare, problema poate fi făcută mai grea prin reducerea deviației standard a gaussianelor țintă, solicitând astfel modelului DNN o precizie de localizare mai mare. Statisticile valorii funcției criteriu sunt monitorizate în permanență, iar gaussianele țintă sunt actualizate în concordanță, pentru a evita un comportament oscilant sau divergent în procesul de antrenare.

Avantajul acestei arhitecturi de bază constă în simplitatea sa. Dezavantajul este că modelul poate produce hărți de segmentare grosiere, deoarece detaliile spațiale fine se pot pierde. Cascada de codificatoare face ca rezoluția hărții caracteristicilor să devină progresiv mai mică, prin urmare este posibil să se sacrifice detaliile spațiale fine, cum ar fi marginile ascuțite. În timp ce decodorul supraeșantionează succesiv hărțile de caracteristici pentru a recrea rezoluția originală, în fiecare etapă sunt transformate numai caracteristicile primite din etapa anterioară; prin urmare, dacă unele detalii spațiale se pierd în straturile precedente din cauza decimării, nu vor putea fi recuperate din pricina naturii secvențiale a acestei arhitecturi. În [7] se arată că reutilizarea în decodor a hărților de caracteristici mai dense (produse de straturile anterioare din cadrul encoder-ului) poate recupera detaliile spațiale și, astfel, poate produce segmentări cu o fidelitate mai bună.

2.5.2.2 Utilizarea Conexiunilor de Salt și a Învățării Multi-sarcină

Pe baza principiilor prezentate în [7], lucrarea [5] introduce "UNet", un DNN complet convoluțional care utilizează conexiuni de salt între fiecare etapă de codificare și decodificare care funcționează la aceeași rezoluție. La fiecare etapă de decodare, harta de caracteristici corespunzătoare a codificatorului este reutilizată prin concatenarea acesteia cu harta de caracteristici curentă a decodificatorului. Supraeșantionarea hărții caracteristicilor de nivel înalt păstrează câmpul vizual mare atașat fiecărui pixel spațial, în timp ce harta caracteristicilor generată de codificatorul corespunzător conține caracteristici de nivel inferior, care sunt mai bogate în detalii spațiale. Prin urmare, concatenarea acestor 2 tipuri de hărți de caracteristici produce tensori care conțin atât informații semantice (în raport cu conținutul global al imaginii), cât și informații despre textură/forme (în raport cu conținutul local al imaginii). Rezultatul final este o rețea capabilă să segmenteze detaliile fine într-un mod coerent (de

exemplu, fără ieșiri cu artefacte ale măștii cum ar fi găuri sau scurgeri).

Sarcina de segmentare a camerelor inimii și localizarea reperelor sunt legate, deoarece în acest exemplu reperele fac parte din conturul adnotat care generează masca țintă. Prin urmare, cele două sarcini au o structură comună, în sensul că caracteristicile învățate pentru o sarcină se pot generaliza și la cealaltă. În loc să se antreneze două instanțe de model diferite, calculele pot fi partajate pentru a obține un singur model care produce valori de ieșire pentru ambele sarcini.

Fig. 2.4 prezintă o arhitectură DNN îmbunătățită care combină ideile de mai sus. Componentele rețelei sunt:

- un codificator cu 6 etape (operații de decimare 2x)
- un strat de tip "bottleneck" care conține straturi conv2D 3x3 și 1x1
- 3 decodificatoare, fiecare cu 6 etape (operații de supraeșantionare biliniară 2x). Primul decoder are 3 canale de ieșire și este responsabil pentru segmentarea camerelor inimii. Ultimele 2 decodere au câte 3 canale de ieșire și sunt responsabile pentru detectarea celor 3 reperi pentru camerele ventricul-stâng și, respectiv, atriu-stâng.

Ieșirea unui etaj de codificare este utilizată de 4 ori: pentru etajul de codificare următor și pentru fiecare decodor ca o conexiune de salt. Se utilizează același criteriu ca în secțiunea anterioară. Primul decoder are o funcție criteriu DICE, în timp ce ultimele 2 au (fiecare) o funcție Adaloss. Valoarea finală a erorii este suma ponderată a celor 3 componente. Datorită conexiunilor de salt, rețeaua poate fi mai adâncă fără a sacrifica fidelitatea spațială a ieșirilor, iar câmpul vizual efectiv (al fiecărui pixel de ieșire) poate crește. Deoarece fiecare pixel de ieșire este calculat în funcție de regiunea sa corespunzătoare din imagine, în mod ideal mărimea acestei regiuni ar trebui să acopere întreaga imagine pentru orice locație din harta de segmentare.

Configurația de învățare multi-sarcină are mai multe avantaje:

- nu este necesar să se deruleze mai multe proceduri de antrenare (deoarece nu există câte o instanță de model pentru fiecare sarcină).
- cerințele privind timpul de execuție a inferenței sunt reduse datorită calculului partajat.
- supervizarea de la o sarcină poate crește performanța pentru celelalte sarcini.
- capetele de decodificare pot produce ieșiri care sunt mai corelate spațial decât în cazul predicțiilor provenite din modele independente. Deoarece capetele împart același encoder, se obține un model cu o probabilitate mai mare de a avea o corelație spațială superioară în ieșirile sale.

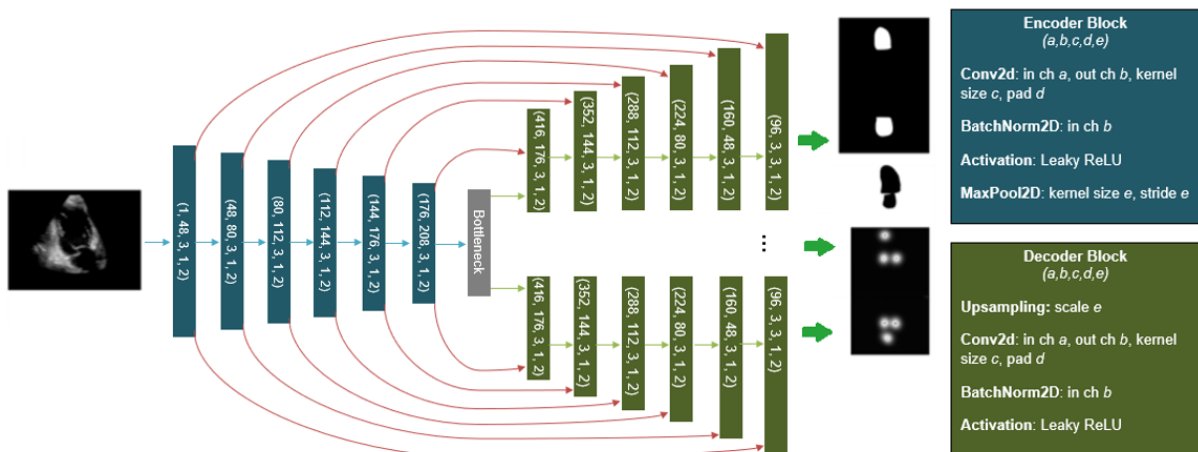


Figura 2.4: Arhitectură DNN îmbunătățită care utilizează conexiuni de salt și învățare multi-task.

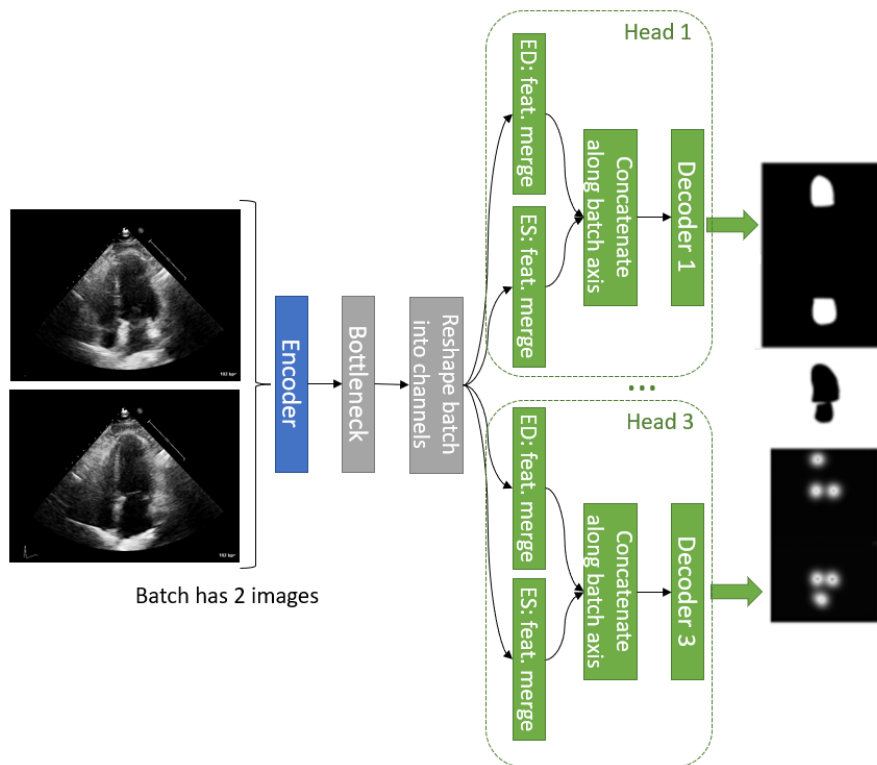


Figura 2.5: Arhitectură DNN care leagă în mod explicit procesarea cadrelor ED și ES. Straturile de fuziune sunt utilizate pentru a construi reprezentări pentru cadrele ED și, respectiv, ES, utilizând hărți de caracteristici din ambele cadre.

2.5.2.3 Predicția Simultană pe Cadrele ED și ES

După cum s-a menționat într-o secțiune anterioară, pentru a obține o estimare a fracției de ejeție (EF), sunt necesare predicții atât pentru cadrele ED, cât și pentru cele ES din aceeași achiziție medicală. Pentru a obține estimări robuste ale EF, sunt necesare estimări robuste ale volumelor individuale. Mai exact, nu ar trebui să existe nicio diferență între modul în care sunt prezise contururile ED și ES. În timpul achiziției video, inima pulsează iar imaginile capturate arată mișcarea anatomică aferentă. Însă, din cauza poziționării traductorului pe pacient, unele structuri cardiace pot fi complet vizibile în timpul unei faze cardiace, dar pot lipsi în cealaltă. Pentru o conturare robustă și coerentă, modelul trebuie să segmenteze în mod similar cadrele ED și ES în această regiune incertă din punct de vedere vizual. De exemplu, ar putea fi posibil să se deducă unele informații lipsă din cadrul ES și să se aplice la segmentarea ED, pentru a maximiza coerența conturului între cele două cadre.

Arhitecturile anterioare nu legau în mod explicit procesarea celor 2 cadre ED/ES, ci doar implicit, prin intermediul straturilor de normalizare a grupurilor, deoarece ED și ES puteau fi puse împreună în același grup. Prin urmare, se propune o nouă arhitectură, construită pe baza celei din sub-secțiunea anterioară. Fig. 2.5 prezintă arhitectura generală. Același codificator este aplicat o singură dată pe un grup format din 2 cadre de intrare. După un etaj de tip "bottleneck", hărțile de caracteristici sunt remodelate prin împărțirea axei de grup în două și dublarea axei canalelor. Se aplică blocuri independente de "fuziune" conv2D 1x1 pentru a obține 2 reprezentări pentru cadrul ED și, respectiv, pentru cadrul ES, pornind de la caracteristicile concatenate ale ambelor cadre. Fiecare cap de decodare are propriile seturi de straturi de "fuziune", care pot acționa la nivelul mai multor scări spațiale.

Procedura de antrenare este aceeași ca în sub-secțiunile anterioare. Modelul poate fi antrenat complet într-o configurație multi-task. Această arhitectură revizuită permite condiționarea reciprocă între valorile ED și ES de ieșire. Rezultatul efectiv al modificărilor arhitecturale propuse este un model capabil de performanțe superioare în cele 2 sarcini principale, care poate oferi un contur robust și consistent, permițând astfel o estimare mai precisă a valorilor fracției de ejeție.

2.5.3 Rezultate

Tab. 2.4 și 2.3 prezintă comparații de performanță pentru cele 3 arhitecturi descrise pe setul de test, pentru segmentare și, respectiv, detectarea reperelor. În comparație cu cel de bază, ambele modele actualizate prezintă performanțe superioare pentru ambele sarcini, pe toate vizualizările, camerele și fazele cardiace.

Chiar dacă modelele actualizate au performanțe similare în ceea ce privește erorile DICE și de reper L_2 , dacă se consideră corelația valorilor EF pe un singur plan și pe două planuri, condiționarea reciprocă a ultimului model produce o îmbunătățire a coeficientului de corelație Pearson, după cum se observă în Tab. 2.5.

În concluzie, sarcina de conturare a camerelor inimii poate fi automatizată cu succes în cadrul unei soluții de estimare a fracției de ejeecție, obținând performanțe bune în analiza cu adnotări de la experți.

2.6 Incertitudinea Rețelelor Neurale Adânci

2.6.1 Introducere

Rețelele neuronale adânci obișnuite au tendința de a fi prea încrezătoare în predicțiile lor, chiar și atunci când acestea sunt complet greșite. Predicția lor este o estimare într-un singur punct, adică doar un singur tensor este furnizat ca ieșire pentru fiecare eșantion de intrare. Deși noțiunea de incertitudine afectează atât gama de valori de ieșire reale posibile, cât și sarcinile din aval, aceasta este complet ignorată de algoritmi DL clasici. Din fericire, extensii simple permit practicienilor să deducă mai multe informații din modele DL deja antrenate.

Tabela 2.3: Compararea performanțelor de detectare a reperelor. Fiecare celulă arată distanțele euclidiene dintre locațiile prezise și cele reale, mediate pentru cele 3 puncte de reper. Erorile sunt măsurate în pixeli (în raport cu rezoluția de intrare de 320x256). (SC = conexiuni de salt, MT = antrenare multi-task, ED/ES = condiționare între cadre)

Vedere	A2C				A3C		A4C				Eroare medie reper
Cameră	LA		LV		LV		LA		LV		
Fază	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	
Model de bază	5.6	5.5	4.9	5.0	5.9	5.6	5.9	5.7	4.7	4.9	5.4
SC & MT	5.2	5.2	4.2	4.4	5.4	5.3	5.4	5.0	4.2	4.4	4.9
SC, MT & ED/ES	5.3	5.1	4.1	4.3	5.4	5.2	5.5	5.0	4.1	4.3	4.8

Tabela 2.4: Compararea performanțelor privind segmentarea camerelor inimii.

Vedere	A2C				A3C		A4C				Global DICE
Cameră	LA		LV		LV		LA		LV		
Fază	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	
Model de bază	90.18 ± 5.2	91.66 ± 4.0	91.72 ± 2.7	90.16 ± 4.1	91.61 ± 3.9	90.09 ± 6.1	88.60 ± 6.7	90.23 ± 5.6	92.02 ± 2.6	89.35 ± 4.3	90.56
SC & MT	90.56 ± 4.6	92.56 ± 3.3	92.66 ± 2.5	91.10 ± 2.8	92.33 ± 3.2	90.54 ± 3.6	89.89 ± 6.4	92.20 ± 5.4	93.5 ± 2.4	90.39 ± 4.0	91.57
SC, MT & ED/ES	90.69 ± 5.0	92.85 ± 2.9	92.54 ± 2.1	90.88 ± 2.9	92.48 ± 3.2	90.51 ± 4.3	90.19 ± 6.5	92.08 ± 5.3	93.66 ± 2.0	90.96 ± 4.1	91.68

Tabela 2.5: Comparație de corelație EF între arhitecturile actualizate.

Arhitectură	plan-unic EF		bi-plan EF
	A2C	A4C	
SC & MT	0.82	0.83	0.84
SC, MT & ED/ES	0.86	0.86	0.88

Incertitudinea în ieșirile prezise depinde de incertitudinea din intrări și de model în sine (de arhitectura și de ponderile învățate). În mod ideal, incertitudinea din ieșiri ar trebui să fie ridicată numai din anumite motive:

- eșantionul de intrare este din afara domeniului (out-of-domain (OoD)), adică se află într-o regiune din spațiul de intrare care nu este acoperită de niciun exemplu de antrenare.
- eșantionul de intrare este plasat în apropierea graniței de separare a claselor învățate.
- în special în cazul intrărilor de dimensiuni mari (pentru care raritatea este și mai pronunțată, având în vedere natura finită a datelor de antrenare [13]), eșantionul de intrare poate fi departe de exemplele de antrenare vecine, chiar dacă este plasat "în interiorul" distribuției datelor.

La un nivel de bază, multe sarcini supervizate de DL pot fi împărțite în două categorii principale:

- **Regresie:** incertitudinea de ieșire este de obicei aproximată folosind o distribuție gaussiană. Modul acesteia reprezintă valoarea cea mai probabilă, în timp ce abaterea standard este o măsură directă a incertitudinii.
- **Clasificare:** entropia distribuției de probabilitate rezultată este un indicator de incertitudine. O entropie ridicată este asociată cu o distribuție uniformă asupra claselor posibile, în timp ce o entropie scăzută apare atunci când cea mai mare parte a masei de probabilitate este plasată doar într-o singură clasă.

Completarea predicției cu informații privind incertitudinea este valoroasă pentru orice flux de lucru DL, în special în domeniul medical. Un software de analiză automată a imaginilor medicale care oferă măsurători precalculate pe baza detecțiilor vizuale ar trebui să fie capabil să evalueze calitatea rezultatului său și a predicțiilor DNN care stau la baza acestuia. Cercetările recente au arătat că estimările de incertitudine pot fi extrase din modele deja antrenate, utilizând tehnici simple. Dacă, de exemplu, în timpul instruirii modelului au fost utilizate straturi de normalizare a grupurilor sau straturi de mascare a activărilor, procedura de inferență în momentul testării poate fi actualizată pentru a oferi, de asemenea, incertitudinile ieșirilor împreună cu predicțiile modelului, prin rularea mai multor inferențe pentru același eșantion de test.

Secțiunile următoare prezintă metode care au avantajul de a necesita o singură inferență pentru a obține atât predicția, cât și incertitudinea aferentă. Segmentarea semantică, ca sarcină, poate fi descompusă într-un set de sarcini de mini-clasificare la nivel de pixel. Incertitudinile pot fi estimate în paralel pentru toți pixelii, deoarece rețeaua în sine emite valori de ieșire simultan pentru toți pixelii.

