



Examensarbete inom teknik
Grundnivå, 15 hp

AI-baserad segmentering av pulmonella blodkärl inom medicinsk bildanalys

AYSE UGUR

SHEIMAA ALCHAMI

Detta examensarbete har utförts i samarbete med
Karolinska Universitetssjukhuset
Handledare på Karolinska Universitetssjukhuset: Jimmy Yu och Jonas Sartorius



AI-baserad segmentering av pulmonella blodkärl inom medicinsk bildanalys

AI-Based Segmentation of Pulmonary Blood Vessels in Medical Image Analysis

AYSE UGUR
SHEIMAA ALCHAMI

Examensarbete inom medicinsk teknik
Grundnivå, 15 hp
Kurshandledare: Tobias Nyberg, Mattias Mårtensson
Examinator: Philip Köck

Kungliga Tekniska Högskolan
Skolan för kemi, bioteknologi och hälsa
Hälsovägen 11 C
SE-141 57 Huddinge, Sverige
<http://www.kth.se/cbh>

2025

Sammanfattning

Manuell analys av datortomografibilder (CT) utan kontrast, med fokus på expiratoriska pulmonella blodkärl, är idag en utmaning. För att underlätta och förbättra diagnostiken vid lungsjukdomar som interstitiella lungsjukdomar (ILD), inklusive lungfibros, finns ett växande behov av kvantitativa och automatiserade bildanalysmetoder. Syftet med denna studie var att utveckla och utvärdera en AI-modell för segmentering av lungkärl baserad på nnUNet, ett självkonfigurerande neuralt nätverk för medicinsk bildanalys. Totalt 33 thorax CT-bilder i exspirationsfas utan kontrastmedel hämtades från Karolinska Universitetssjukhuset i Solna via databasen PACS. Bilderna segmenterades i 3D Slicer och användes för träning av modellen med nnUNet i flera stegvisa träningsrundor. Modellens prestanda utvärderades kvantitativt med Dice Similarity Coefficient (DSC) mot manuellt segmenterade referensbilder från sju slumpmässigt utvalda fall. Resultaten visade att modellen successivt förbättrades och kunde identifiera en större del av kärlträdet speciellt i de centrala lungregionerna. Den slutliga modellen efter tre träningsrundor uppnådde ett medelvärde för DSC på 0,9911. Segmenteringen var dock mindre exakt i de perifera kärlområdena. nnUNet visar lovande resultat för automatiserad segmentering av pulmonella kärl i expiratoriska CT-bilder utan kontrast och har potential för framtida kliniska tillämpningar. Fortsatt utveckling och utökning av träningsdata krävs för att förbättra noggrannheten särskilt i de perifera delarna av lungorna.

Nyckelord:

Datortomografi (CT), Pulmonella blodkärl, 3D Slicer, nnUNet, Neurala nätverk, Medicinsk bildanalys, Automatisk segmentering, AI i radiologi

Abstract

Manual analysis of non-contrast computed tomography (CT) images, focusing on expiratory pulmonary blood vessels, remains a significant challenge. To facilitate and improve diagnosis in lung diseases such as interstitial lung diseases (ILD), including pulmonary fibrosis, there is an increasing need for quantitative and automated image analysis methods. The aim of this study was to develop and evaluate an AI model for vessel segmentation based on nnUNet, a self-configuring neural network for medical image analysis. A total of 33 non-contrast thoracic CT scans in the expiratory phase were retrieved from Karolinska University Hospital in Solna via the PACS database. The images were segmented in 3D Slicer and used to train the model with nnUNet through several iterative training rounds. The model's performance was quantitatively evaluated using the Dice Similarity Coefficient (DSC), compared against manually segmented reference images from seven randomly selected cases. The results showed a gradual improvement in the model's ability to identify an increasing portion of the vascular tree, particularly in central lung regions. The final model, after three training rounds, achieved a mean DSC of 0.991. However, segmentation was less accurate in peripheral vessel regions. nnUNet shows promising results for automated segmentation of pulmonary vessels in expiratory, non-contrast CT images and has potential for future clinical applications. Further development and expansion of training data are needed to improve accuracy, especially in peripheral lung areas.