2.6.2 Modele bazate pe Energie

S-a arătat în [10] că un model de clasificare având o activare finală softmax conține implicit un estimator al densității de intrare. Se propune o funcție de energie care utilizează aceleași valori de intrare ca și stratul softmax:

$$E(x; f) = -T \log \sum_i^K \exp \left(\frac{f_i(x)}{T} \right) \quad (2.6)$$

unde T este temperatura softmax și $E(x; f)$ este valoarea energiei atașate a intrării x sub modelul f .

Scorul de energie este aliniat liniar cu densitatea de log-probabilitate a eșantioanelor de intrare:

$$\log p(x) = \frac{-E(x; f)}{T} - \log Z \quad (2.7)$$

unde $\log Z$ este constant pentru toate intrările x .

Un criteriu simplu pentru detecția OoD poate fi formulat pe baza scorurilor energetice observate pe setul de validare: un prag τ poate fi derivat și comparat cu scorurile energetice ale eșantioanelor de testare. De obicei, se utilizează scoruri energetice negative, pentru a se alinia la definiția convențională în care eșantioanele pozitive (din distribuția reală) au scoruri mai mari [10]:

$$G(x; \tau, f) = \begin{cases} 0 \text{ (OoD)}, & \text{if } -E(x; f) \leq \tau, \\ 1 \text{ (in-distrib)}, & \text{if } -E(x; f) > \tau. \end{cases} \quad (2.8)$$

Scorul de energie poate fi, de asemenea, încorporat în obiectivul de antrenare, împreună cu funcția criteriu pentru clasificare bazată pe entropia categorială. Setul de date de antrenare va fi format din 2 părți:

- un subset în-distribuție, pe care se calculează eroarea de clasificare și pentru care se minimizează scorul de energie.
- un subset OoD, pe care se calculează și se maximizează doar scorul energetic.

Metoda energiei este aplicabilă unui model deja antrenat, extinzându-l ca și clasificator OoD. Un model de segmentare și de detectare a reperelor antrenat pe ecocardiografiile apicale BMode (cu arhitectura sa prezentată în secțiunea 2.5.2.2) a fost aplicat pe un set de test ce conține ecocardiografiile apicale gradate pe o scară de la 1 la 5 în ceea ce privește calitatea imaginii regiunii LV (1 - cea mai bună calitate a imaginii, 5 - cea mai proastă calitate a imaginii). Au fost luate în considerare două grupări:

- gradul 1 & 2 & 3: Achiziții în care ventriculul stâng este destul de vizibil, cu zgomot de pixel redus, etc. Doar aceste niveluri de calitate au fost utilizate în setul de antrenare original al modelului.
- gradul 4 & 5: achiziții în care unele părți ale regiunilor peretelui ventriculului stâng nu sunt vizibile, zgomot de imagine ridicat, poziționare slabă a traductorului, etc. Datele de această calitate scăzută nu au fost niciodată văzute de model în timpul antrenării, prin urmare ar trebui să fie în afara distribuției în raport cu datele de calitate superioară.

A fost luat în considerare doar capul de segmentare al modelului pre-antrenat (adică nu a fost reglat în funcție de scorurile energetice). Deoarece segmentarea semantică implică clasificări la nivel de pixel pe întreaga imagine de intrare, valorile pre-softmax pe fiecare locație-pixel pot fi utilizate pentru a calcula scorurile energetice ale pixelilor. Fig. 2.6 prezintă un exemplu de segmentare cu harta energetică atașată. Ec. 2.7 afirmă că scorul energetic este aliniat liniar cu probabilitatea eșantionului de intrare. În contextul segmentării, fiecare pixel este clasificat pe baza câmpului vizual efectiv din imaginea de intrare.

Prin urmare, scorul de energie al unui pixel indică probabilitatea ca regiunea aferentă câmpului său receptiv (din imaginea de intrare) să se afle în interiorul distribuției de regiuni de imagine observate în timpul antrenării. Pentru a calcula un scor de energie care descrie întreaga regiune LV, s-a construit o regiune de interes (RoI) deasupra camerei LV folosind conturul LV și reperatele prezise de model. Scorurile energetice aferente pixelilor din interiorul acestei RoI pot fi mediate pentru a obține o estimare a incertitudinii întregii regiuni. A fost testată capacitatea de clasificare a modelului între cele două grupări de calitate vizuală a regiunii LV. Prin variația pragului utilizat pentru a separa cele două distribuții de scoruri, metrica AuRoC obținută de modelul pre-antrenat a fost de ~ 0.81 .

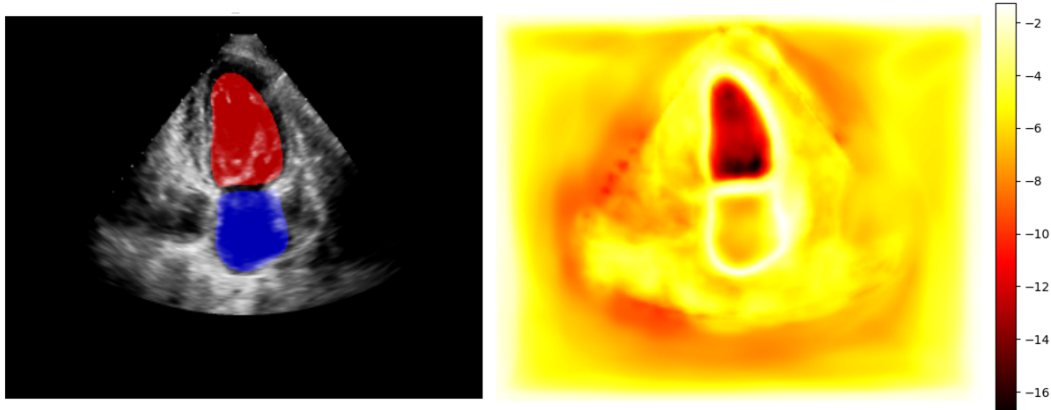


Figura 2.6: Exemplu de segmentare LV (masca albastră) și LA (masca roșie) pentru o achiziție A2C (stânga). Sub-plotul din dreapta arată scorurile energetice ale pixelilor. În jurul marginilor măștilor segmentate, scorurile energetice au valori relativ mari.

Scorurile de energie la nivel de pixel pot fi, de asemenea, mediate pe întreaga imagine, obținându-se o măsură globală a incertitudinii, utilă atunci când se testează pe seturi de date care pot fi foarte diferite în comparație cu setul de antrenare. Pe astfel de date, este posibil ca predicțiile obișnuite (de exemplu, segmentarea camerei LV) să nu fie efectuate din cauza lipsei unui conținut semantic adecvat. Această procedură poate acționa ca un mecanism de filtrare a datelor.

2.6.3 Procese Gaussiene Rare

2.6.3.1 Formularea Matematică

Un Proces Gaussian (GP) reprezintă o distribuție notată \mathcal{GP} peste funcțiile cu valori reale $f(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ definite pe un domeniu de intrare \mathcal{X} [12]. Analog unei Gaussiene multivariate, care reprezintă o distribuție pe vectori cu dimensiuni finite, un GP reprezintă o distribuție pe funcții cu o infinitate nenumărabilă de dimensiuni: indicii vectoriali din variabilele aleatoare Gaussiene multi-variate corespund conceptual unor puncte de evaluare specifice $X \in \mathcal{X}$ în funcțiile aleatoare GP [12].

Un GP este definit prin 2 funcții cu valori reale:

$$f(\cdot) \sim \mathcal{GP}(\mu(\cdot), k(\cdot, \cdot')) \quad (2.9)$$

unde $\mu(\cdot)$ este valoarea medie a funcției f și $k(\cdot, \cdot')$ formează matricea de covarianță a evaluărilor funcției f în punctele (\cdot) .

Considerând o partiție a setului infinit nenumărabil de variabile aleatoare reprezentat de un GP în două seturi:

- unul care conține un subset finit notat \mathbf{u} evaluat la un set finit de puncte $\mathbf{Z} = Z_1, \dots, Z_M \in \mathcal{X}$, astfel încât $\mathbf{u} = f(\mathbf{Z})$, cu media $\mu_{\mathbf{u}}$ și matricea de covarianță $\mathbf{K}_{\mathbf{u}\mathbf{u}}$.
- unul care conține restul setului infinit nenumărabil de variabile aleatoare notate ca $f(\cdot)$ evaluate în toate locațiile din \mathcal{X} , cu excepția punctelor din \mathbf{Z} .

Un GP poate fi rescris astfel:

$$\begin{pmatrix} f(\cdot) \\ \mathbf{u} \end{pmatrix} \sim \mathcal{GP} \left(\begin{pmatrix} \mu(\cdot) \\ \mu_{\mathbf{u}} \end{pmatrix}, \begin{pmatrix} k(\cdot, \cdot') & \mathbf{k}_{\cdot\mathbf{u}} \\ \mathbf{k}_{\mathbf{u}\cdot'} & \mathbf{K}_{\mathbf{u}\mathbf{u}} \end{pmatrix} \right) \quad (2.10)$$

unde $\mathbf{k}_{\cdot\mathbf{u}}$ și $\mathbf{k}_{\mathbf{u}\cdot'}$ reprezintă funcții cu valori vectoriale care exprimă inter-covarianța dintre variabila aleatoare finită-dimensională u și variabila aleatoare nenumărabilă infinit-dimensională $f(\cdot)$; și anume, $\mathbf{k}_{\cdot\mathbf{u}}$ este o funcție cu valoare vectorială egală cu $k(\cdot, Z_m)$.

Procesul Gaussian condițional al $f(\cdot)$ condiționat pe \mathbf{u} este:

$$f(\cdot)|\mathbf{u} \sim \mathcal{GP}(\mu(\cdot) + \mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{u}\mathbf{u}}^{-1}(\mathbf{u} - \mu_{\mathbf{u}}), k(\cdot, \cdot') - \mathbf{k}_{\cdot\mathbf{u}}\mathbf{K}_{\mathbf{u}\mathbf{u}}^{-1}\mathbf{k}_{\mathbf{u}\cdot'}) \quad (2.11)$$

În ecuația de mai sus, \mathbf{u} poate fi un vector de măsurători scalare (de exemplu, valorile ieșire țintă) care corespund unui set \mathbf{Z} de eșantioane de intrare de antrenare. $\mu(\cdot)$ poate fi fixat la zero, iar kernel-ul k se aplică pe toate perechile (\cdot, Z_i) , obținându-se matricile necesare.

2.6.3.2 Estimarea Deterministă a Incertitudinii în Învățarea Profundă

Având în vedere seturile de date de imagini, aplicarea kernel-urilor standard k direct pe perechile (X_i, X_j) de imagini ar duce la o expresivitate slabă a modelării, deoarece toate operațiile nucleului ar fi efectuate în spațiul pixelilor. În schimb, Deep Kernel Learning (DKL) [14] încearcă să combine expresivitatea rețelelor neuronale adânci cu capacitatea de predicție probabilistică a GP-urilor [11] prin utilizarea unui extractor de caracteristici bazat pe DNN pentru a alimenta un strat de ieșire GP. Prin urmare, nucleul ar conține un extractor adânc de caracteristici [11]:

$$k_{l,\theta} \rightarrow \bar{k}_l(f_\theta(x_i), f_\theta(x_j)) \quad (2.12)$$

unde $f_\theta(\cdot)$ este un DNN parametrizat de θ , \bar{k}_l poate fi un nucleu standard (cum ar fi funcția de bază radială sau Matérn) cu hiper-parametri l (de exemplu, lungimea nucleului și scala de ieșire).

Procedura de instruire prin inferență variațională poate fi utilizată pentru a evita stocarea reprezentărilor caracteristice ale tuturor celor N exemple de antrenare și calcularea inversei unei matrice N -pe- N . În schimb, se învață M puncte de inducție și M variabile latente de inducție în spațiul caracteristic, cu $M \ll N$.

Colapsul caracteristicilor este un fenomen [11] în care reprezentările exemplurilor de intrare pot fi plasate pe un sub-spațiu cu dimensiuni reduse, micșorând efectiv anumite regiuni ale spațiului de intrare în vecinătăți foarte înguste în spațiul caracteristicilor și, prin urmare, făcând imposibilă distincția între anumite puncte de intrare, chiar dacă acestea pot fi inițial plasate la o distanță foarte mare una de cealaltă. În [11] se propune ca extractorul de caracteristici să fie constrâns să fie bi-Lipschitz pentru a atenua colapsul caracteristicilor:

$$\frac{1}{K}d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2)) \leq Kd_X(x_1, x_2) \quad (2.13)$$

unde $K \geq 1$, $f : X \rightarrow Y$ este extractorul de caracteristici DNN, iar d_X și d_Y sunt funcții de distanță pe setul X și, respectiv, Y . Prin urmare, extractorul de caracteristici trebuie să aibă 2 proprietăți: sensibilitate și netezime.

Ieșirea unui strat final GP este o distribuție gaussiană. În ceea ce privește segmentarea semantică, ieșirea este o distribuție de valori de apartenență la clasă nenormalizate, per pixel. Cu toate acestea, aplicarea directă a unui strat de ieșire GP asupra reprezentărilor fiecărui pixel din întreaga imagine de intrare este problematică din cauza constrângerilor de memorie și de calcul. În schimb, un cap de segmentare DNN clasic poate preselecta regiuni mai mici din imaginea de intrare care prezintă interes pentru o anumită sarcină și poate elimina celelalte zone irelevante pentru a economisi memorie și calcul.

A fost realizat un experiment în care un model format dintr-un extractor adânc de caracteristici bi-Lipschitz care alimentează două capete de decodificare a fost antrenat pe sarcina de segmentare a camerei LV din ecocardiografii. Primul cap este responsabil pentru realizarea unei segmentări semantice a imaginii complete, pe baza căreia sunt extrase regiunile de interes pentru GP. În acest experiment, primul cap de decodificare segmentează camera LV, iar conturul acesteia este extras și dilatat pentru a obține o bandă mai groasă în jurul LV. Este probabil ca incertitudinea să fie mai semnificativă în interiorul acestui "inel" care înconjoară LV, deoarece pot exista mai multe trasee plauzibile de contur care să segmenteze LV, pentru o anumită imagine de intrare.

Reprezentările pixelilor din interiorul "inelului" sunt extrase din harta caracteristicilor acestora și remodelate în b tensori de forma (p_i, c) , unde b este numărul de eşantioane din grupul de intrare al extractorului de caracteristici, p_i este numărul de pixeli selectați pentru exemplul i și c este numărul de canale din harta caracteristicilor, adică dimensiunea reprezentării pentru fiecare pixel. Această procedură de preselecție modelează în comun pixeli din aceeași imagine, dar acționează independent asupra fiecărei imagini de intrare.

În timpul fiecărei etape de antrenare, cele două capete de decodificare împart caracteristici intermediare calculate de către encoder-ul adânc bi-Lipschitz (care constă într-un UNet rezidual normalizat spectral). Primul cap are o funcție criteriu DICE, în timp ce al doilea cap folosește media funcției criteriu ELBO calculate pe imaginile din întregul grup de intrare curent. Pentru a evita denaturarea datelor din cauza procedurii de preselecție, cel de-al doilea cap se aplică, de asemenea, pe seturi de reprezentări din alte locații de pixeli alese aleatoriu. În acest fel, GP nu este antrenat doar pe regiuni din jurul LV, ci acoperă (în medie) întregul spațiu al imaginii.

Modelul poate fi antrenat complet și concomitent. În timpul testării, se poate utiliza aceeași procedură de preselecție. Fig. 2.7 prezintă un exemplu de ieșire. Partea interioară a măștii prezise de capul clasic poate fi cusută cu componenta exterioară a măștii (de-a lungul inelului gros LV) prezisă de capul GP. Se poate observa că incertitudinea este concentrată de-a lungul mijlocului benzii de pixeli modelate de GP, adică de-a lungul marginii măștii finale prezise. Deoarece rezultatul GP este o distribuție, extragerea mai multor eşantioane poate fi utilizată pentru a construi mai multe componente exterioare ale măștii și, prin îmbinarea cu partea interioară constantă, poate produce mai multe variante de măști LV prezise.

2.6.4 Concluzii

Metodele de estimare a incertitudinii oferă o perspectivă valoroasă asupra calității potențiale a predicției unui model DNN. Astfel de metode pot fi, de asemenea, utilizate pentru a îmbunătăți în continuare modelul, prin detectarea subseturilor de date pentru care incertitudinea modelului este în mod consistent ridicată. Setul de antrenare poate fi actualizat cu mai multe date de acest tip, iar o re-antrenare ar trebui să reducă incertitudinea, îmbunătățind în același timp precizia predicției. Unele metode de incertitudine au avantajul de a fi ușor de aplicat modelelor deja antrenate, adăugând astfel valoare la soluțiile de inteligență artificială preexistente, permițându-le să detecteze intrările defectuoase sau ieșirile cu o incertitudine prea mare, din diverse etape din cadrul fluxului interior de lucru.

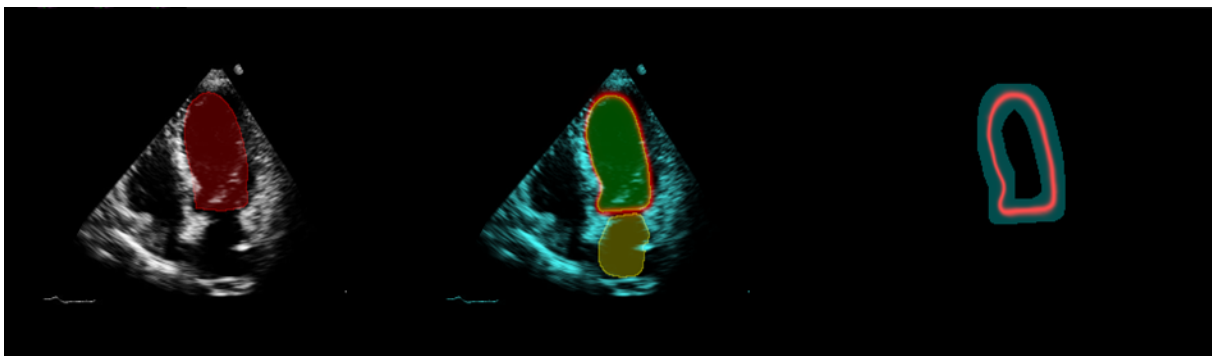


Figura 2.7: Exemplu de ieșire de la un GP rar variațional antrenat pentru sarcina de segmentare a LV. Subplotul din stânga prezintă datele de intrare împreună cu masca țintă LV; subplotul din mijloc prezintă măștile LV (verde) și LA (galben) prezise; subplotul din dreapta prezintă harta de incertitudine a măștii prezise (roșu) și zona pixelilor modelați de GP (bandă albastră).