Keywords: Computed Tomography (CT), Pulmonary vessels, 3D Slicer, nnUNet, Neural networks, Medical image analysis, Automatic segmentation, AI in radiology

Innehållsförteckning

1	Introduktion.....	1
1.1	Mål.....	1
2	Bakgrund.....	3
2.1	Bildkvalitet och påverkande faktorer.....	3
2.2	Tekniska inställningar vid CT.....	3
2.3	Skivtjocklek (eng. slice thickness).....	4
2.4	Kontrastmedlets roll vid bildanalys.....	5
2.5	Artificiell intelligens i medicinsk bildbehandling.....	5
2.6	Lungans anatomi.....	6
2.7	Utveckling av kvantitativ bildanalys och AI-metoder.....	6
2.8	Skillnader mellan inspiratoriska och expiratoriska CT-bilder.....	6
2.9	nnUNet.....	7
2.10	Utvärderingsmått.....	8
2.11	Tolkning av träningskurvor från nnUNet.....	8
3	Metod.....	11
4	Resultat.....	15
4.1	Träningsrundan 1 - Första modellträningen.....	15
4.2	Träningsrunda 2 - Utökad träning med 8 bilder.....	16
4.3	Träningsrunda 3 - Utökad modellträning med 25 bilder.....	18
4.4	DSC Resultat.....	19
4.5	Resultat av de olika konfigurationsvalen.....	20
5	Diskussion.....	21
5.1	Valet av modellens konfiguration.....	21
5.2	Effekter av bildkvalitet på modellens prestanda.....	22
5.3	Datamängdens betydelse.....	22
5.4	Begränsningar med DSC.....	23
5.5	Feltyper och felkällor i segmenteringen.....	23
6	Slutsats.....	25
7	Referenser.....	27
Bilagor		
Bilaga 1: nnUNet kommando i Terminal		
Bilaga 2: Konfigurationsfilen dataset.json för träning i nnUNet		
Bilaga 3: MATLAB KOD för DSC		

Förkortningar

AI – Artificiell intelligens (eng. Artificiell Intelligence)

CT – Datortomografi (eng. Computed Tomography)

DPLD – Diffusa parenkymala lungsjukdomar (eng. Diffuse parenchymal lung diseases)

DSC – Dice Similarity Coefficient

IVK – Intravenöst kontrastmedel (eng. Intravenous contrast)

Nativ – Bilder tagna utan kontrastmedel (eng. non-contrast imaging)

PACS – Picture Archiving and Communication System

QI – Kvantitativ avbildning (eng. Quantitative Imaging)

VSC – Visual Studio Code

1 Introduktion

Datortomografi (CT, eng. Computed tomography) är en av de viktigaste metoderna för att undersöka och diagnostisera lungsjukdomar [1]. Genom CT undersökningar kan detaljerade tvärsnittsbilder av lungorna framställas, som gör det möjligt att upptäcka förändringar i kärlstrukturer och andra anatomiska avvikelser [1]. För att få en så tydlig bild som möjligt används ofta kontrastmedel i kombination med bilder tagna när patienten andas in djupt, så kallade inspiratoriska bilder. Detta skapar skarpa gränser mellan blodkärl och omgivande vävnad och förenklar därmed både visuell bedömning och automatisk segmentering [2].

Majoriteten av dagens artificiella intelligens (AI) baserade segmenteringsmodeller är utvecklade för inspiratoriska, kontrastförstärkta CT bilder, där de tydliga vävnadsgränserna underlättar för algoritmer att lära sig och segmentera strukturer korrekt [3]. Däremot presterar dessa modeller sämre på expiratoriska bilder utan kontrastmedel, där lungorna är mindre utspända och vävnadsskillnaderna är mindre tydliga [4]. Studier har visat att expiration minskar kontrastförstärkningen och försvårar tolkningen av lungparenkymet vilket påverkar den diagnostiska noggrannheten och segmenteringsresultaten [4, 5].