3. Detectarea Punctelor de Reper în Ecocardiografiile 3D utilizând Învățarea prin Recompensă

3.1 Introducere

Învățarea prin recompensă (RL) este o categorie de metode de învățare automată în care un agent interacționează cu un mediu pe baza unui set predefinit de acțiuni (continue sau discrete), își observă starea curentă și primește o recompensă dată de mediu ca efect al unei acțiuni specifice întreprinse într-o anumită stare.

În [21] a fost introdus o metodă de căutare a reperelor pe mai multe scale bazat pe învățarea profundă prin recompensă pentru a rezolva sarcina de detectare a reperelor anatomice în scanările de tip CT. Aceasta a atins o precizie de ultimă generație, fiind în același timp cu câteva ordine de mărime mai rapidă decât metodele de referință. Agenții artificiali sunt antrenați nu numai să distingă aspectul organelor, ci și să navigheze spre reperele țintă de-a lungul unor traiectorii optime.

În acest capitol, metoda din [21] este adaptată pentru detectarea a 6 repere ale ventriculului stâng (LV) pe ecocardiografiile TTE BMode 3D. Trei agenți multi-scală sunt antrenați să urmeze traiectorii optime în timp ce caută reperele în interiorul volumului ecografic. Suita de agenți RL reduce necesitatea de a rula un DNN convoluțional 3D pe întregul volum ecografic, ceea ce duce la economii de calculații și timp de execuție.

3.2 Metode

3.2.1 Învățarea prin Recompensă

Modelul matematic clasic care descrie funcționarea agentului în interiorul mediului este un proces de decizie Markov (MDP), care este un proces de control stocastic în timp discret. Acesta constă în [20]: un set de stări ale mediului: $S_t \in \mathcal{S}$; un set de acțiuni ale agentului (posibile în orice stare dată s): $A_t \in \mathcal{A}(s)$; un set de recompense numerice: $R_{t+1} \in \mathcal{R} \subset \mathbb{R}$; o funcție p care descrie dinamica mediului. Având în vedere o stare curentă s și o acțiune aleasă a , p oferă probabilitatea următoarei stări s' și a recompensei r :

$$p(s', r|s, a) \doteq \Pr \{S_t = s', R_t = r | S_{t-1} = s, A_{t-1} = a\} \quad (3.1)$$

Dacă seturile \mathcal{S} , \mathcal{A} și \mathcal{R} au un număr finit de elemente, atunci MDP-ul descris este finit. Indicele t se referă la momentul discret de timp. O traiectorie a MDP este o secvență de stări, acțiuni și recompense obținute prin rularea agentului în interiorul mediului:

$$S_0, A_0, R_1, S_1, A_1, R_2, S_2, A_2, R_3, \dots \quad (3.2)$$

Scopul agentului este de a maximiza recompensa sa cumulativă. Din cauza posibilei existențe a unor episoade foarte lungi, suma recompenselor poate tinde spre valori foarte mari în timpul funcționării agentului, de aceea se utilizează conceptul de amortizare pentru a calcula rentabilitatea amor-

tizată, cu condiția ca agentul să încerce să aleagă acțiuni care să maximizeze recompensele:

$$G_t \doteq R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \quad (3.3)$$

unde γ este coeficientul de amortizare cu $0 \leq \gamma \leq 1$.

Rezolvarea unei sarcini de învățare prin recompensă implică găsirea unei politici bune π care să obțină recompense viitoare mari. O politică π este considerată mai bună decât (sau egală cu) o altă politică π' dacă randamentul său mediat este mai mare decât (sau egal cu) cel al lui π' pentru toate stările:

$$\pi \geq \pi' \text{ dacă și numai dacă } v_{\pi}(s) \geq v_{\pi'}(s), \forall s \in \mathcal{S} \quad (3.4)$$

O politică optimă este mai bună sau egală cu toate celelalte politici. Pot exista multiple politici optime. Toate acestea sunt notate cu π_* și au aceeași funcție de valoare (optimă) a stării:

$$v_*(s) \doteq \max_{\pi} v_{\pi}(s), \forall s \in \mathcal{S} \quad (3.5)$$

Politicile optime au, de asemenea, aceeași funcție optimă acțiune-valoare:

$$q_*(s, a) \doteq \max_{\pi} q_{\pi}(s, a), \forall s \in \mathcal{S}, a \in \mathcal{A} \quad (3.6)$$

Pentru o stare s și o acțiune efectuată a , această funcție oferă randamentul așteptat în cazul în care politica optimă este urmată ulterior:

$$q_*(s, a) = \mathbb{E}[R_{t+1} + \gamma v_*(S_{t+1}) | S_t = s, A_t = a] \quad (3.7)$$

3.2.2 Agenți Multi-scală pentru Ecocardiografii 3D

În această secțiune, metoda din [21] a fost adaptată pentru detectarea a 6 repere LV pe ecocardiografiile TTE BMode 3D: 4 repere de anulus și 2 repere de apex. Dat fiind un volum ecocardiografic de $128 \times 128 \times 128$, metoda de detecție implică aplicarea succesivă a agenților antrenați la mai multe scări: $32 \times 32 \times 32$ (cea mai grosieră), $64 \times 64 \times 64$ și $128 \times 128 \times 128$ (cea mai fină).

Mediul este grila care indexează volumul ecografic, iar o stare este complet descrisă de poziția curentă a agentului în interiorul volumului său la scară. Setul de acțiuni \mathcal{A} are 6 acțiuni posibile (a se vedea Fig. 3.1): sus/jos, stânga/dreapta și față/spate, care deplasează poziția agentului cu un increment de-a lungul direcțiilor specificate.

O stare este considerată terminală dacă este suficient de aproape de locația țintă (GT) sau dacă este în afara limitelor volumului ecografic de intrare. Agentul observă o regiune $23 \times 23 \times 23$ centrată pe

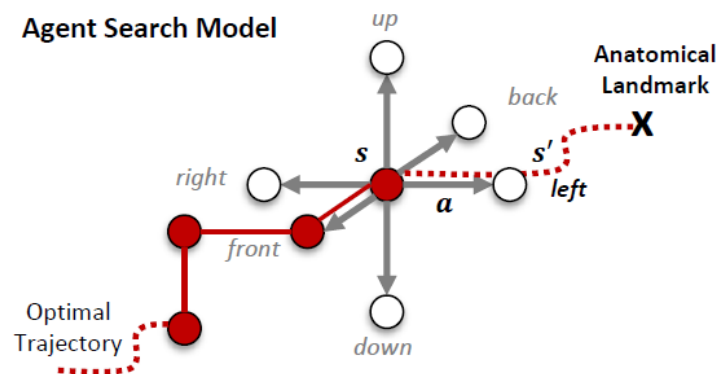


Figura 3.1: Modelul de căutare a reperelor folosit de agenți. Există 6 acțiuni posibile în fiecare stare. Tensorul de stare este reprezentat de o regiune de volum ecografic centrată pe poziția curentă a agentului. Figura preluată din [21].

locația sa curentă. Fragmentul regiunii este extras din volumul de ecocardiografie scalat și introdus în rețeaua Q a agentului, care prezice perechi acțiune-valoare pentru toate cele 6 acțiuni posibile. În cazul în care se află aproape de granița volumului scalat, fragmentul regiunii observat utilizează valori surogat de zero la indexarea în afara volumului de intrare.

Sistemul de recompensare adoptat a fost o sumă de 3 componente: (i) una proporțională cu modificarea relativă a distanței față de GT; (ii) un cost fix pentru orice acțiune; (iii) o recompensă specială ori de câte ori GT este atins (o recompensă pozitivă) sau ori de câte ori agentul iese din limita volumului (o recompensă negativă).

Acest sistem de recompensare face ca funcția de valoare a stării să scadă pe măsură ce stările se îndepărtează de GT. Dacă agentul ajunge la locația GT se acordă o recompensă mare, în timp ce dacă pășește în afara volumului este puternic penalizat. Deoarece recompensa este întotdeauna negativă (cu excepția cazului în care se ajunge la GT terminal), o politică de control optimă este aceea de a face cât mai puțini pași posibil spre GT.

Arhitectura rețelei Q a fost inspirată din MobileNetV2 [22] și a fost concepută pentru a fi foarte mică, având doar 87k parametri. Pentru a reduce timpul de execuție la doar câteva milisecunde, fiecare etapă a avut doar un singur bloc rezidual inversat și, prin urmare, nu au fost posibile conexiuni de salt. În schimb, activări de tip LeakyReLU au fost plasate între toate etapele. Rețeaua a avut 2 canale de intrare: pixelii volumului ecocardiografic și o mască de segmentare a pereților LV (furnizată de un model de segmentare 3D plasat în amonte).

Trei agenți au fost antrenați pe două volume de 0.25x și de 0.5x cu sub-eșantionare și pe volumul original. Primul agent efectuează o localizare grosieră. Deoarece dimensiunea intrării sale este de 23x23x23 în interiorul unui volum de 32x32x32, agentul are un câmp vizual relativ de $\sim 72\%$, similar straturilor de tip "bottleneck" dintr-un model DNN clasic de tip imagine-la-imagine. În timpul testării, agenții următori rafinează progresiv locația de căutare, deoarece punctul de pornire al etapei următoare este punctul de convergență al etapei anterioare (a se vedea Fig. 3.2). Similar straturilor finale ale unei rețele imagine-la-imagine care se ocupă de detalii spațiale progresiv mai fine, pe măsură ce scara volumului ecografic crește spre scara originală, agenții multi-scală învață să interpreteze de la detalii spațiale grosiere pe câmpuri vizuale mari (atunci când operează la scări mai mici) la detalii spațiale fine pe câmpuri vizuale mici (atunci când operează la scări mai mari).

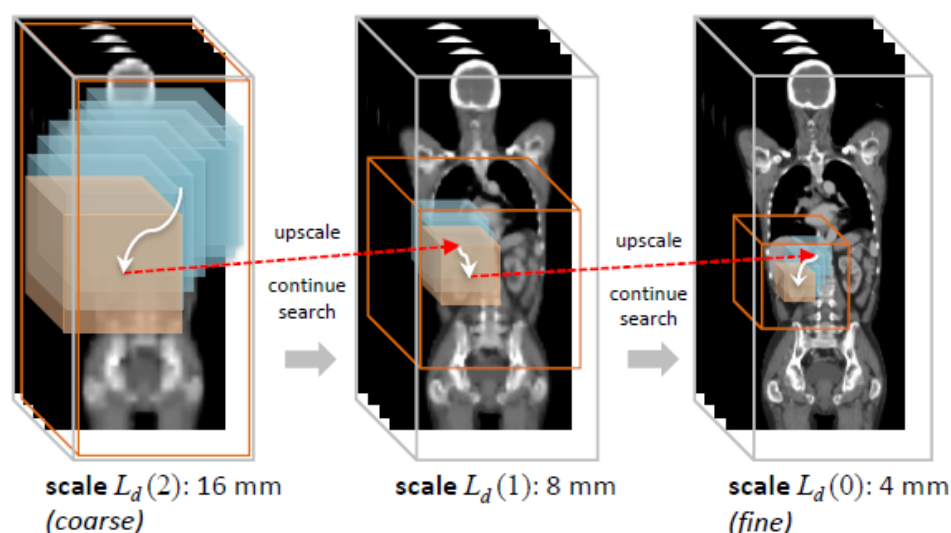


Figura 3.2: Conducța de detectare a reperelor pe mai multe scări. Locația finală a etapei anterioare este utilizată ca punct de plecare pentru etapa următoare. Fiecare etapă implică o creștere de 2 ori a rezoluției spațiale. Săgețile albe reprezintă traiectorii de căutare locală. Cuburile albastre și maro denotă regiuni de stare din volumul ecografic de-a lungul traiectoriei agentului (adică tensori de intrare în rețeaua Q a agentului). Figura preluată din [21].

La toate scările volumului ecografic, fiecare episod de antrenare alege puncte de pornire aleatoare în interiorul unei vecinătăți cubice predefinite de dimensiuni fixe n_{start} în jurul locației GT, în timp ce regiunea de deplasare permisă n_{travel} în timpul unui episod este mai mare decât n_{start} cu o poziție de 1 pixel pe fiecare dintre cele 6 laturi. Atingerea unei stări/locații în afara lui n_{travel} a fost, de asemenea, considerată terminală.

Fiecare episod de antrenare a avut o lungime maximă impusă de 36 de pași, un număr proporțional cu pașii necesari agentului cu cel mai grosier volum pentru a ajunge la GT în cel mai rău caz. O epocă de antrenare a reunit etapele de antrenare simultană ale tuturor agenților și a fost dimensionată astfel încât 5% din setul de antrenare să fie acoperit în timpul unei epoci complete. O procedură completă de antrenare a implicat 200 de epoci, acoperind astfel de 10 ori întregul set de antrenare (set ce a inclus 2000 de pacienți, fiecare având două volume corespunzătoare punctelor de timp ED și ES). Pe măsură ce antrenarea a progresat, coeficientul ϵ utilizat în politica de tip ϵ -greedy a fost micșorat în mod liniar de la o valoare inițială de 1 la o valoare minimă de 0.05 pe parcursul primelor 80% din epoci. Ca urmare, politica a acționat progresiv mai determinist în timpul antrenării, iar procentul de episoade convergente a crescut invers proporțional cu ϵ .

În momentul testării, GT este utilizat doar pentru calculul metricii distanței și, prin urmare, nu poate fi utilizat pentru a marca o anumită locație ca fiind terminală în interiorul volumelor ecografice scalate. Prin urmare, s-a utilizat un criteriu de terminare pentru a detecta convergența agentului:

- fiecare locație vizitată a fost plasată în interiorul unei liste sub forma unui tuplu de indici de volum (x, y, z) .
- dacă locația curentă a fost vizitată anterior de către agent, atunci s-a produs un ciclu pe traiectoria sa. Pentru a obține estimarea finală a locației reperului (la scara volumului ecografic curent), s-a calculat media tuturor locațiilor care fac parte din ciclu.

Punctul de plecare pentru cei mai grosieri agenți a fost centrul de masă al măștii peretelui LV, pentru toate cele 6 reperi. Agenții succesivi au folosit ca punct de plecare locația reperului convergent din etapa anterioară, ca în Fig. 3.2.

3.3 Rezultate

Tab. 3.1 prezintă metrici de eroare pe un set de test pentru toate cele 6 reperi LV atunci când se utilizează soluția de căutare multi-scală bazată pe RL. Unele reperi prezintă o performanță mai bună de detectare, deși toate au o eroare medie a reperelor de $\leq 2.1\%$ din dimensiunea volumului ecografic original. Mai mult de 90% din cazuri sunt detectate la o distanță de sub 5 pixeli față de GT (la o rezoluție originală de 128x128x128), pentru toate reperele.

Tabela 3.1: Metrici de detecție a reperelor TTE LV (pe volume originale de dimensiune 128x128x128) utilizând agenți RL multi-scală. Ultima coloană arată procentul de cazuri de test pentru care predicția a fost în interiorul unei sfere cu raza de 5 pixeli în jurul GT.

ID reper	medie	std	max	min	mediana	% <5px
0	2.53	1.39	6.93	0.77	2.38	91.8
1	2.18	1.17	5.68	0.45	1.74	97.9
2	2.64	1.51	8.28	0.75	2.25	91.8
3	2.05	0.95	4.95	0.23	1.88	100
4	1.94	1.36	7.40	0.27	1.69	93.8
5	2.46	1.60	8.75	0.43	2.06	95.9

3.4 Concluzii

Acest capitol a prezentat o abordare a detectării reperelor 3D folosind o cascadă de agenți RL la multiple scale. Utilizând metoda versatilă a învățării Q profunde, trei agenți au învățat să urmeze traiectorii optime către locațiile țintă, navigând în interiorul volumului ecografic. Această metodă a avut mai multe avantaje, cum ar fi eficiența datelor și stabilitatea antrenării. O arhitectură DNN mică a fost investigată pentru o predicție rapidă și robustă a valorii acțiunilor.

4. Detectarea Segmentărilor Incorecte ale Lumenului în Angiografiile Coronariene prin Tomografie Computerizată¹

4.1 Introducere

Angiografia coronariană prin tomografie computerizată (CCTA) este o metodă eficientă de imagistică, din ce în ce mai acceptată ca un test de primă linie pentru a diagnosticul bolii coronariene (CAD). Progresele în domeniul CCTA au permis o expunere minimă la radiații, o caracterizare coronariană eficientă și imagistică detaliată a aterosclerozei în timp. Datorită numărului tot mai mare de dovezi care arată eficacitatea CCTA [28, 29], ghidurile recente ACC/AHA privind durerea toracică recomandă CCTA ca test de primă linie pentru pacienții cu dureri toracice stabile și acute.

În timp ce performanța metodelor bazate pe inteligență artificială s-a îmbunătățit considerabil de-a lungul anilor, având în vedere importanța unei detecții precise a lumenului, în prezent se folosesc încă abordări semiautomate. Astfel, lumenul este mai întâi detectat automat, iar apoi inspectat și editat manual de către radiolog, dacă se consideră necesar. Acest proces, împreună cu editarea liniei centrale a arterei coronare, a necesitat, de exemplu, între 10 și 60 de minute într-un studiu de evaluare a performanței de diagnosticare a CT-FFR bazate pe ML pentru detectarea bolii coronariene obstructive din punct de vedere funcțional [30]. O abordare potențială pentru reducerea semnificativă a timpului necesar pentru o analiză semiautomată a lumenului CCTA este de a preselecția locațiile suspecte să necesite inspecție și editare, și să se prezinte radiologului doar acestea pentru examinare. Având în vedere că o rețea neurală adâncă (DNN) este responsabilă de generarea măștilor de segmentare a lumenului, această etapă de preselecție poate fi legată de subiectul incertitudinii și a detectării intrărilor aberante (OoD) în metodele de Deep Learning.