Eftersom de flesta existerande AI modeller inte är utvecklade för att hantera expiratoriska CT bilder tagna utan kontrastmedel, uppstår betydande utmaningar när de tillämpas på sådant bildmaterial. Den lägre bildkvaliteten och vävnadskontrasten i dessa bilder ställer högre krav på modellernas anpassningsförmåga [4]. Trots detta finns det fortfarande få modeller som är specifikt anpassade för segmentering av expiratoriska bilder tagna utan kontrastmedel [6].

En möjlig väg framåt är användningen av nnUNet. NnUNet är ett självkonfigurerande och flexibelt djupinlärningssystem som tidigare har visat mycket goda resultat inom flera områden av medicinsk bildsegmentering [7]. Till skillnad från traditionella nätverk, som ofta kräver manuell finjustering för varje nytt dataset, kan nnUNet automatiskt analysera datasetet och anpassa sin struktur och träningsstrategi därefter [3,7]. Detta gör nnUNet särskilt lämpligt för utmanande segmenteringssuppgifter där bildkvaliteten och vävnadskontrasten varierar.

Tillförlitligheten och noggrannheten i analysen kan förbättras genom att anpassa och vidareutveckla nnUNet för segmentering av lungkärl på expiratoriska CT bilder tagna utan kontrastmedel. Eftersom nnUNet har visat sig klara av att hantera svårare bildmaterial och varierande bildkvalitet [7], framstår det som ett särskilt lovande alternativ för just den här typen av segmenteringssuppgifter. Syftet med detta arbete var därför att anpassa och optimera nnUNet för att möjliggöra tillförlitlig segmentering av lungkärl på expiratoriska CT-bilder tagna utan kontrastmedel, med avsikten att förbättra den kliniska diagnostiken och underlätta patienthandläggningen vid lungsjukdomar.

1.1 Mål

Målet med projektet var att utveckla en nnUNet-modell för klinisk användning vid segmentering av expiratoriska CT-bilder tagna utan kontrastmedel.

2 Bakgrund

Detta kapitel presenterar relevant bakgrundsteori och tekniska begrepp. Dessa är viktiga för att förstå de metoder och analyser som tillämpats i detta projekt.

2.1 Bildkvalitet och påverkande faktorer

CT bilders kvalitet påverkas av flera faktorer, där de fyra mest centrala är spatial upplösning, bildkontrast, bildbrus och artefakter [8,9]. Spatial upplösning beskriver systemets förmåga att återge små detaljer i bilden och påverkas av tekniska parametrar såsom detektorupplösning och storleken på röntgenrörets fokuspunkt. Bildkontrast avgörs av skillnader i intensitet mellan olika vävnader och är viktig för att kunna särskilja anatomiska strukturer.

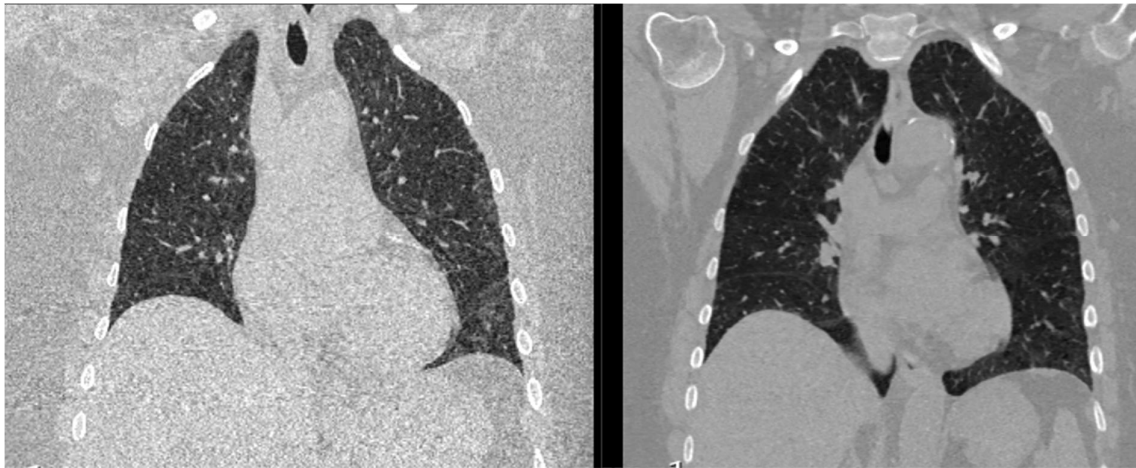
Artefakter kan uppstå på grund av olika felkällor, exempelvis rörelser från patienten eller inre organ (som hjärtats aktivitet), och kan leda till att bilden inte korrekt speglar den faktiska anatomin. Bildbrus uppstår huvudsakligen på grund av variationer i antalet detekterade röntgenfoton, så kallat kvantbrus, vilket minskar den diagnostiska tydligheten i bilden [9].

2.2 Tekniska inställningar vid CT

En viktig faktor som påverkar CT-bildens kvalitet är valet av rekonstruktionsfilter, även kallat kernel eller convolution kernel [10]. När en CT-skanner samlar in data registreras en stor mängd råa röntgenprojektioner från olika vinklar runt kroppen. Dessa projektioner måste rekonstrueras till tvärsnittsbilder och det görs med hjälp av rekonstruktionsfilter vilka används för att förstärka detaljer (skärpa) eller reducera brus [10].

Hårda filter, t.ex. Br70, Br80, förstärker högfrekventa signaler det vill säga att små, snabba förändringar i intensitet t.ex. kanter och små strukturer blir tydligare, som används för att förbättra den spatiala upplösningen [11]. Det är därför mest användbart vid bedömning av ben eller lungparenkym. Nackdelen är att de också förstärker bildbrus. Enligt Jimmy Yu, specialistläkare i thoraxradiologi vid Karolinska Universitetssjukhuset, dämpar mjuka filter, t.ex. Br30, Br40, högfrekventa komponenter och förstärker lågfrekventa signaler. Detta leder till reducerat brus, men medför även en försämring av den spatiala upplösningen, eftersom finare strukturer kan framträda mindre tydligt.

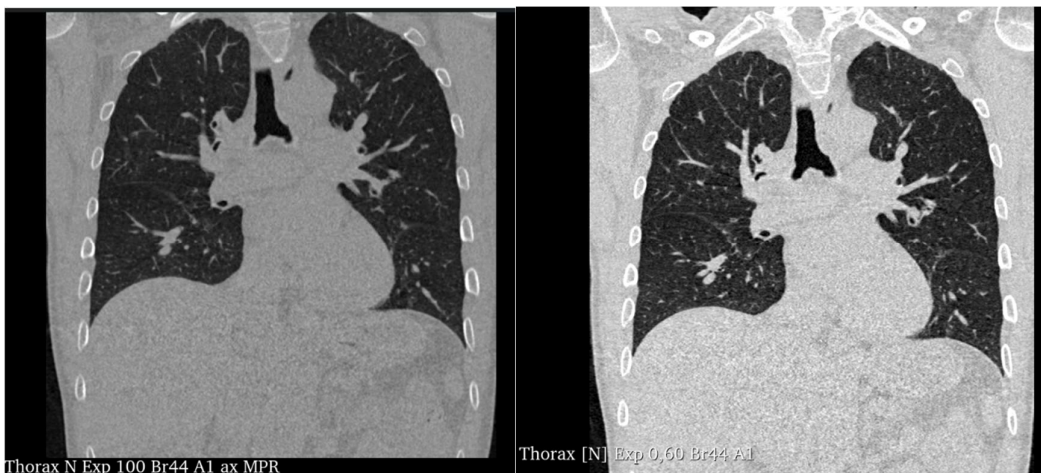
Jimmy Yu berättar vidare att Br40 och Br44 ofta används vid thoraxundersökningar. Yu nämner att Br40 är ett något mjukare filter som ger en bra balans mellan brusreduktion och detaljåtergivning och att Br44 är något hårdare och används när man vill ha bättre kantdefinition, exempelvis vid lunganalys, men det genererar också mer brus, se figur 1. Valet av filter är alltså en avvägning mellan kontrast, brus och upplösning, och styrs av vad som ska analyseras i bilden.