Modelele Normalizing Flow (NF) pot fi antrenate în mod explicit pentru a modela densitățile de probabilitate ale datelor de intrare. Având în vedere o sarcină țintă din aval T , în cazul în care doar datele de intrare ale acestora sunt utilizate pentru construirea modelului NF, atunci estimarea probabilității eşantioanelor de intrare pentru sarcina țintă poate fi obținută prin intermediul modelului NF. Eşantioanele de intrare cu probabilități scăzute pot fi marcate ca fiind din afara distribuției, iar ieșirea modelului țintă ar trebui să fie considerată nesigură, deoarece ar funcționa în afara distribuției sale de antrenare. Un model NF poate fi construit, de asemenea, prin concatenarea eşantioanelor de intrare cu ieșirea GT țintă a acestora. În acest fel, modelul NF poate fi plasat în aval de sarcina țintă și poate acționa ca un model de audit, detectând cazurile în care modelul anterior a furnizat predicții eronate. În ambele scenarii, NF este un model separat și, prin urmare, nu impune nicio constrângere asupra modelului responsabil pentru sarcina țintă. De obicei, NF poate funcționa eficient în ambele direcții: direcția înainte (sau de inferență), în care eşantioanele de intrare x din domeniul de intrare X sunt transformate în reprezentări z care sunt probabile în cadrul unei distribuții alese Z . La fiecare strat,

¹Următorul capitol descrie experimente și analize publicate în:

■ **Ciusdel, C.F.**, et al., 2022. Normalizing Flows for Out-of-Distribution Detection: Application to Coronary Artery Segmentation. *Appl. Sci.* 12, 3839. <https://doi.org/10.3390/app12083839>

Unele secțiuni au fost citate textual din referința de mai sus, care reprezintă o lucrare publicată anterior de către autor, în cadrul programului de cercetare doctorală.

datele de intrare sunt modificate către Z , iar valoarea $\log \text{Det}$ (adică $\ln \left(\left| \det \left(\frac{\partial f}{\partial x} \right) \right| \right)$, unde f este NF) se însumează cu contribuția stratului curent. Direcția inversă (sau generativă) utilizează proprietatea de bijecție a NF pentru a transforma o reprezentare z într-un eșantion sintetic x_{new} .

În această lucrare se prezintă o abordare bazată pe NF pentru detectarea OoD a segmentărilor lumenului coronarian. Modelele NF care sunt construite din straturi de cuplare, așa cum se propune în [23, 24], tind să se concentreze pe corelațiile locale ale pixelilor în locul semnificației semantice globale [25, 26] și, ca urmare, eșantioanele OoD pot produce, de fapt, valori mai mari ale log-probabilității decât datele de tip în-distribuție. Se investighează utilizarea unui nou tip de strat de cuplare, care utilizează convoluții reversibile 1x1 în care parametrii de filtrare sunt calculați pe baza componentelor nemodificate. Se compară arhitectura propusă cu o arhitectură de tip Glow în ceea ce privește sarcina de detectare a perechilor greșite de imagini ale lumenului CCTA și a segmentărilor corespunzătoare de lumen. Imaginile lumenului coronarian și măștile sunt volume 3D suprapuse de-a lungul axei canalului. De asemenea, se utilizează perturbații sintetice asupra măștilor binare și se folosesc eșantioanele perturbate ca valori aberante explicite pentru a modela densitatea de probabilitate învățată a perechilor imagine-mască "corecte". Scopul final este de a semnaliza acele eșantioane pentru care segmentarea dată nu se potrivește corect cu imaginea lumenului. Se evaluează performanța modelelor NF astfel: (i) în raport cu perturbațiile măștii sintetice și (ii) utilizând adnotări de la experți.

4.2 Metode

4.2.1 Arhitecturi NF

S-a investigat utilizarea unei arhitecturi NF de tip Glow, combinând straturile introduse anterior în [23, 24], cum ar fi straturile de cuplaj de tip "tabla de șah" și de mascare a canalelor, convoluțiile 1x1 inversabile, straturile de Divizare și Redimensionare. Rețeaua de bază este reprezentată în Fig. 4.1 și descrisă în Tab. 4.1. S-au utilizat straturi de cuplaj afine ca în ecuația (4.1), unde x și y sunt tensorii de intrare și, respectiv, de ieșire. Indicii a și b desemnează în mod obișnuit cele două jumătăți ale tensorilor: una care este transmisă neschimbată și cealaltă care este actualizată în mod liniar în raport cu ea însăși, dar într-un mod foarte neliniar în raport cu prima jumătate, prin intermediul funcțiilor s și t (care sunt rețele neuronale adânci).

$$\begin{aligned} y_a &= x_a \\ y_b &= (x_b - t_{DNN}(x_a)) s_{DNN}(x_a) \end{aligned} \quad (4.1)$$

Rețelele s și t sunt, în acest caz, un CNN 3D cu două capete. Funcția de activare finală a capului s a fost aleasă ca $\exp(\tanh(x))$ pentru a calcula cu ușurință contribuția la $\log \text{Det}$ (ca $\sum \tanh(x)$ peste toate dimensiunile spațiale și toate canalele) și pentru a oferi o limită de $[e^{-1}, e^1]$ pentru scalarea efectuată la fiecare strat de cuplare, asigurând stabilitatea numerică și o valoare maximă globală limitată a $\log \text{Det}$.

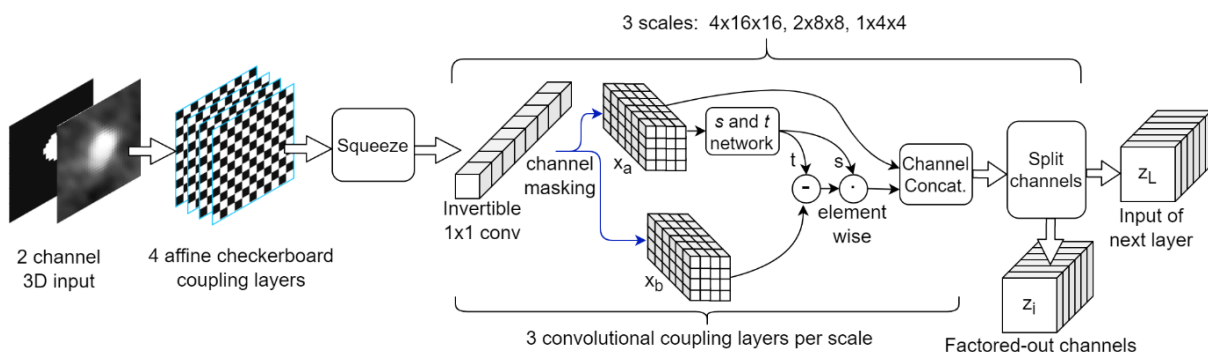


Figura 4.1: Arhitectura modelului de bază. Straturile "Normalizare Activare" nu sunt afișate.

Tabela 4.1: Arhitectura de bază de tip Glow

Etapa	Nr. Blocuri	Descriere bloc	Rezoluție	Nr. canale	Nr total de parametri
1	4	Strat de cuplare afină folosind masca de tip "tablă de șah"; Normalizare Activare (nu pe ultima etapă)	8x32x32	2	~ 2 milioane
2 3 4	1	Operație 3D de redimensionare	Etapa 2: 4x16x16 Stage 3: 2x8x8 Etapa 4: 1x4x4	Etapa 2: 16 Etapa 3: 64 Etapa 4: 256	
	3	Normalizare Activare; Convoluție 1x1 inversabilă; Strat de cuplare afină utilizând mascarea canalelor			
	1	Divizarea canalelor			

Eșantioanele de intrare constau în grupuri de 8 secțiuni transversale adiacente (sub-eșantionate la o rezoluție de 32x32) și 2 canale (concatenarea angiografiei și a măștii binare).

În [25] s-a demonstrat că NF care utilizează straturi de cuplare afine sunt predispuse să se concentreze mai mult pe corelațiile locale ale pixelilor decât pe conținutul semantic și să exploateze coadaptarea straturilor de cuplare pentru a maximiza log-probabilitatea finală. Motivați de aceste constatări, propunem utilizarea unui nou tip de strat de cuplare, unul care poate funcționa eficient pentru ambele direcții ale NF, nu se concentrează pe corelațiile locale ale pixelilor și are o structură inductivă similară cu cea a CNN-urilor convenționale. Stratul seamănă cu o secvență standard de tip Glow de 1x1 Convoluție inversabilă, mascare a canalelor, strat de cuplare afină. Cu toate acestea, ultimul pas este înlocuit cu o convoluție 1x1 (cu tensor de polarizare) ai cărei parametri sunt calculați pe baza canalelor nemodificate, ca în [27]. Polarizarea este aplicată în toate pozițiile spațiale, prin urmare are aceleași valori pe toată lățimea, înălțimea și adâncimea tensorului rezultat, ceea ce înseamnă că stratul nu mai este capabil să reproducă valorile pixelilor mascați, așa cum reiese din [25]. Același filtru de convoluție (specific fiecărui eșantion) se aplică în toate pozițiile spațiale, spre deosebire de calculul per-element efectuat în (4.1). Acest comportament este similar cu cel al CNN-urilor clasice, cu excepția faptului că acum ponderile filtrelor nu sunt aceleași pentru toate eșantioanele. Ecuația (4.2) descrie funcționarea stratului, cu o notație simplificată: $*$ înseamnă o convoluție 1x1 cu filtru k și $+$ este o sumă de difuzare. k este calculat de un CNN și are forma $C_{modif}-pe-C_{modif}$, unde C_{modif} este numărul de canale care sunt actualizate. b este un vector de lungime C_{modif} .

$$y_a = x_a$$

$$y_b = x_a * k(x_a) + b(x_a) \quad (4.2)$$

A fost proiectată o nouă arhitectură NF care utilizează stratul de cuplare de mai sus. Prima etapă este o secvență de straturi de cuplaj aditiv cu mascare de tip "checkerboard". Conform [25], aceste straturi se vor concentra în principal pe corelațiile locale ale pixelilor, dar acest lucru este echivalent cu funcționarea primelor straturi din CNN-urile clasice, unde câmpul vizual receptiv este mic, iar filtrele tind să caute modele simple, cum ar fi colțuri, muchii, texturi etc. Spre deosebire de cuplajele afine, cuplajele aditive conservă volumul, adică nu contribuie direct la logDet și la log($p(x)$) final, ci indirect prin intermediul straturilor din amonte.

Etapele următoare constau în cascade de straturi de cuplare, așa cum sunt descrise în Fig. 4.2 și Tab. 4.2. Spre deosebire de un CNN clasic, în care filtrele de formă 3x3 (sau mai mari) și pasurile mai mari de 1 sunt utilizate (fie în straturile convoluționale, fie în cele de decimare prin maximizare) pentru a crește câmpul vizual efectiv (FoV), în arhitectura propusă, FoV-ul este crescut doar prin operațiile

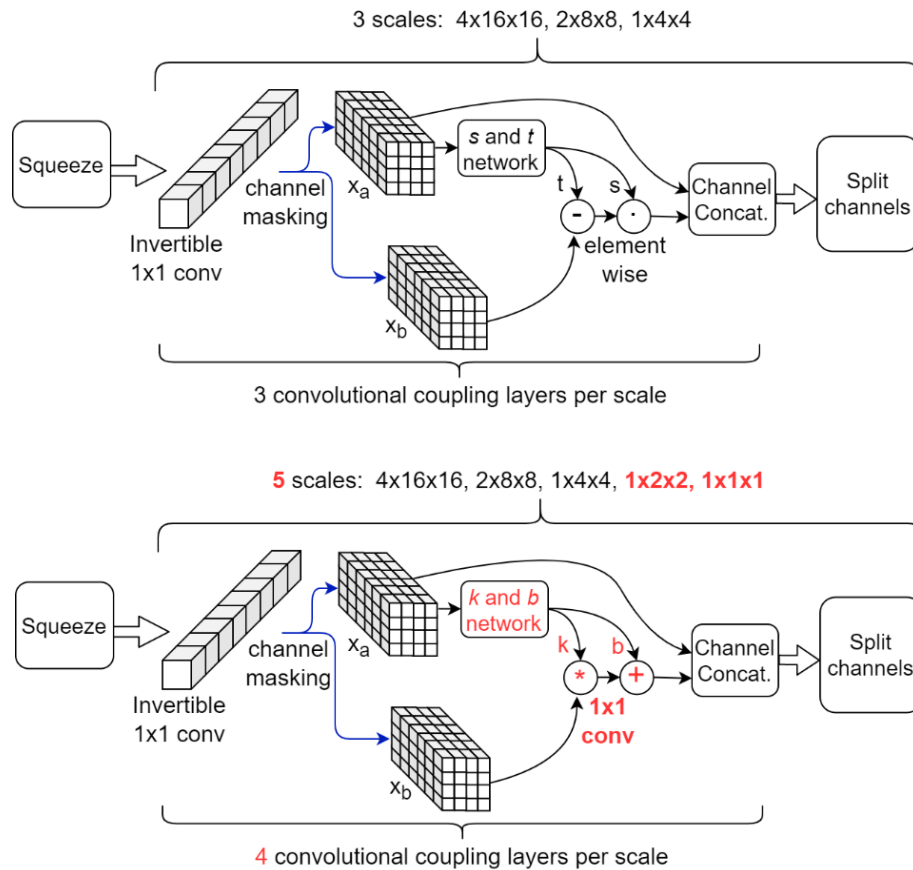


Figura 4.2: Comparație între arhitectura interioară de bază (sus) și cea propusă (jos) care utilizează noul strat de cuplare. Părțile actualizate sunt evidențiate cu roșu. Straturile de normalizare nu sunt reprezentate.

Tabela 4.2: Arhitectura NF îmbunătățită prin utilizarea straturilor de cuplare noi.

Etapă	Nr. Blocuri	Descriere bloc	Rezoluție	Nr. Canale	Nr total de parametri
1	4	Strat de cuplare aditivă folosind masca de tip "tablă de șah"; BatchNorm (nu pe ultimul bloc)	8x32x32	2	~8.7 milioane
2	1	Operație 3D de redimensionare		Etapă 2: 16	
3	4	BatchNorm; Convoluție 1x1 inversabilă; strat de cuplaj convoluțional folosind mascarea canalelor	Etapă 2: 4x16x16	Etapă 3: 64	
4				Etapă 3: 2x8x8	
5				Etapă 4: 1x4x4	
6	1	Divizarea canalelor	Etapă 5: 1x2x2	Etapă 5: 512	
			Etapă 6: 1x1x1	Etapă 6: 1024	
				După etapa 2: 8	
				După etapa 3: 32	
				După etapa 4: 128	
				După etapa 5: 256	

de comprimare.

După o astfel de redimensionare, se formează o regiune de 1x1x1 pixeli dintr-un fragment de 2x2x2 pixeli care au fost aplatizați spațial către dimensiunea canalelor. Prin urmare, FoV se dublează pe fiecare axă spațială pentru fiecare etapă de comprimare. Acest lucru permite convoluțiilor 1x1 să

opereze pe un FoV din ce în ce mai mare, similar cu funcționarea unui CNN clasic, păstrând în același timp capacitatea de calcul eficient al NF pentru transformările înainte/înapoi. Există suficiente operații de comprimare astfel încât rezoluția în ultima etapă să scadă la 1x1x1. Firește, se restricționează dimensiunile spațiale de intrare pentru a fi puteri ale lui 2.

În toate experimentele efectuate, ponderile rețelei sunt inițializate astfel încât straturile să fie o funcție identitate la începutul antrenării, așa cum se propune în [24]. S-a utilizat platforma PyTorch DL cu optimizatorul Adam cu o rată de învățare de $1e-4$ iar modelele au fost antrenate până când funcția criteriu pe setul de validare s-a stabilizat.

4.2.2 Perturbații Sintetice ale Măștii

Scopul aplicației este de a detecta perechile incorecte de (imagine angiografie, mască lumen), adică eșantioane în care segmentarea nu este în deplină concordanță cu imaginea. Pentru a testa modelele dezvoltate, a fost concepută o metodă de obținere a exemplurilor "greșite" (adică a eșantioanelor care nu se află în distribuția perechilor imagine-mască "corecte") pornind de la datele inițiale (considerate a fi "corecte").

S-au mărit seturile de date prin aplicarea unor perturbații prestabilite asupra măștii binare de segmentare a lumenului, păstrând în același timp imaginea angiografică neatinsă. Au fost utilizate trei tipuri de perturbări ale măștii:

- **zoom:** s-au aplicat operații de mărire/micșorare a măștii în raport cu centrul acesteia, astfel încât masca rezultată să fie în continuare aliniată cu angiografia, dar mai mare/mică decât înainte.
- **morfologie:** s-au aplicat dilatări sau eroziuni de-a lungul a 4 direcții în planul înălțime*lățime: stânga-dreapta, sus-jos, susStânga-josDreapta și susDreapta-josStânga. Această perturbare afectează doar o parte a măștii (partea erodată sau dilatăată), în timp ce cealaltă parte rămâne neatinsă. Prin convenție, nivelurile negative și pozitive se referă la cele două sensuri în direcția selectată, zero însemnând poziția originală a măștii (nivelurile sunt exprimate ca raporturi ale dimensiunii originale a măștii de-a lungul direcției alese). La fiecare nivel, se poate aplica fie dilatarea (rezultând măști prelungite), fie eroziunea (rezultând măști scurtate).
- **translații:** în aceleași 4 direcții pe planul înălțime*lățime, s-au translat imagini întregi ale măștii. Fiecare creștere de nivel înseamnă o deplasare de pixeli.

Pentru fiecare arhitectură de rețea, s-au efectuat două proceduri de antrenare: una care utilizează doar datele originale (neperturbate) și una care utilizează un set de date format din datele originale și versiunea perturbată a acestora. Perturbațiile sunt aplicate în timpul antrenării, similar tehnicilor de augmentare a datelor, astfel încât fiecare eșantion din datele originale este perturbat pe toate tipurile, nivelurile și direcțiile de perturbare pe parcursul epocilor de antrenare. La fiecare epocă, raportul dintre datele neatinsă și cele perturbate este de 1 la 1.

Antrenarea doar pe datele originale și apoi testarea pe perturbații sintetice oferă o imagine clară a capacității de detectare a OoD, care provine pur și simplu din structura inductivă a arhitecturii NF. De asemenea, considerăm că modelele NF, fiind o clasă de modele generative, oferă o formă de explicabilitate prin capacitatea de a produce eșantioane din densitatea de probabilitate învățată. Prin eșantionarea repetată din model și prin calcularea log-probabilităților asociate, se poate observa tipul de eșantioane pe care modelul le consideră ca fiind în distribuție.

4.3 Rezultate și Discuție

4.3.1 Evaluarea pe Perturbații Sintetice ale Măștii

În primă fază, s-au evaluat modelul de referință și cel propus după antrenarea pe date originale (neperturbate). S-au aplicat perturbațiile sintetice pe setul de test în niveluri crescânde de severitate

și s-a măsurat cât de bine pot distinge modelele între scorurile eșantioanelor originale și cele ale eșantioanelor perturbate. La fiecare nivel de perturbare, s-a calculat aria de sub curba RoC. Utilizăm AuRoC ca o măsură pentru a evalua cât de bine pot fi separate două distribuții de date individuale prin utilizarea unui prag de probabilitate.