Figur 1: CT-bilder av thorax rekonstruerade med två olika filterinställningar för att illustrera effekten på kärlsynlighet i lungparenkymet. Vänster bild visar rekonstruktion med ett hårt filter, Br44, vilket ger skarpare konturer av benstrukturer men med ökat bildbrus och sämre differentiering av lungans mjukvävnad. Höger bild är rekonstruerad med ett mjukare filter, Br40, vilket resulterar i förbättrad visualisering av blodkärl i lungan tack vare ökad kontrastupplösning och minskat brus. Skillnaden i filterinställning illustrerar hur valet av rekonstruktionsalgoritm påverkar bildernas diagnostiska information. Bildanvändning godkänd av Jimmy Yu.

2.3 Skivtjocklek

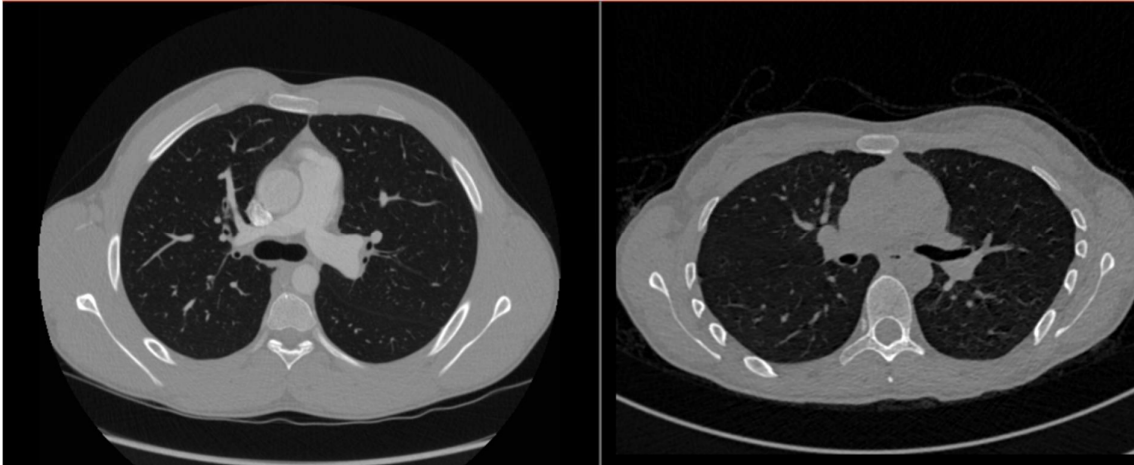
Skivtjocklek (eng. slice thickness), är en annan central parameter som påverkar både bildkvalitet och filstorlek. Den avser tjockleken på de rekonstruerade bildskivorna i z-axeln, patientens längdriktning. Tunna snitt (t.ex. 0,6–1,5 mm) ger högre detaljupplösning och är särskilt användbara för tredimensionell visualisering, volymberäkningar eller segmentering [2]. Dock medför de mer brus och större datamängder. Tjockare snitt (t.ex. 3–5 mm) ger en mer homogen bild med mindre brus, men med sämre förmåga att återge små strukturer eller subtila förändringar [8]. Figur 2 presenterar en CT-bild som den till vänster visar en snitttjocklek på 1 mm, medan bilden till höger är rekonstruerad med 0,6 mm snitttjocklek.



Figur 2: CT-Bilder av thorax med olika snitttjocklek för att illustrera effekten på bilddetaljer och kärlsynlighet i lungparenkymet. Vänster: CT-bild rekonstruerad med 1 mm snitttjocklek, vilket ger en övergripande god anatomisk översikt men med begränsad upplösning av finare strukturer såsom små perifera blodkärl. Höger: CT-bild med 0,6 mm snitttjocklek, vilket möjliggör högre rumslig upplösning och därmed förbättrad visualisering av tunna kärl i lungorna samt mer detaljerad anatomisk information. Bilderna är hämtade från PACS, bildanvändning godkänd av Jimmy Yu.

2.4 Kontrastmedlets roll vid bildanalys

En signifikant skillnad föreligger mellan CT-bilder tagna med respektive utan kontrastmedel, se Figur 3. Intravenöst kontrastmedel (IVK) ökar attenueringen av röntgenstrålning i kärlstrukturer, enligt Jimmy Yu är IVK vanligen baserad på jodhaltiga medel.



Figur 3: CT-bilder tagna från Karolinska Universitetssjukhusets PACS-system. Bilderna visar axiella snitt av lungorna, med IVK till vänster och NATIV till höger. Bildanvändning godkänd av Jimmy Yu.

Röntgen attenueringen, det vill säga hur mycket strålning som absorberas eller sprids bort från sin bana, är starkt beroende av materialets atomnummer. Ju högre atomnummer, desto mer sannolikt är det att fotoner interagerar med materialet via processer som fotoelektrisk absorption och Compton-spridning. Vid de energinivåer som används i medicinsk CT (vanligtvis runt 70-140 keV) [12], sker mycket av attenueringen genom fotoelektrisk absorption, där sannolikheten för interaktion är ungefär proportionell mot atomnumret i kubik (Z^3) [13]. Eftersom jod har ett högt atomnummer jämfört med kroppens naturliga vävnader, som huvudsakligen består av kol, syre, väte och kväve med mycket lägre atomnummer, leder tillsatsen av jod till kraftigt ökad absorption av röntgenstrålar i de vävnader där kontrastmedlet ansamlas speciellt i blodkärl. Detta gör att blodkärl och andra strukturer som innehåller kontrastmedel får högre attenuering på CT bilder som i sin tur förbättrar vävnadskontrasten och gör det lättare att skilja kärl och andra kontrastfyllda strukturer från omgivande vävnad.

Vid undersökningar utan kontrastmedel (nativ) försvåras differentiering mellan kärl och omgivande parenkym speciellt i områden med hög densitet vilket påverkar bildanalysens noggrannhet negativt. Enligt Jimmy Yu kan administrering av kontrastmedel vara kontraindicerad, exempelvis vid jod överkänslighet eller nedsatt njurfunktion, därför föreligger ett kliniskt behov av metodutveckling för analys av icke-kontrastförstärkta CT-bilder.

2.5 Artificiell intelligens i medicinsk bildbehandling

Artificiell intelligens (AI) har genom åren fått en mycket viktigare roll inom sjukvården och används idag i flera olika medicinska tillämpningar. Detta inkluderar diagnostik, behandling och patientvård [14]. Stora mängder medicinska data kan analyseras snabbare än vad det tidigare varit möjligt och medför fördelar för radiologin. Inom sjukvården är AI ett

hjälpmedel för att automatiskt detektera patologiska förändringar i röntgen och datortomografi bilder som bidrar till tidig diagnostik av exempelvis lungcancer, kroniskt obstruktiv lungsjukdom (KOL) och interstitiella lungsjukdomar [15].