Se poate observa că modelul propus are performanțe superioare pentru toate tipurile și nivelurile de perturbații. O transformare sub 1.0x generează de fapt valori mai mari ale scorurilor pentru modelul de bază, ceea ce duce la valori AuRoC sub 0.5. Chiar și la niveluri mici de perturbare a măștii (de exemplu, zoom de 0.9x/1.1x, ± 2 pixeli de translație), modelul propus are o sensibilitate mult mai mare în detectarea modificărilor măștii (chiar dacă nu a fost antrenat în mod explicit pentru a face acest lucru), spre deosebire de modelul de bază, unde scorurile încep să scadă mai semnificativ doar la niveluri de perturbare mai mari. Morfologia măștii este cea mai grea de detectat, deoarece o parte a măștii rămâne neschimbată. Astfel, modelul de bază este în mare măsură insensibil la acest tip de perturbare, deoarece AuRoC maxim la un nivel de perturbare ridicat de 60% este sub 0.65. În comparație, AuRoC pentru modelul propus are o variație mult mai rapidă în cazul creșterii severității perturbației, atingând valori de peste 0.9 pentru unele direcții la o severitate de morfologie de 60%.

Pentru a obține un semnal de log-probabilitate care descrie un întreg segment de vas, a fost utilizată o abordare de tip fereastră glisantă în care grupuri suprapuse formate din 8 secțiuni transversale adiacente sunt introduse în modelul NF pentru a obține valorile log-prob. pentru fiecare grup. Utilizând această procedură, secțiunile transversale din mijloc pot participa la cel mult 8 grupuri, prin urmare pot exista până la 8 valori de log-probabilitate prezise legate de fiecare secțiune transversală din mijloc. Se utilizează o schemă de vot bazată pe medie, în care valoarea log-prob. finală pentru fiecare secțiune transversală este calculată prin medierea scorurilor atașate. Fig. 4.3 prezintă un astfel de exemplu, în care se aplică o perturbare sintetică cu o variație cunoscută a severității. Modelul NF propus detectează atunci când severitatea perturbației este suficient de mare, în timp ce produce valori ridicate ale log-prob. atunci când perturbația este neglijabilă.

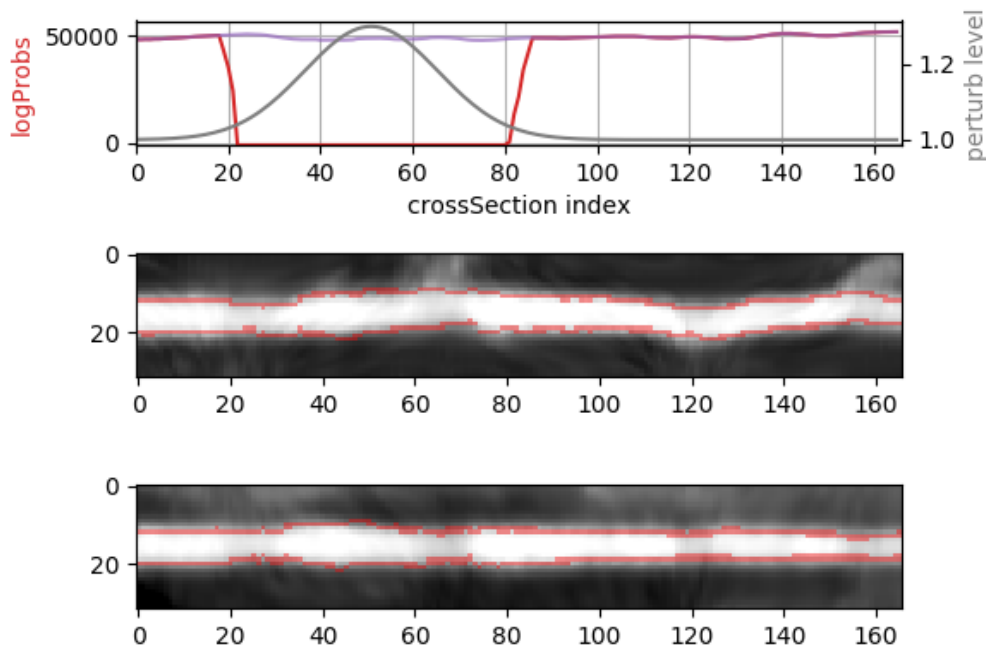


Figura 4.3: Predicția întregului segment de vas folosind o abordare cu ferestre glisante și arhitectura propusă. Se aplică o perturbare de zoom cu o variație cunoscută de severitate (graficul de sus, semnal gri) (se reține că nivelul de zoom 1.0x este o transformare identitate). Semnalul log-prob. rezultat (graficul de sus, semnal roșu) scade ori de câte ori perturbarea este suficient de severă, în comparație cu semnalul log-prob. original în absența oricărei perturbari (graficul de sus, semnal violet). Cele 2 diagrame din partea de jos afișează două vederi laterale ale segmentului de vas (2 proiecții pe axe diferite), cu conturul măștii perturbate suprapus.

Tabela 4.3: Metrici pe setul secundar de date pentru modelul de bază și cel propus. Modelul propus depășește în mod constant performanța modelului de referință și are valori metrice apropiate de consensul inter-experti.

Metrică	Consens Inter-Expert Medie [Min, Max]	Modelul de bază	Modelul propus
Acuratețe	0.81 [0.79, 0.86]	0.64	0.79
Sensibilitate	0.79 [0.70, 0.87]	0.48	0.76
Specificitate	0.83 [0.76, 0.90]	0.77	0.81
PPV	0.79 [0.70, 0.87]	0.63	0.76
NPV	0.83 [0.76, 0.90]	0.65	0.81

4.3.2 Evaluare pe baza Adnotărilor Experților

Un set de test selectat a fost adnotat manual și independent de către trei experți la nivel de leziune: fiecare leziune a fost marcată ca fiind "corect" sau "incorect adnotată", pe baza următoarelor instrucțiuni: o leziune trebuie marcată ca fiind "incorect adnotată" dacă *cel puțin un* contur al vreunei secțiunii transversale necesită editare, în caz contrar, dacă *niciun* contur al secțiunii transversale nu necesită editare, atunci leziunea trebuie marcată ca fiind "corect adnotată". Instrucțiunile pentru adnotatori au fost concepute astfel încât să corespundă procedurii utilizate pentru a construi acest set de testare separat, cu scopul de a putea compara direct predicțiile din modelul NF cu cele furnizate de către adnotatori.

Evaluarea celor două modele NF în raport cu adnotările umane a fost formulată ca o problemă de clasificare binară. Tab. 4.3 rezumă metricile relevante (acuratețe, sensibilitate, specificitate, PPV și NPV) pentru modelul propus și cel de bază. Consensul de adnotare a fost obținut printr-un vot majoritar între cei trei adnotatori. Valorile medii ale metricilor între utilizatori au fost obținute prin calcularea mediei tuturor celor 6 valori posibile ale metricilor aferente perechilor de adnotatori, de exemplu, Adnotator_1 (ca GT) versus Adnotator_3 (ca predicție), Adnotator_3 (ca GT) versus Adnotator_1 (ca predicție) etc. Atunci când este comparat cu consensul de adnotare, modelul propus are performanțe mai mari decât cel de bază, pe toate metricile considerate.

Se observă că modelul propus are o sensibilitate de 76.0%, aproape de valoarea inter-utilizator de 79.0%, în timp ce modelul de bază atinge doar 48%. De asemenea, scorul general de acuratețe crește la 78.6% (aproape de valoarea inter-utilizator de 80.9%) pentru modelul propus, în comparație cu o valoare de acuratețe de 64.3% pentru modelul de bază de tip Glow. Aceste rezultate întăresc observația conform căreia modelul de bază nu este capabil să capteze pe deplin conținutul semantic, în timp ce modelul propus reușește. Un comportament similar a fost observat în secțiunea anterioară, unde modelul propus a avut valori AuRoC mai bune în detectarea perturbațiilor sintetice atunci când a fost antrenat doar pe datele originale.

4.3.3 Eșantionarea folosind Modelele Antrenate

Au fost utilizate modelele antrenate pe setul de antrenare augmentat pentru a genera eșantioane noi. Similar cu procedurile de eșantionare din [24], s-a folosit $\mathcal{N}(\mathbf{0}, 0.6 \cdot I)$ în locul distribuției anterioare reale (adică distribuția standard normală multivariată) pentru a produce eșantioane cu log-probabilități mai mari și care să pară mai realiste. Fiecare eșantion nou a fost re-rulat prin model în direcția înainte pentru a calcula log-probabilitățile, confirmând că eșantionul este de fapt în distribuție (procedura de eșantionare poate genera rareori eșantioane cu o log-probabilitate mai mică). Fig. 4.4 prezintă eșantioane din cele două modele.

După cum s-a arătat deja în [25], modelul de bază tinde să se concentreze mai mult pe texturi și nu este capabil să capteze semantica datelor de antrenare. S-a observat că, în majoritatea eșantioanelor generate, masca de segmentare lipsește (adică sunt generate doar zerouri pe canalul măștii). De asemenea, forma rotundă obișnuită a lumenului nu se poate distinge în canalul de imagine. În schimb,

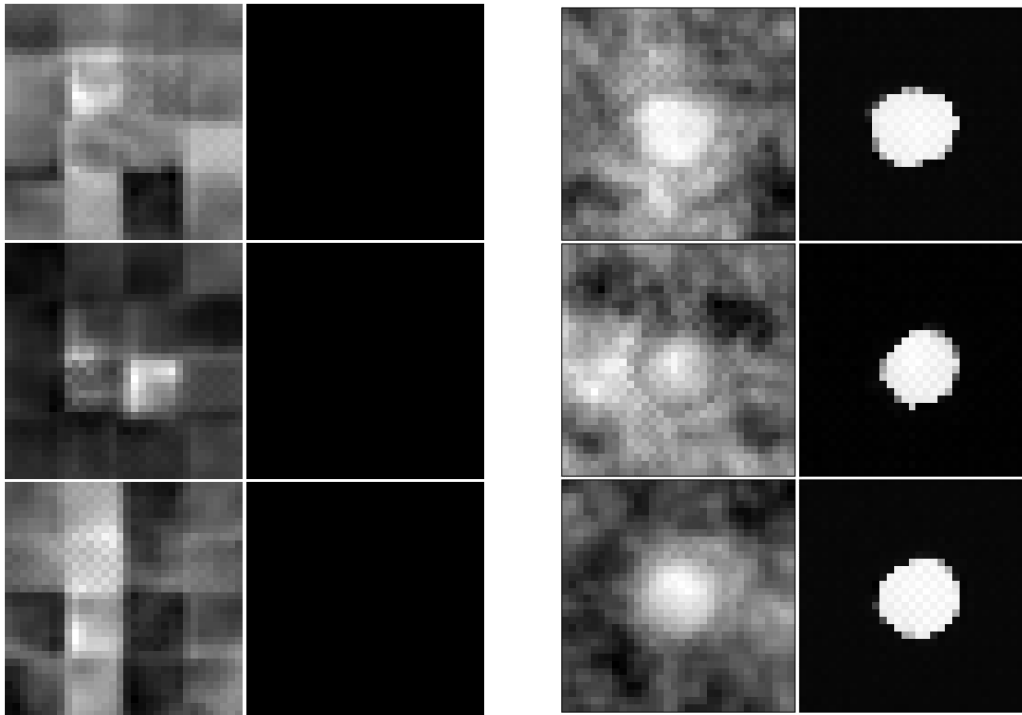


Figura 4.4: Trei eșantioane (perechi de imagine a lumenului și mască de segmentare) generate de modelul propus (dreapta) și de cel de bază (stânga).

arhitectura propusă reușește să capteze conținutul semantic al unui exemplu obișnuit: lumenul are forma tipică în canalul de imagine, masca de segmentare este prezentă (cu valori de pixel plauzibile, de exemplu aproape de 0 sau 1) și respectă forma și poziția lumenului din canalul de imagine.

Considerăm că inspectarea eșantioanelor generate este un mecanism de explicabilitate care oferă o perspectivă asupra distribuției de probabilitate învățate, adică modelul poate furniza exemple de intrări care sunt foarte probabile în cadrul densității învățate și, prin repetarea procedurii de eșantionare de suficiente ori, se poate construi o aproximare a setului tipic al distribuției învățate. În cazul în care un model generativ produce în mod constant eșantioane cu probabilități logaritmice asociate ridicate, dar care au o calitate scăzută în urma unei inspecții manuale și care nu sunt plauzibile având în vedere subiectul/domeniul specific, atunci aceasta este o dovadă că densitatea de probabilitate învățată nu este o bună aproximare a adevăratei densități de probabilitate și, prin urmare, modelul nu poate fi utilizat în mod fiabil pentru detectarea intrărilor aberante (OoD).

4.4 Concluzii

În timp ce performanța metodelor bazate pe inteligență artificială s-a îmbunătățit considerabil de-a lungul anilor, abordările semiautomate sunt încă folosite în prezent. O abordare potențială pentru reducerea semnificativă a timpului de procesare este de a preselecta regiunile de interes care probabil vor necesita inspecție și editare manuală. În acest studiu, s-a legat această etapă de preselectie de subiectul incertitudinii și al detectării intrărilor din afara distribuției, pe baza NF. Utilizarea unui nou strat de cuplare care prezintă o structură inductivă care favorizează exploatarea caracteristicilor semantice în locul corelațiilor locale ale pixelilor a fost investigată în sarcina de a detecta perechile de imagini CCTA ale lumenului și segmentările necorespunzătoare acestora. O arhitectură de rețea care utilizează astfel de straturi a fost testată în comparație cu una de referință de tip Glow. Rețeaua propusă a arătat o performanță mai bună în detectarea OoD atunci când a fost testată împotriva perturbațiilor sintetice, în timp ce sensibilitatea de detectare a adnotărilor greșite a fost apropiată de consensul inter-experti. Eșantioanele din model confirmă faptul că densitatea de probabilitate

învățată a reușit să capteze conținutul informațional relevant din exemplele de antrenare, în loc să modeleze doar texturi simple.

Metoda propusă în această lucrare permite luarea unei decizii mai de încredere folosind doar imagistica CCTA. Utilizând metoda propusă de detectare a exemplilor din afara domeniului, zona gri din interpretarea clinică poate fi potențial redusă.

5. Detectarea Fazelor Cardiace în Angiografiile Coronariene Invazive¹

5.1 Introducere

Angiografia coronariană invazivă (ICA) reprezintă standardul de aur în imagistica CAD [31], oferind informații importante despre structura și funcția inimii (anual, peste un milion de proceduri ICA sunt efectuate numai în SUA [32]). Ea permite evaluarea severității anatomice a stenozelor coronariene fie vizual, fie prin angiografie coronariană cantitativă asistată de calculator (QCA) [33]. Având în vedere limitările evaluării anatomice pure a CAD și având în vedere recentele progrese tehnologice, au fost introduse și validate metode de evaluare funcțională a stenozelor coronariene bazate pe imagini și pe ICA, de exemplu, calculul rezervei de debit fracțional (FFR) bazat pe imagini [34, 35, 36]. În acest context clinic și în alte contexte clinice bazate pe utilizarea ICA, detectarea cadrelor de sfârșit de diastolă (EDF) și, în general, detectarea fazei cardiace pe fiecare cadru temporal al unei achiziții coronarografice este de o importanță semnificativă.

În prezent, selectarea EDF și identificarea fazei cardiace se realizează fie manual, fie automat, pe baza semnalelor ECG achiziționate simultan [35]. Acest lucru prezintă o serie de dezavantaje: este posibil ca semnalele ECG să nu fie întotdeauna disponibile, iar detectarea fazei cardiace pe baza ECG poate fi dificilă dacă raportul semnal-zgomot este prea mic pentru a detecta cu precizie sfârșitul diastolei sau dacă semnalul prezintă artefacte [37], [38].

Principala provocare abordată în cadrul acestei lucrări este dezvoltarea unei metode bazate exclusiv pe imagine pentru detectarea fazei cardiace la nivel de cadru și detectarea EDF pe angiografiile coronariene invazive.

În plus, s-au abordat două provocări suplimentare legate de cea principală. Prima a fost aceea de a demonstra că metodologia introdusă în prezentul document oferă rezultate la fel de bune în toate tipurile de angiografii coronariene, cu variații în ceea ce privește vederea (artera coronară stângă / dreaptă), unghiurile de achiziție primară și secundară, frecvența cardiacă, precum și tipul și starea pacientului (stabil / acut, intervenție coronariană percutanată (PCI) anterioară, grefă de bypass coronarian (CABG) anterioară, ocluzii coronariene totale etc.). A doua provocare suplimentară a fost valorificarea setului de date foarte mare (77435 de imagini de angiografii coronariene achiziționate de la 13081 de pacienți) fără a efectua adnotări manuale. Exploatarea unui set de date atât de mare este benefică din cauza variabilității extreme a angiografiei coronariene cu raze X, dar nu este fezabilă adnotarea manuală a tuturor datelor.

Prin urmare, se prezintă aici o metodologie bazată pe rețele neuronale adânci (DNN) pentru detectarea fazelor cardiace (sistole sau diastole) și a EDF pe imagini coronariene cu raze X. Ieșirile țintă, utilizate pentru antrenarea modelului, sunt derivate din semnalele ECG achiziționate simultan.

¹Următorul capitol descrie experimente și analize publicate în:

- **Ciușdel, C.**, et al., 2020. Deep Neural Networks for ECG-free Cardiac Phase and End-Diastolic Frame Detection on Coronary Angiographies. *Comput. Med. Imaging Graph.* 84, 101749, <https://doi.org/10.1016/j.compedimag.2020.101749>

Unele secțiuni au fost citate textual din referința de mai sus, care reprezintă o lucrare publicată anterior de către autor, în cadrul programului de cercetare doctorală.

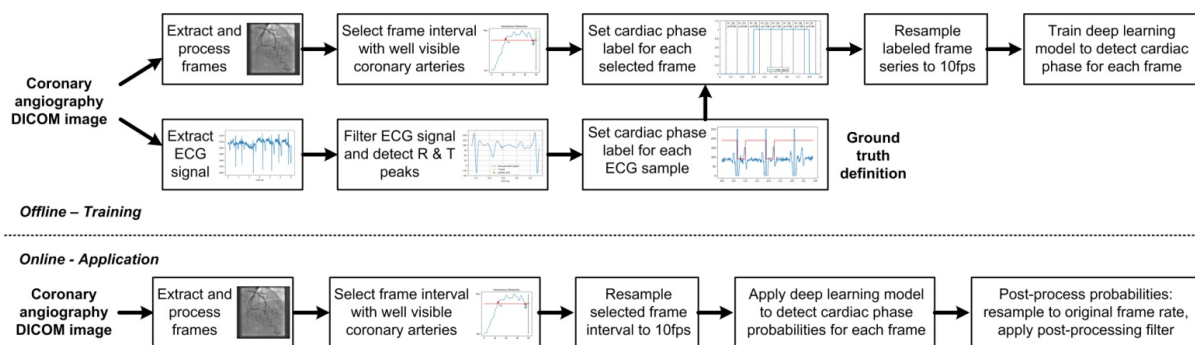


Figura 5.1: Fluxul general de lucru al metodologiei propuse pentru antrenarea offline și aplicarea online.