2.6 Lungans anatomi

Lungorna är de primära organen för gasutbyte i människokroppen och är avgörande för att upprätthålla syre- och koldioxidnivåer i blodet [16]. De består av två huvudsakliga lobber, en i höger lunga och två i vänster lunga, och är belägna i brösthålan. Inom lungvävnaden finns ett finförgrenat nätverk av blodkärl, där syrefattigt blod från kroppen transporteras via lungartärerna till lungorna. I de perifera delarna av lungan övergår artärerna i kapillärnätverk som omsluter alveolerna, små luftblåsor där själva gasutbytet sker. Här diffunderar syre från inandningsluften in i blodet, medan koldioxid från blodet passerar ut i alveolerna för att andas ut. Efter gasutbytet förs det syresatta blodet tillbaka till hjärtat via lungvenerna för vidare distribution i kroppen [16].

2.7 Utveckling av kvantitativ bildanalys och AI-metoder

Mycket av dagens bildtolkning förlitar sig på en manuell bedömning som är tidskrävande och observatörberoende. Enligt Jimmy Yu är en detaljerad kartläggning av kärlstrukturen viktig för diagnostiken, men samtidigt svår att åstadkomma enbart genom manuell bedömning. Yu säger dessutom att en specifik utmaning är analysen av strukturella förändringar i lungornas kärlträd, som en viktig del av diagnostiken för en mängd lungsjukdomar. Vidare nämner Yu att det är svårt att enbart basera bedömningen på ord eller kliniska observationer, eftersom sjukdomens svårighetsgrad inte alltid framgår tydligt utan en mer detaljerad och kvalitativ analys. Därför behövs avancerade metoder för att kvantifiera och visualisera dessa förändringar för att bättre förstå sjukdomens omfattning och för att ge en mer exakt bedömning.

För att hantera dessa utmaningar har Jimmy Yu och Jonas Sartorius, medicinekandidat projektassistent, på Karolinska Universitetssjukhuset och Karolinska Institutet utvecklat en metod som kombinerar kvantitativ bildanalys (eng. QI, Quantitative Imaging) med AI drivna neurala nätverk. Metoden innefattar först en automatiserad segmentering av lungkärnen på CT bilder, följt av omvandling av dessa segmenteringar till en träd-graf. Ur denna graf kan objektiva mått extraheras, såsom kärllängder, kärlvolymer och förgreningsegenskaper. Dessa kvantitativa data kan ge nya insikter om kärlförändringar vid lungsjukdom och bidra till en högre diagnostisk precision.

2.8 Skillnader mellan inspiratoriska och expiratoriska CT-bilder

Enligt Jimmy Yu är befintliga AI modellen dock främst tränad på inspiratoriska CT-bilder tagna med kontrastmedel. Detta begränsar dess användbarhet för expiratoriska bilder tagna utan kontrastmedel. Figur 5 visar skillnaden mellan en inspiratorisk och en expiratorisk lunga i en CT-bild. Vid inspiration är lungvolymen större och de pulmonella strukturerna, inklusive blodkärlen, är mer utspända och tydligare avgränsade. I den expiratoriska fasen är lungvolymen mindre, som gör blodkärlen svårare att urskilja på grund av ökad vävnadskompression och högre attenuering. Denna skillnad i volym och attenuering påverkar både bildkvalitet och möjligheten att segmentera strukturer automatiskt med hjälp av AI.

Vid inspiration är förhållandena däremot gynnsamma, eftersom det stora luftfyllda utrymmet i lungorna skapar ett högt luft-till-vävnad förhållande, något som ger upphov till tydligare kontraster i CT- bilderna. I den expiratoriska fasen minskas mängden luft i lungorna vilket leder till att vävnaderna komprimeras och skillnaden i attenuering minskas mellan olika strukturer, som i sin tur påverkar bildens skärpa och tydlighet negativt.



Figur 5: Koronala CT-bilder av thorax tagna under två olika andningsfaser. Vänster: Bild tagen under inspiration med IVK som resulterar i större lungvolym, tydligt avgränsade pulmonella strukturer och ökad kontrast mellan luft och vävnad. Höger: Bild tagen under nativ, där minskad lungvolym och vävnadskompression leder till minskad skillnad i attenuering mellan olika strukturer. Detta försämrar kontrasten och gör det svårare att urskilja exempelvis blodkärl i parenkymet. Bilderna illustrerar hur andningsfas och kontrastanvändning påverkar bildkvaliteten och den diagnostiska tydligheten i CT-bilden. Bilderna är hämtade ur PACS, bildanvändning godkänd av Jimmy Yu.