5.2 Metode

Arterele coronare prezintă o mișcare semnificativă pe imaginile angiografice în timpul unui ciclu cardiac. Mișcarea este efectul combinat al contracției cardiace, al mișcării respiratorii și, eventual, al deplasării pacientului sau a mesei, apărând de obicei sub formă de panoramare în imaginile angiografice. Obiectivul principal al fluxului de lucru propus este de a utiliza tehnici bazate pe învățare profundă pentru a determina faza cardiacă a fiecărui cadru angiografic prin analiza implicită a mișcării arterelor și a structurilor vizibile pe cadre angiografice consecutive și fără a utiliza informații ECG asociate.

Fig. 5.1 afișează fluxul de lucru general al metodologiei propuse pentru detectarea fazei cardiace, inclusiv antrenarea offline și aplicarea online. Pe scurt, pentru aplicarea online, un prim model, antrenat pentru a detecta arterele coronare, este utilizat pentru a preseleciona un subset de cadre în care arterele coronare sunt clar vizibile. Un al doilea model prezice probabilitățile de fază cardiacă pentru fiecare cadru: acesta ia ca intrare secvențe de câte 10 cadre din intervalul preselectat, efectuează convoluții spațiale și temporale și emite predicții pentru cele 4 cadre din mijloc ale secvenței. Cel de-al doilea DNN este aplicat cu o abordare de tip fereastră glisantă. Numai în fazele de antrenare și evaluare pentru cel de-al doilea DNN, semnalele ECG înregistrate simultan cu imaginile angiografice sunt utilizate pentru a furniza valori țintă pentru ieșiri: Vârful unde R și ale unde T, utilizate pentru a determina debutul și, respectiv, sfârșitul sistolei, sunt detectate algoritmic, iar etichetele corespunzătoare fazei cardiace sunt propagate fiecărui cadru angiografic component. Etapa preliminară de procesare a secvențelor angiografice pentru a extrage cadrele în care vasele coronariene sunt clar vizibile se aplică, de asemenea, atunci când se construiește baza de date de antrenare pentru metoda de învățare profundă care determină faza cardiacă a fiecărui cadru. Deși semnalul ECG este de obicei disponibil pentru întreaga achiziție angiografică, s-a ales să se detecteze faza cardiacă doar pentru acele cadre în care arterele coronare sunt bine vizibile, deoarece acestea sunt cadrele care sunt relevante în procesul de luare a deciziilor clinice. În plus, permițând modelului de detectare a fazei cardiace să învețe doar pe baza cadrelor cu artere bine vizibile, se presupune că se obține un model mai specializat și, astfel, mai performant.

Principala contribuție metodologică în această lucrare este abordarea pentru antrenarea offline și aplicarea online (vezi Fig. 5.1).

- Sunt selectate automat intervale de cadre angiografice coronariene relevante din punct de vedere clinic. Acest lucru are un impact pozitiv atât asupra procesului de antrenare (permițând modelului de predicție a fazei cardiace să se specializeze pe cadrele relevante, cu artere bine vizibile, așadar precizia predicției sale crește pe aceste cadre), cât și pentru aplicația online (de exemplu, vor fi selectate numai cadrele relevante pentru EDF).
- În fluxul de lucru nu sunt incluse etape intermediare sau sarcini de detectare, localizare sau segmentare a structurilor anatomiche. În plus, nu este necesară definirea vreunei măsuri de

substituție. Modelele iau direct imaginile ca intrare și dau ca rezultat vasele detectate / faza cardiacă.

- Valorile țintă de ieșire sunt derivate automat din semnalele ECG

5.2.1 Procesul Offline de Antrenare

5.2.1.1 Preprocesarea Semnalelor ECG și a Cadrelor Angiografice

Ca date de intrare sunt utilizate achiziții coronariene angiografice anonimizate în format DICOM. În etapa de preprocesare, fiecare fișier DICOM este analizat pentru a extrage datele imagistice și semnalul ECG brut utilizat pentru definirea ieșirilor țintă (Fig. 5.2 stânga). Tranzițiile diastolă-sistolă și diastolă-sistolă sunt detectate pentru a defini ieșirile țintă. Pentru prima, se utilizează BioSPPy, o soluție software cu sursă deschisă pentru procesarea semnalelor biologice, [39]. În primul rând, BioSPPy aplică o etapă implicită de procesare semnalului ECG original, bazat pe un filtru liniar cu răspuns finit la impuls, cu întârziere de fază zero, folosind două treceri. Apoi, vârfurile R sunt detectate ca surogat pentru tranziția diastolă-sistolă (Fig. 5.2 mijloc). Condiția prealabilă pentru identificarea tranziției diastolă-sistolă este detectarea vârfului T. Oscilațiile de înaltă frecvență sunt de obicei încă prezente după prima etapă de filtrare, împiedicând detectarea precisă a vârfului T. Prin urmare, se aplică o altă etapă de filtrare trece-jos cu zero întârziere de fază. În cele din urmă, vârfurile T sunt identificate în conformitate cu următoarele reguli (Fig. 5.2 dreapta):

- Momentul de vârf T trebuie să se situeze între două limite predefinite, exprimate în raport cu durata bătăilor inimii: în acest caz, au fost selectate limitele de 20% și 65% [40]
- Vârful T trebuie să fie un maxim sau un minim local al semnalului ECG filtrat.
- Vârful T trebuie să fie maximul sau minimul local cu cea mai mare întindere temporală, în raport cu punctele de extrem învecinate, în interiorul ferestrei considerate.

În cele din urmă, punctul de timp corespunzător tranziției diastolă-sistolă este definit ca fiind primul minim sau maxim local după vârful T, sau ca fiind punctul de timp în care semnalul filtrat scade sau crește sub valoarea medie a ferestrei de 20%-65%, oricare dintre aceste momente de timp este întâlnit primul [40]. Se generează apoi un semnal binar de clasificare pe baza punctelor de timp detectate: "0" pentru sistole și "1" pentru diastole.

Scopul metodei propuse este de a prezice corect etichetele de fază cardiacă și de a detecta EDF pentru cadrele angiografice în care arterele coronare sunt bine vizibile, adică pline de substanță de contrast, deoarece acestea corespund setului de cadre relevante din punct de vedere clinic pentru aplicațiile avute în vedere. Pentru fiecare cadru, vizibilitatea arterelor coronare este determinată cu

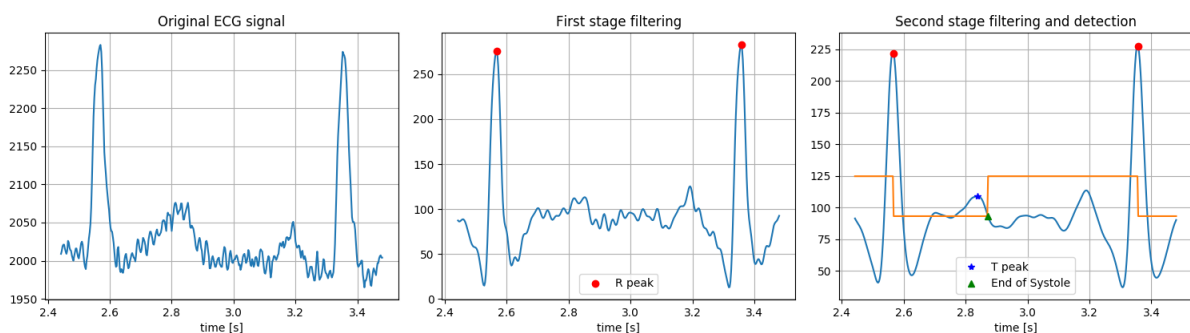


Figura 5.2: Prelucrarea semnalelor ECG: semnalul ECG original (stânga); prima etapă de filtrare, folosind filtru liniar trece-bandă cu zero întârziere de fază, și detectarea vârfului R (centru); a doua etapă de filtrare, folosind filtru trece-jos cu zero întârziere de fază și detectarea vârfului T și a sfârșitului sistolei (dreapta).

ajutorul unui clasificator de învățare profundă antrenat să detecteze pixelii care reprezintă arterele coronare, adică să estimeze "gradul de vascularitate" al imaginii. Toate cadrele cu o vizibilitate a vaselor mai mare decât un anumit prag, definit în raport cu numărul maxim de pixeli per cadru, sunt marcate ca fiind candidate; cea mai lungă secvență continuă de cadre candidate este selectată pentru prelucrarea ulterioară.

În plus, pentru a dezvolta un model de detecție a fazelor cardiace care este independent de frecvența cadrelor din achiziție, seria cadrelor de intrare este reeșantionată la 10 fps.

5.2.1.2 Arhitectura Modelului de Segmentare a Vaselor și Procesul de Antrenare

Un model bazat pe învățare profundă este utilizat pentru a determina intervalul de cadre angiografice în care arterele coronare sunt bine vizibile, adică umplute cu agent de contrast. Clasificatorul este o rețea neuronală adâncă având o arhitectură similară cu U-Net [41], antrenată pentru a clasifica pixelii folosind optimizatorul Adam [42] prin minimizarea unei funcții criteriu personalizate, derivată din indicele Jaccard:

$$L = 1 - \frac{\mu + \sum P_i T_i}{\mu + \sum P_i^2 + \sum T_i^2 - \sum P_i T_i}$$

unde P_i și T_i sunt probabilitatea prezisă și cea țintă ca al i -lea pixel să facă parte dintr-o arteră coronariană, iar $\mu = 0.1$ este un factor de netezire.

Datele de antrenare sunt cadre angiografice în care vasele coronariene sunt adnotate manual. Fiecare arteră este adnotată ca un set de puncte centrale, iar fiecărui punct îi este asociată o estimare aproximativă a razei locale a vasului. Utilizând aceste adnotări, se generează o mască binară pentru fiecare cadru angiografic, în care pixelii care se găsesc în vecinătatea punctelor liniei centrale sunt setați la unu; toți ceilalți pixeli sunt setați la zero.

În timpul inferenței, modelul procesează independent fiecare cadru angiografic și emite o hartă de probabilitate de aceeași dimensiune ca și imaginea de intrare, unde valoarea fiecărui pixel reprezintă probabilitatea ca locația asociată să facă parte dintr-o arteră coronariană. Prin însumarea tuturor probabilităților pixelilor, se determină un scor global de vascularitate al cadrului, adică probabilitatea cadrului de a conține vase coronariene.

Pentru fiecare cadru angiografic din serie, se determină scorul de vascularitate. Cadrele care au un scor mai mic de 40% din scorul maxim de vascularitate din serie sunt eliminate. Din cadrele rămase, se selectează cea mai lungă sub-secvență de cadre consecutive pentru procesarea ulterioară.

5.2.1.3 Arhitectura Modelului de Detecție a Fazelor Cardiace și Procesul de Antrenare

Predictorul de fază cardiacă este o rețea neuronală adâncă care ia ca intrare o secvență de cadre angiografice (Fig. 5.3). Ieșirea rețelei este clasificarea cadrelor din mijlocul secvenței de intrare, în funcție de probabilitatea fiecărui cadru de a fi diastolic sau sistolic. Deoarece se are în vedere o problemă de clasificare binară, clasele diastolice și sistolice fiind asociate cu indicele de clasă 1 și, respectiv, 0, probabilitatea de ieșire poate fi interpretată ca indicele de clasă prezis. S-a proiectat rețeaua pentru a procesa o secvență de 10 cadre, acoperind exact o secundă de achiziție. Această alegere a fost făcută ca un compromis între precizia detecției și necesarul de memorie în timpul procesului de antrenare. Rețeaua emite clasificarea numai pentru cele 4 cadre din mijlocul secvenței, astfel încât, pentru fiecare cadru, sarcina de clasificare utilizează informații atât din cadrele anterioare, cât și din cele următoare.

După cum se arată în figura 5.3, o primă rețea neuronală convoluțională (CNN) este utilizată pentru a efectua convoluții spațiale, mapând fiecare imagine de intrare de dimensiune 512 x 512 într-un vector de caracteristici 64-dimensional. CNN-ul spațial utilizează o structură clasică de straturi convoluționale 2D, urmate de straturi de maximizare și, în final, de un strat complet conectat. Același CNN este aplicat în mod independent pe fiecare cadru, obținându-se o matrice de caracteristici de mărime 10 x 64. Un al doilea CNN efectuează convoluții temporale pe această matrice, independent pentru toate cele 64 de caracteristici ale imaginii. Rezultatul său este o matrice de 64 x 8, care conține 8 caracteristici temporale pentru fiecare dintre cele 64 de caracteristici spațiale. Această matrice

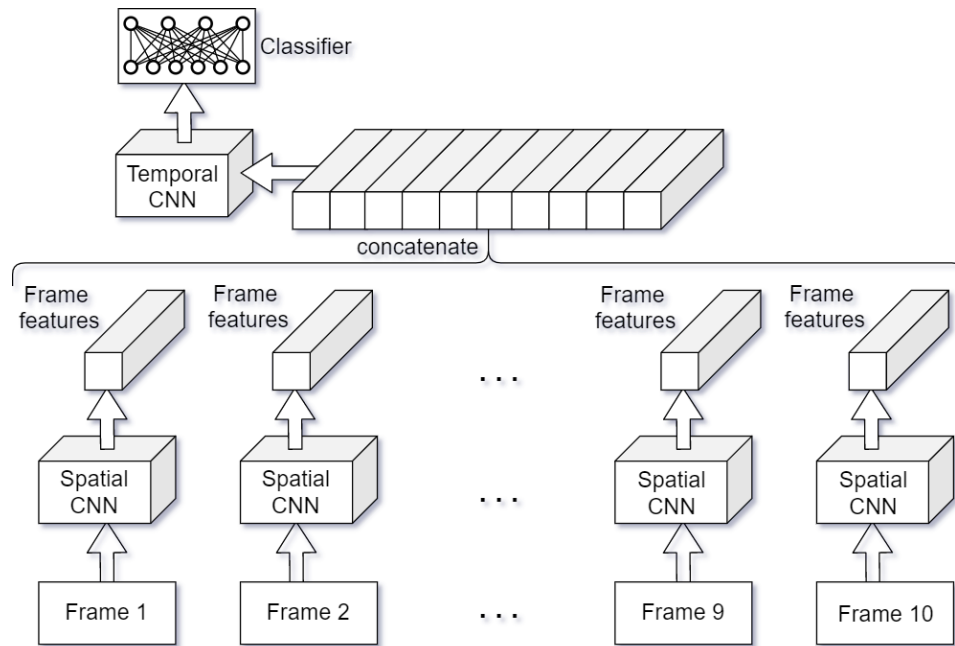


Figura 5.3: Prezentare generală a rețelei neuronale adânci utilizate pentru detectarea fazei cardiace.

este apoi transmisă unui clasificator dens, care calculează probabilitățile de ieșire pentru cele 4 cadre din mijlocul secvenței de intrare (cadrele 4-7).

După selectarea cadrelor relevante cu ajutorul modelului de vascularitate și reeșantionarea la 10 fps, seturile de date de antrenare sunt generate ca secvențe de câte 10 cadre. Ieșirile țintă pentru cele 4 cadre din mijlocul fiecărei secvențe sunt obținute prin procesarea semnalului ECG asociat, așa cum este descris în secțiunea anterioară.

Antrenarea se realizează utilizând optimizatorul Adam [42] și o funcție criteriu personalizată, bazată pe divergența distribuțiilor Poisson [43] care penalizează probabilitatea de ieșire mai puțin sever dacă clasa este prezisă corect în raport cu valoarea de prag de 0.5.

5.2.2 Procesul Online al Aplicației

În timpul aplicării online a modelului antrenat pentru predicția fazelor cardiace, nu se utilizează informații ECG. Cadrele angiografice sunt preprocesate cu aceleași metode utilizate pentru popularea bazei de date de antrenare: sunt decupate și redimensionate, iar estimatorul de vascularitate bazat pe DL este utilizat pentru a selecta intervalul de cadre relevant. În continuare, cadrele din intervalul ales sunt reeșantionate la 10 fps și sunt furnizate ca intrare în predictorul de fază cardiacă.

Pentru a minimiza numărul de operații redundante, mai întâi se aplică CNN-ul spațial pe toate cadrele de intrare, iar reprezentările calculate sunt stocate pentru utilizarea ulterioară. În continuare, CNN-ul temporal și clasificatorul de ieșire sunt aplicate cu o abordare de tip fereastră glisantă, cu un pas de 1, folosind ca intrare reprezentările pre-calculate ale cadrelor. Deoarece modelul generează probabilități pentru cele patru cadre din mijloc, mai multe cadre ajung să aibă mai multe predicții, iar pentru predicțiile finale ale fazei cardiace se iau în considerare doar cadrele pentru care au fost generate cel puțin două probabilități. Predicția finală este selectată ca fiind valoarea probabilității cea mai apropiată de unul dintre indicii de clasă. Într-o etapă suplimentară de postprocesare, secvența rezultată de valori de probabilitate este reeșantionată la rata originală a cadrelor utilizând interpolarea liniară. Rezultatul este un semnal binar de clasificare, în care "0" corespunde sistolei și "1" diastolei.

5.3 Rezultate

5.3.1 Detectația Vaselor

Modelul care efectuează detectarea vaselor a fost antrenat pe 1174 de angiografii coronariene și validat pe 293. Adnotările au fost efectuate pentru cinci cadre consecutive pentru fiecare angiografie, rezultând 5870 de cadre de antrenare și 1465 de cadre de validare. În plus, s-a efectuat o augmentare a datelor pe setul de antrenare, constând în zoom aleator cu un factor între 0.8 și 1.2, modificări gaussiene de intensitate cu media 0 și abaterea standard 0.1 și zgomot gaussian cu media 0 și abaterea standard 0.2. Pe un set separat de test, format din cadre extrase din 91 de angiografii coronariene (adică 455 de cadre în total), modelul a obținut un scor Dice mediu de 0.86. Seturile de antrenare, validare și testare conțin achiziții care descriu atât coronare drepte (RCA), cât și stângi (LCA), din diferite unghiuri de achiziție.

5.3.2 Detectația Fazelor Cardiace

Fluxul de lucru pentru detectarea fazei cardiace a fost antrenat și evaluat utilizând seturi de date independente. Setul de date de antrenare a constat din 56655 de angiografii coronariene obținute de la 6820 de pacienți. O împărțire 90 – 10 % a fost aplicată la nivel de pacient în timpul antrenării pentru stabilirea seturilor de date finale de antrenare și validare.

Setul principal de date de evaluare a constat din 20780 de angiografii coronariene de la 6261 pacienți și a fost preluat de la un alt centru clinic. Acest set de date complet independent a fost utilizat pentru evaluarea cantitativă a performanțelor fluxului de lucru pentru detectarea fazelor cardiace și a cadrului final-diastolă și pentru efectuarea diferitelor analize pe subgrupuri pe baza unghiurilor de achiziție, a ritmului cardiac și a vasului de interes. În plus, semnalele ECG ale unui subset de angiografii coronariene, extrase din setul principal de date de evaluare, au fost adnotate de examinatori clinici experimentați, iar performanța detectării fazei cardiace a fost evaluată separat pe acest subset.

Nu au fost formulate criteriile de excludere (în setul de date de antrenare și în setul principal de date de evaluare) legate de unghiurile de achiziție, frecvența cardiacă, intensitatea agentului de contrast, caracteristicile și diagnosticul pacientului.

5.3.2.1 Performanța de Detectare a Fazelor Cardiace pe Setul Principal de Date de Evaluare

Pentru evaluarea cantitativă a performanțelor fluxului de lucru de detectare a fazei cardiace, am exclus cadrele cu valori intermediare ale ieșirilor țintă și se prezintă în cele ce urmează rezultatele pentru cadrele rămase, precum și pentru cazul în care sunt excluse și cadrele adiacente unei tranziții de fază cardiacă.

Tab. 5.1 afișează, pentru aceste două configurații, măsurile statistice pentru detectarea fazei cardiace pe setul de date de evaluare, în două setări: (a) toate angiografiile coronariene au aceeași pondere, adică statisticile sunt mai întâi calculate independent pentru fiecare angiografie coronariană, iar apoi se calculează o valoare medie pe întreg setul și (b) toate cadrele au aceeași pondere.

Pentru a evalua performanța detecției bazate pe imagine în mod independent pentru cele două tranziții de fază cardiacă, de la sistole la diastole și de la diastole la sistole, s-au determinat erorile de clasificare a fazei cardiace la nivel de cadru în jurul fiecărui tip de tranziție pentru angiografiile coronariene cu semnale ECG adnotate. S-a constatat că 65.0% / 65.2% din erori au corespuns tranziției sistole-diastole, iar 35.0% / 34.8% din erori au corespuns tranziției diastole-sistole.

5.3.2.1.1 Performanța de Detectare a Fazei Cardiace în Funcție de Vasul de Interes Angiografiile coronariene sunt înregistrate fie pentru artera coronară stângă (LCA), fie pentru artera coronară dreaptă (RCA). Cu ajutorul unui clasificator bazat pe DL (98.0% de acuratețe) s-a determinat artera coronariană vizibilă în fiecare angiografie și s-a calculat acuratețea medie pentru fiecare clasă: 98.9%

pentru LCA, adică 73.7% din toate angiografiile coronariene, și 98.5% pentru RCA, adică 26.3% din toate angiografiile coronariene. O distribuție LCA / RCA similară a putut fi observată în setul de antrenare.

5.3.2.1.2 Performanța de Detectare a Fazei Cardiace față de Adnotările Experților Pentru a valida în continuare abordarea propusă, s-a luat în considerare un subset de 300 de pacienți aleși aleatoriu din setul principal de evaluare. Pentru fiecare dintre pacienți s-a ales aleator o achiziție angiografică care îndeplinea criteriile de includere, iar doi examinatori clinici experimentați au adnotat independent toate vârfurile R și T de pe semnalele ECG corespunzătoare. Deoarece vârfurile T nu au putut fi întotdeauna detectate în mod fiabil pe semnalele ECG extrase din imaginile angiografice, în cele din urmă s-au reținut doar acele achiziții angiografice pe care ambii examinatori clinici au identificat vârfurile R și T, ceea ce a dus la un set final de evaluare format din 207 angiografii coronariene. Pornind de la adnotările vârfurilor R și T, s-au definit etichetele țintă de fază cardiacă, așa cum este descris în secțiunea "Metode". Tab. 5.2 prezintă măsurile statistice pentru detectarea fazei cardiace pe acest set de evaluare.

5.3.2.1.3 Performanța de Detectare a Cadrului Final-Diastolic pe Setul Principal de Date de Evaluare Unul dintre obiectivele pentru efectuarea detectării fazei cardiace pe angiografiile coronariene este detectarea EDF-urilor adecvate pentru QCA 3D și segmentarea arterelor coronare. Tab. 5.3 afișează măsurile statistice pentru detectarea EDF pe setul principal de date de evaluare, obținute cu un interval de toleranță de ± 1 cadre în jurul EDF-ului de referință, așa cum a fost detectat pe baza semnalului ECG. Pentru predicțiile bazate pe imagine, fiecare cadru diastolic urmat de un cadru sistolic este considerat a fi un EDF. Pentru semnalul de referință, fiecare cadru diastolic care este urmat fie de un cadru de tranziție, fie de un cadru sistolic, este considerat a fi un EDF.

Tabela 5.1: Măsuri statistice pe setul de evaluare, calculate în două variante: (a) toate angiografiile coronariene au aceeași pondere și (b) toate cadrele au aceeași pondere.

Măsură statistică	Incluzând cadrele adiacente tranzițiilor		Excluzând cadrele adiacente tranzițiilor	
	Ponderi identice la nivelul angiografiilor	Ponderi identice la nivelul cadrelor	Ponderi identice la nivelul angiografiilor	Ponderi identice la nivelul cadrelor
Acuratețe	96.4%	96.6%	98.8%	98.8%
Sensibilitate	97.1%	97.3%	99.3%	99.4%
Specificitate	95.2%	95.3%	97.6%	97.5%
PPV	97.0%	97.1%	98.9%	98.9%
NPV	95.5%	95.7%	98.6%	98.7%

Tabela 5.2: Măsuri statistice pe setul de evaluare format din angiografii coronariene pentru care semnalele ECG au fost adnotate de doi examinatori clinici experimentați.

Măsură statistică	Ref. definită de examinatorul clinic 1	Ref. definită de examinatorul clinic 2
Acuratețe	97.6%	97.6%
Sensibilitate	97.2%	97.3%
Specificitate	98.3%	98.3%
PPV	99.3%	99.3%
NPV	93.6%	93.5%

Tabela 5.3: Măsurile statistice pentru detectarea EDF, cu un interval de toleranță de ± 1 cadre în jurul EDF-urilor de referință detectate pe baza semnalului ECG.

Măsură statistică	Performanță
Precizie	98.4%
Recuperare	97.9%
Scor F1	98.2%

5.4 Discuție

Detectarea fazelor cardiace și a cadrului diastolic final sunt etape esențiale pentru procesarea cantitativă a angiografiilor coronariene. În prezent, selectarea cadrului final-diaistolă se realizează fie manual, fie automat, pe baza semnalului ECG achiziționat simultan [35]. Este posibil ca semnalul ECG să nu fie întotdeauna disponibil, iar detectarea fazei cardiace bazată pe ECG prezintă mai multe dezavantaje: raportul semnal-zgomot poate fi prea mic pentru a detecta cu precizie sfârșitul diastolei sau semnalul poate prezenta artefacte [37, 38]. În acest articol, raportăm, după cunoștințele noastre, primul flux de lucru bazat pe învățare profundă pentru clasificarea fazei cardiace bazat doar pe cadrele angiografice, validat pe un set de date mare și realist.

Datele de antrenare au inclus, de asemenea, variația naturală a achizițiilor angiografice din fiecare vedere, inclusiv variațiile în ceea ce privește calitatea imaginii. S-a evitat în mod special idealizarea seturilor de date de antrenare și evaluare, pentru a se asigura că modelul este aplicabil în practica clinică zilnică cu acuratețea raportată în prezentul document. Atât setul de antrenare, cât și cel de evaluare au fost foarte mari, constând într-un total de 77435 de angiografii coronariene achiziționate de la 13081 de pacienți. Acest lucru asigură atât generalizarea modelului, cât și statistici de predicție precise. În plus, seturile de date de antrenare / validare și evaluare au fost dobândite din diferite locații clinice. Deși s-au utilizat exclusiv CNN-uri pentru detectarea fazei cardiace, rezultate similare pot fi obținute cu rețele neuronale recurente (RNN) pentru acest tip de sarcină.

5.5 Concluzii

Având în vedere seturile de date foarte mari utilizate în timpul antrenării și evaluării, s-a adoptat o strategie în care ieșirile țintă de fază cardiacă sunt determinate automat din semnalele ECG achiziționate simultan (au fost selectate doar acele achiziții pentru care semnalul ECG era fiabil). Această strategie a permis reducerea drastică a costurilor de adnotare (atât în termeni de timp, cât și de cheltuieli).

S-a concluzionat că fluxul de lucru propus, bazat pe imagini și care utilizează rețele neuronale adânci, a demonstrat o performanță bună, evitând astfel necesitatea selecției manuale a cadrelor și a achiziției ECG, reprezentând un pas relevant către evaluarea automată a CAD.

6. Concluzii Finale

6.1 Concluzii

Teza a fost axată pe dezvoltarea, aplicarea și evaluarea metodelor bazate pe Deep Learning pe seturi mari de date formate din achiziții de imagistică medicală, pentru a demonstra utilitatea unor astfel de metode în evaluarea și diagnosticarea pacienților cu boli cardiovasculare. Modelele DL au fost adaptate arhitectural, în timp ce protocoalele de antrenare au fost personalizate în funcție de specificul sarcinilor, având în vedere parametri eterogeni cum ar fi cantitatea și tipul de date, complexitatea predicției și constrângerile de timp de execuție.

Segmentarea semantică este o sarcină fundamentală, în special în domeniul medical. Cercetările recente au propus noi arhitecturi DNN (de exemplu, UNet) care au atins performanțe de vârf pentru segmentare și detectare în cazul datelor medicale. În această teză, acuratețea a fost îmbunătățită progresiv prin utilizarea unor astfel de arhitecturi noi și prin cercetarea unei serii de modificări personalizate, cum ar fi unirea segmentării și a detectării punctelor de reper în cadrul unei probleme de învățare de tip "multitask", utilizarea unor funcții criteriu noi (de exemplu Adaloss) și modelarea în comun a perechilor de cadre ED și ES în ecocardiografii pentru a îmbunătăți performanța finală a unei soluții DL pentru estimarea fracției de ejeție.

Tendința în domeniul imagisticii medicale este de a include algoritmi de inteligență artificială în echipamentul de imagistică, pentru a sprijini diagnosticarea și pentru a automatiza sarcinile repetitive, ceea ce duce la economii de timp și resurse. Prin urmare, performanța predictivă a unor astfel de algoritmi are un impact direct asupra capacității de utilizare a unor întregi soluții software. Un atribut valoros, pe lângă performanța reală de predicție, este estimarea incertitudinii. Simpla tratare a modelelor DNN ca niște cutii negre afectează capacitatea de explicare și fiabilitatea unei soluții de analiză medicală automată. Atunci când o incertitudine mare este asociată cu predicțiile modelului, datele de intrare pot fi marcate și prezentate cititorilor experți care iau o decizie privind următoarele etape ale procesului sau validează/corectează predicția modelului. Estimarea incertitudinii este un subiect de cercetare intensă. Multiple metode au fost investigate în contextul segmentării semantice, cu rezultate interesante. Regiunile din imaginea de intrare care au o calitate necorespunzătoare pot fi marcate, iar multiple variante de predicție pot fi propuse medicilor pentru o ajustare simplă și eficientă a rezultatul final.

Atunci când nu sunt disponibile adnotări pe scară largă, pixelii imaginilor brute pot fi totuși utilizați pentru a pre-antrena modele DNN mari. Sarcini euristice pretext au fost investigate pentru învățarea autosupravegheată pe ecocardiografii. Deși este un cadru simplu cu un potențial ridicat, trebuie să se acorde o atenție deosebită artefactelor din date care au o corelație ridicată cu sarcina pretext. Comportamentul DNN într-un regim autosupravegheat a fost investigat atunci când au fost injectate artefacte sintetice în setul de antrenare, ceea ce a arătat că astfel de sarcini pretext simple oferă modelelor posibilitatea de a trișa, în loc să le încurajeze să deducă caracteristicile reale relevante ale datelor. Au fost descrise abordări mai robuste, cum ar fi tehnicile de învățare contrastivă.

Când operează pe volume de intrare 3D, rețelele convoluționale tind să necesite multe resurse, atât în ceea ce privește cantitatea de memorie necesară, cât și în ceea ce privește volumul FLOP. Pentru o problemă de localizare a reperelor în ecocardiografiile 3D, au fost studiate tehnicile de învățare prin recompensă (eng. RL) și a fost dezvoltată o suită de agenți RL multi-scală pentru a înlocui CNN-ul echivalent de mari dimensiuni necesar pentru o astfel de sarcină. În loc să opereze pe în-

tregul domeniu de intrare, fiecare agent învață o traiectorie care maximizează recompensa. Astfel, se evită calculele inutile pe regiuni irelevante ale volumului de intrare. Deoarece fiecare agent are la bază un model DNN relativ mic, au fost atinse viteze rapide de execuție, în ciuda naturii secvențiale a algoritmului RL.

Învățarea nesupravegheată este un subdomeniu important în sfera învățării profunde. Modelarea generativă implică învățarea, implicită sau explicită, a distribuției de bază a datelor din cadrul setului de antrenare. În această teză au fost cercetate modele de tipul "Normalizing Flows" pentru a modela în mod explicit și eficient densitatea de probabilitate a perechilor de imagini lumen CT și măști de segmentare. Cercetările anterioare au evidențiat unele defecte în funcționarea modelelor NF profunde, limitându-le în captarea și concentrarea asupra conținutului semantic. A fost propusă și investigată o arhitectură revizuită care a prezentat performanțe mult îmbunătățite în detectarea valorilor aberante în comparație cu o metodă standard. Transformările afine individuale au fost înlocuite cu transformări convolutive, similare cu mecanismul din rețelele convoluționale obișnuite. Această modificare structurală a avut un impact pozitiv asupra capacităților semantice ale modelului NF propus. O altă opțiune este modelarea implicită a distribuției de probabilitate a datelor, folosind rețelele generative adversative (eng. GAN). Astfel de modele pot fi condiționate de semnale externe care dictează modul în care sunt sintetizate noile eșantioane. S-a investigat un strat de (de)normalizare adaptiv-spațial pentru construirea unui model care produce ecocardiografii apicale realiste condiționate de măștile de segmentare a camerelor furnizate de utilizator. Arhitectura a fost extinsă în continuare pentru a genera simultan perechi de cadre ecografice ED și ES, ca și cum acestea ar aparține aceleiași achiziții medicale. Astfel de modele generative puternice permit generarea de seturi de date sintetice cu proprietăți prescrise și personalizate, cum ar fi valorile EF și dimensiunea, plasarea și/sau orientarea camerelor.

Pentru prelucrarea cantitativă a angiografiilor coronariene, detectarea fazelor cardiace și a cadrelor final-diastolice sunt considerate etape esențiale. Pe baza unui set mare de date, a fost dezvoltat un algoritm de detectare a fazei cardiace bazat exclusiv pe imagini, care a demonstrat performanțe ridicate într-o mare varietate de unghiuri de achiziție, frecvențe cardiace și vizualizări. Pentru setările mai eterogene din ecocardiografii, a fost cercetată o extensie care utilizează straturi neuronale recurente, care a obținut, de asemenea, performanțe bune și robustețe în mai multe vizualizări. Acești algoritmi bazați pe imagini ameliorează problemele cunoscute legate de detectarea fazei cardiace pe bază de ECG.

În concluzie, au fost propuse și analizate mai multe abordări de învățare profundă pentru rezolvarea problemelor și subiectelor existente în domeniul diagnosticului bazat pe imagistică medicală pentru pacienții cu boli cardiovasculare. Au fost cercetate și testate diverse modificări care au dat naștere la soluții noi și personalizate pentru sarcini DL de mare importanță.

6.2 Contribuții Originale

Contribuțiile personale pot fi grupate în funcție de modalitatea de imagistică medicală. Fiecare soluție propusă este un pas spre un obiectiv comun în aplicațiile din domeniul sănătății: automatizarea fiabilă și de încredere a procedurilor de diagnosticare medicală, ceea ce duce la îmbunătățirea îngrijirii pacienților și a calității serviciilor.

6.2.1 Ecocardiografie

Ultrasunetele (US) sunt o tehnică de imagistică utilizată pe scară largă datorită caracterului său neinvaziv, în timp real și cu costuri reduse. Prin urmare, metodele de inteligență artificială dezvoltate pentru US pot ajunge la o bază mare de clienți, iar performanța lor poate avea un impact potențial mare.

Dezvoltarea modelelor DNN necesită, de obicei, cantități relativ mari de date adnotate. Este cunoscut faptul că metodele de pre-antrenare oferă o creștere a performanței atunci când reglarea su-

pervizată pe setul de date țintă pornește de la vectori parametrici pre-optimizați; astfel de parametri au fost obținuți prin forțarea modelului să învețe reprezentări expresive ale datelor neetichetate care fie sunt invariante față de unele augmentări, fie pot face distincție între ele. O contribuție originală este investigarea sarcinilor euristice pretext pentru obținerea de modele pre-antrenate pe ecocardiografii. Oglindirea pe orizontală a imaginii și ordonarea cadrelor sistolice au fost propuse ca două sarcini pretext adecvate pentru ecocardiografiile apicale BMode, deoarece acestea necesită ca modelul să deducă caracteristici despre amplasarea și volumul camerelor și despre deschiderea valvei mitrale. Astfel de caracteristici generale sunt exploatate în sarcinile supravegheate din aval: deoarece un model nu mai trebuie să învețe aceste caracteristici generale de la zero, acesta necesită un set de antrenare mai mic pentru a obține performanțe relativ bune.

O altă metodă de îmbunătățire a performanțelor modelului în fața unor volume mici de date (de un anumit tip) este generarea lor sintetică. Date sintetice din categoriile sub-reprezentate pot fi generate condiționat pentru a corespunde semanticii și proprietăților furnizate de utilizator. Un astfel de cadru a fost adaptat pentru a rula pe ecocardiografii apicale 2D. O contribuție originală sub forma unei extensii arhitecturale a permis unui GAN să eșantioneze perechi de cadre ED/ES în loc de cadre individuale, ca și cum acestea ar aparține aceleiași achiziții și aceluiași ciclu cardiac. Măștile de segmentare a camerelor au acționat ca semnale de condiționare, iar eșantioanele generate au respectat aspectul impuls. Această metodă permite generarea de perechi sintetice ED/ES cu EF controlat de utilizator, care pot fi de mare folos atunci când se încearcă construirea de clasificatoare pe seturi de antrenare dezechilibrate, prin creșterea populației intervalelor slab reprezentate.

Detectarea fazei cardiace este un pas crucial în soluțiile automate de calcul al EF. O contribuție originală a constat în dezvoltarea unei arhitecturi bazate pe RNN capabile să gestioneze natura eterogenă a ecocardiografiilor. Modelul a prezentat performanțe bune pe secvențe de diferite rate de cadre, lungimi și vizualizări.

O sarcină esențială în cadrul soluțiilor de evaluare cardiacă este auto-conturarea camerelor inimii, pentru a estima cantități precum volumele, tensiunile longitudinale globale etc. O altă contribuție originală a fost aceea de a investiga și de a extinde diverse arhitecturi de segmentare și de detectare a reperelor. Unirea celor două tipuri de predicție în cadrul unei funcții criteriu de tip multi-task și utilizarea conexiunilor de ocolire a straturilor au oferit un spor de performanță. O extensie arhitecturală propusă pentru predicția simultană pe cadre ED/ES cu condiționare reciprocă, a îmbunătățit robustețea modelului în fața calității variabile a achiziției.

S-a raportat în literatura de specialitate că modelele DNN obișnuite tind să fie prea încrezătoare în predicțiile lor, chiar și atunci când sunt greșite. Un astfel de comportament nu este benefic, în special în cazul soluțiilor automate de imagistică medicală. O altă contribuție originală a fost adaptarea metodelor recente de estimare a incertitudinii pentru sarcina de segmentare semantică. De exemplu, atunci când se utilizează procese gaussiene (GP), un cap de decodare DNN clasic a fost utilizat pentru a calcula o regiune de interes pe care să se aplice capul de decodare GP. Această regiune a fost mult mai mică decât întreaga imagine de intrare și, prin urmare, a permis o reducere de câteva ordine de mărime a resurselor de calcul necesare. O metodă bazată pe energie a fost, de asemenea, adaptată pentru estimarea incertitudinii conturului din jurul ventriculului stâng. Analizele au arătat că există o corelație moderată între estimările de incertitudine realizate prin diferite metode, chiar și atunci când se utilizează metrici euristice pentru estimarea incertitudinii globale.

În final, a fost investigată detectarea punctelor de reper 3D pe volumele de ecocardiografii utilizând agenți RL multi-scală. O contribuție originală a constat în reglarea procedurii de antrenare și a hiper-parametrilor săi, și a strategiei de implementare în producție pentru a obține agenți rapizi și robuști. Rezultatele indică o performanță bună de localizare, cu cerințe minime de timp de execuție și de calcul.

6.2.2 Angiografie Coronariană cu Tomografie Computerizată

Anumite soluții automate de imagistică pot avea o etapă opțională în care medicul poate introduce adnotări. Altele pot utiliza cascade de modele DNN. Se poate observa că o adnotare/predicție

defectuoasă se poate propaga de-a lungul întregului proces și poate afecta direct calitatea rezultatului final. Un model de audit poate ajuta la asigurarea faptului că datele de intrare, într-o anumită etapă, sunt valide. Această validitate poate fi formulată ca o problemă de detecție OoD, deoarece poate avea capacități de generalizare mai bune decât, de ex., utilizarea unui clasificator obișnuit care verifică un set limitat de defecte posibile.

Fluxurile de normalizare sunt candidați buni, deoarece permit un calcul rapid și eficient al densităților de probabilitate. Cu toate acestea, s-a raportat în literatura de specialitate că arhitecturile standard care se bazează pe straturi de cuplare afine tind să se concentreze mai mult pe texturi decât pe conținutul semantic, ceea ce duce la performanțe slabe de detecție OoD. O contribuție originală a fost propunerea unei noi arhitecturi NF care utilizează straturi de cuplare convoluționale pentru detectarea perechilor defectuoase de imagini CT coronariene și segmentări ale lumenului. O pereche a fost considerată defectuoasă dacă masca nu este complet aliniată cu imaginea lumenului. Au fost propuse perturbări ale măștii (de exemplu: zoom, dilatații, translații etc.) pentru a genera anomalii artificiale. La compararea cu un model de referință, analizele arată că modelul propus are capacități de interpretare semantică mai bune și prin urmare performanțe de detectare superioare. Experimentele de eșantionare au arătat că modelul de bază se concentrează într-adevăr pe textură și nu este capabil să producă eșantioane realiste, în timp ce modelul propus reușește să genereze eșantioane sintetice care au un aspect realist și sunt coerente din punct de vedere semantic între canale (de exemplu, masca lumenului corespunde imaginii CT).

6.2.3 Angiografie Coronariană Invazivă

Detectarea fazei cardiace reprezintă o sarcină importantă de preprocesare a angiografiilor coronariene invazive. În loc să se utilizeze semnalul ECG (care poate fi zgomotos sau absent), se propune o soluție bazată exclusiv pe imagini. Achizițiile video sunt pre-procesate și reeșantionate la o frecvență constantă a cadrelor și sunt introduse într-o rețea neuronală convoluțională. Într-o manieră autosupravegheată, ECG-ul este procesat pentru a extrage un semnal binar care descrie cele două faze cardiace; ECG-ul este utilizat doar în timpul antrenării pentru a oferi supervizare modelului. Deși procesarea ECG-ului nu este o procedură complet robustă, setul de antrenament este foarte mare și, astfel, mediază erorile locale în extragerea fazelor pe baza ECG-ului. Modelul este aplicat în mod glijant, utilizând un mecanism de vot majoritar, și poate fi aplicat, prin urmare, pe secvențe de lungime arbitrară. Analizele relevă performanțe bune și robuste de predicție a modelului, pentru o gamă largă de unghiuri de vizualizare și pentru multiple tipuri de vizualizări. Un set mare de evaluare (colectat dintr-o locație diferită de cea a setului de antrenare) a fost utilizat pentru a testa robustețea detecției fazelor cardiace în cadrul mai multor scenarii.

6.3 Diseminarea Rezultatelor Cercetării

Pe parcursul programului de doctorat, cercetările efectuate au condus la 4 publicații în calitate de autor sau coautor. Două articole de jurnal au fost publicate ca prim autor:

- **Ciusdel, C.**, et al., 2020. Deep Neural Networks for ECG-free Cardiac Phase and End-Diastolic Frame Detection on Coronary Angiographies. *Comput. Med. Imaging Graph.* 84, 101749, <https://doi.org/10.1016/j.compmedimag.2020.101749> (factor de impact 4.79, jurnal Q1)
- **Ciusdel, C.**, et al., 2022. Normalizing Flows for Out-of-Distribution Detection: Application to Coronary Artery Segmentation. *Appl. Sci.* 12, 3839. <https://doi.org/10.3390/app12083839> (factor de impact 2.68, jurnal Q2)

Pe parcursul colaborării în cadrul consorțiului proiectului european ITFoC (Information Technology for the Future Of Cancer), a fost publicat următorul articol în calitate de co-autor:

- Tsopra, R., ..., **Ciusdel, C.**, et al., 2021. A framework for validating AI in precision medicine: considerations from the European ITFoC consortium. BMC Med Inform Decis Mak 21:274, <https://doi.org/10.1186/s12911-021-01634-3> (factor de impact 2.80, jurnal Q3)

Un articol a fost publicat în calitate de coautor în prelegerile unei conferințe internaționale:

- Danu, M., **Ciusdel, C.**, Itu, L., 2020. Deep learning models based on automatic labeling with application in echocardiography. 24th Intl Conf. on System Theory, Control and Computing (ICSTCC). DOI: 10.1109/ICSTCC50638.2020.9259701

Următorul manuscris a fost trimis unui jurnal, în calitate de co-autor. Stadiul actual al manuscrisului este de "revizie majoră":

- Hatfaludi, CA., Tache, IA., **Ciusdel, C.**, et al., 2022. Towards a deep-learning approach for prediction of fractional flow reserve from optical coherence tomography. Appl. Sci. (factor de impact 2.68, jurnal Q2)

Următorul manuscris a fost trimis la recenzare (în calitate de co-autor) în cadrul unei conferințe internaționale:

- Hatfaludi, CA., **Ciusdel, C.**, Toma, A., Itu, L.M., 2022. Deep Learning based Aortic Valve Detection and State Classification on Echocardiographies. 20th IEEE Intl Power Electronics and Motion Control Conf.

Bibliografie

- [1] Cireşan, D., et al., 2011. Flexible, High Performance Convolutional Neural Networks for Image Classification. Proc 22nd Intl Joint Conf on Artificial Intelligence. Vol. 2: 1237–1242.
- [2] Krizhevsky, A., et al., 2012. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 60 (6): 84–90. doi:10.1145/3065386.
- [3] Anaya-Isaza, A., et al., 2021. An overview of deep learning in medical imaging. Informatics in Medicine Unlocked. Vol. 26. <https://doi.org/10.1016/j.imu.2021.100723>
- [4] Stokes M.B., Roberts-Thomson R., 2017. The role of cardiac imaging in clinical practice. Aust Prescr. Vol. 40:151-5. <https://doi.org/10.18773/austprescr.2017.045>
- [5] Ronneberger, O., et al., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597
- [6] Teixeira, B., et al., 2019. Adaloss: Adaptive Loss Function for Landmark Localization. arXiv:1908.01070
- [7] Long, J., et al., 2015. Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.
- [8] Ghorbani, A., et al., 2020. Deep learning interpretation of echocardiograms. npj Digital Medicine (2020) 3:10. doi.org/10.1038/s41746-019-0216-8
- [9] Danu, M., **Ciusdel, C.**, Itu, L., 2020. Deep learning models based on automatic labeling with application in echocardiography. 24th Intl Conf. on System Theory, Control and Computing (ICSTCC). DOI: 10.1109/ICSTCC50638.2020.9259701
- [10] Liu, W., et al., 2020. Energy-based Out-of-distribution Detection. Proc. NeurIPS
- [11] van Amersfoort, J., et al., 2021. Improving Deterministic Uncertainty Estimation in Deep Learning for Classification and Regression. arXiv:2102.11409v1
- [12] Leibfried, F., et al., 2020. A Tutorial on Sparse Gaussian Processes and Variational Inference. arXiv:2012.13962v11
- [13] Theodoridis, S. Machine Learning: A Bayesian and Optimization Perspective. 2nd Edition. Chapter 3. Elsevier, 2020. <https://doi.org/10.1016/B978-0-12-818803-3.00015-5>
- [14] Wilson, A.G., et al., 2016. Deep Kernel Learning. Proc. 19th Int. Conf. on AI and Stats (AISTATS, Spain 2016). JMLR: vol 51.
- [15] Park, T., et al., 2019. Semantic Image Synthesis With Spatially-Adaptive Normalization. 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2332–2341. doi: 10.1109/CVPR.2019.00244.

- [16] Arjovsky, M., et al., 2017. Wasserstein GAN. arXiv:1701.07875
- [17] Foster, D., 2019. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. O'Reilly Media, USA. ISBN: 978-1-492-04194-8
- [18] Gulrajani, I., et al., 2017. Improved training of wasserstein GANs. Proc. 31st Intl Conf. on Neural Information Processing Systems (NIPS'17).
- [19] Miyato, T., et al., 2018. Spectral normalization for generative adversarial networks. Proc. Intl Conf. on Learning Representations 2018.
- [20] Sutton, R., Barto, A., 2018. Reinforcement Learning: An Introduction. 2nd Ed. MIT Press.
- [21] Ghesu, F., et al., 2017. Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.41:1. doi:10.1109/TPAMI.2017.2782687
- [22] Sandler, M., et al., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. IEEE Conf on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520.
- [23] Dinh, L., et al., 2017. Density Estimation using Real NVP. Proc. ICLR.
- [24] Kingma, D.P., Dhariwal, P., 2018. Glow: Generative Flow with Invertible 1x1 Convolutions. Proc. NeurIPS.
- [25] Kirichenko, P., et al., 2020. Why Normalizing Flows Fail to Detect Out-of-Distribution Data. Proc. NeurIPS.
- [26] Nalisnick, E., et al., 2019. Do Deep Generative Models Know What They Don't Know?. Proc. ICLR.
- [27] Karami, M., et al., 2019. Invertible Convolutional Flow. Proc. NeurIPS.
- [28] Mark, D.B., et al., 2016. Economic outcomes with anatomical versus functional diagnostic testing for coronary artery disease. *Ann Intern Med*;165:94–102.
- [29] Levin, D.C., et al., 2019. Coronary CT angiography: reversal of earlier utilization trends. *J Am Coll Radiol*;16:147–55.
- [30] Coenen, A., et al., 2018. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: result from the MACHINE consortium. *Circulation: Cardiovascular Imaging*, Vol. 11.
- [31] Ryan, T.J., 2002. The coronary angiogram and its seminal contributions to cardiovascular medicine over five decades. *Circulation*. 106, 752-756.
- [32] Mozaffarian, D. et al., 2015. Heart disease and stroke statistics-2015 update: a report from the American Heart Association. *Circulation*. 131, e29-322. <http://doi.org/10.1161/CIR.000000000000152>.
- [33] Ng, V.G., Lansky, A.J., 2011. Novel QCA methodologies and angiographic scores. *Int J Cardiovasc Imaging*. 27, 157-165. <http://doi.org/10.1007/s10554-010-9787-9>.
- [34] Tu, S. et al., 2014. Fractional Flow Reserve calculation from 3-dimensional quantitative coronary angiography and TIMI frame count: A fast computer model to quantify the functional significance of moderately obstructed coronary arteries. *JACC Cardiovasc Interv*. 7, 768-777. <http://doi.org/10.1016/j.jcin.2014.03.004>.

- [35] Tröbs, M., et al., 2016. Comparison of Fractional Flow Reserve Based on computational fluid dynamics modeling using coronary angiographic vessel morphology versus invasively measured Fractional Flow Reserve. *Am J Cardiol.* 117, 29-35. <http://doi.org/10.1016/j.amjcard.2015.10.008>.
- [36] Itu, L.M. et al., 2016. A Machine Learning Approach for Computation of Fractional Flow Reserve from Coronary Computed Tomography. *J App Physiol.* 121, 42-52. <http://doi.org/10.1152/japplphysiol.00752.2015>
- [37] Kroft, L.J.M. et al., 2007. Artifacts in ECG-Synchronized MDCT Coronary Angiography. *Cardiac Imaging Review.* 189, 581-591.
- [38] Dehkordi, M.T., 2016. Extraction of the Best Frames in Coronary Angiograms for Diagnosis and Analysis. *J Med Signals Sens.* 6, 150–157. <http://doi.org/10.4103/2228-7477.186887>
- [39] Carreiras, C. et al., 2015. BioSPPy - Biosignal Processing in Python, <https://github.com/PIA-Group/BioSPPy/>.
- [40] Marzencki, M. et al., 2014. Diastolic Timed Vibrator: Noninvasive Pre-Hospitalization Treatment of Acute Coronary Ischemia. *IEEE Trans Biomed. Circ Syst.* 8, 313-324. <http://doi.org/10.1109/TBCAS.2013.2270181>.
- [41] Ronneberger, O. et al., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proc. MICCAI 2015.* https://doi.org/10.1007/978-3-319-24574-4_28
- [42] Kingma, D. et al., 2014. Adam: A Method for Stochastic Optimization. *Proc. Conf Learning Representations*, 2014.
- [43] Lin, T.Y. et al., 2017. Focal Loss for Dense Object Detection. *Proc. IEEE ICCV.* <http://doi.org/10.1109/TPAMI.2018.2858826>

Rezumat

Progresul din ultimii ani al procesoarelor masiv-paralele și disponibilitatea unor colecții mari de date au accelerat cercetarea metodelor de învățare profundă. Există o tendință de a include algoritmi bazați pe inteligența artificială în cadrul soluțiilor de imagistică medicală sau chiar ca parte a scannerului medical în sine. Având în vedere complexitatea specifică a sarcinilor de predicție și volumele mari de achiziții medicale, modelele de Deep Learning și-au câștigat pe bună dreptate locul în centrul soluțiilor de asistare a diagnosticului. Această teză investighează și extinde metode din DL și arhitecturi de modele neurale pentru a răspunde aspectelor specifice ale mai multor aplicații medicale.

Structura acestei teze urmează fluxul clinic pentru un pacient suspectat de boli cardiovasculare. Pornind de la imagistica non-invazivă de tip ultrasunete 2D, metodele de pre-antrenare auto-supervizată care utilizează sarcini euristice pretext sunt investigate pentru utilitatea și robustețea lor în construirea unor modele prototip generalizabile din date neadnotate. Generarea condiționată a ecocardiografiilor pe baza măștilor prescrise de segmentare ale camerelor inimii utilizând rețele adversative generative este cercetată pentru calitatea sa de sinteză a datelor noi și personalizate. O sarcină de clasificare video sub forma detectării fazei cardiace pe întregi achiziții este rezolvată de un nou model DNN. De asemenea, sunt investigate segmentarea semantică și modelarea incertitudinilor de predicție. Arhitecturi personalizate sunt testate pentru capacitatea lor de a rezolva cerințe specifice, cum ar fi coerența conturului între cadrele final-sistolice și final-diastolice. Metodele de incertitudine, cum ar fi procesele gaussiene și modelele bazate pe energie, sunt aplicate în contextul segmentării semantice. Pentru imagistica 3D cu ultrasunete, se investighează o optimizare bazată pe învățarea prin recompensă pentru o sarcină de localizare a punctelor de reper, pentru efectele sale în economia timp de execuție și resurse. În continuare, tehnicile de învățare nesupravegheată, cum ar fi fluxurile de normalizare, sunt explorate pentru capacitatea lor de modelare explicită a densității, într-o aplicație de detectare a punctelor din afara distribuției adevărate, pentru a semnaliza segmentările incorecte ale lumenului în angiografiile coronariene prin tomografie computerizată. O nouă arhitectură de model arată performanțe de detecție superioare prin exploatarea caracteristicilor semantice. În cele din urmă, este abordat subiectul important al detectării fazei cardiace pentru angiografiile coronariene invazive, pentru care a fost dezvoltată o soluție automată și robustă.

Soluțiile la toate aceste sarcini deschid noi perspective în medicina modernă, cum ar fi sistemele de diagnosticare și de prognoză complet automatizate.