2.9 nnUNet

En av de ledande AI modellerna för segmentering av medicinska bilder är nnUNet, en självkonfigurerande djuplärningsbaserad modell för medicinsk bildsegmentering [7]. Modellen erbjuder flera konfigurationslägen för träning och initial segmentering, så kallad inferens, däribland 2D, 3D low resolution (3d_lowres), 3D cascade (3d_cascade_fullres) och 3D full resolution (3d_fullres). Dessa lägen är utformade för att hantera variationer i bildstorlek, snittets tjocklek och hårdvarubegränsningar [17]. 2D-konfigurationen 2d tränar modellen på enskilda snitt istället för volymer, vilket minskar kraven på minne och beräkningskapacitet. 3d_lowres är en enklare och snabbare variant, detta tränar modellen på nerskalade volymer [18]. Den används när hårdvaruresurserna är begränsade eller när bildvolymer är mycket stora. Fördelen är minskad minnesanvändning och snabbare träning, men detta sker på bekostnad av reducerad spatiell upplösning i segmenteringsresultatet. 3d_fullres tränar modellen på volymer i full upplösning. Det gör att små anatomiska strukturer kan återges mer detaljerat, samtidigt som modellen kan utnyttja det rumsliga sammanhanget mellan snitten. Den är särskilt lämplig när upplösningen är relativt enhetlig och när hög noggrannhet krävs. Däremot kräver den mer GPU-minne och beräkningskraft än 2D. 3d_cascade_fullres kombinerar 3d_lowres och 3d_fullres i två steg [19]. Först används 3d_lowres för att skapa en grov förutsägelse, som sedan förfinas av en separat 3d_fullres-modell. Denna kaskadstrategi kan förbättra segmenteringen vid svåra uppgifter eller stora variationer i strukturstorlek, men är också den mest resurskrävande av konfigurationerna eftersom två modeller måste tränas och köras sekventiellt [18].

Vid träning av nnUNet krävs en specificering av vilken fold som ska användas i korsvalideringen. nnU-Net delar automatiskt upp träningsdata i fem delar (folds 0 till 4) för att möjliggöra femfaldig korsvalidering. Ett värde mellan -f 0 och -f 4 innebär att träningen sker på en enskild fold, där modellen tränas på 4/5 av data och valideras på den återstående femtedelen. Om man vill träna på alla folds vilket ger fem modeller, en per fold, måste -f all anges för att träna på hela datasetet utan valideringsuppdelning, t.ex. inför slutgiltig inferens [20].

2.10 Utvärderingsmått

Det används olika metoder för att mäta prestanda vid träning och utvärdering av segmenteringsmodeller. Dice Similarity Coefficient (DSC) är ett mått som används för att mäta hur väl modellen lyckas segmentera blodkärl i expiratoriska CT-bilder tagna utan kontrastmedel, jämfört med en manuellt skapad referenssegmentering (ground truth) [21]. Det resulterade värdet av DSC ligger mellan 0 och 1, där ett högre värde indikerar större överensstämmelse mellan modellens segmentering och referensen. Ett DSC-värde på 1 innebär perfekt överlappning, medan 0 innebär att segmenten inte överlappar alls [22].

nnUNet beräknar under träningen efter varje epok en egen så kallad pseudo dice, se exempel på träningslog i filen bifogat på DiVA. En epok motsvarar ett genomlopp av all träningsdata. Det innebär att modellen har behandlat varje bild i träningsuppsättningen en gång. Antalet epoker används för att ange hur långt modellen har tränats. Pseudo dice värde används för att följa modellens förbättring över epok och visas i den träningskurva som nnUNet automatiskt genererar [7].

Den teoretiska formeln som används för att beräkna DSC beskrivs enligt ekvation 1 [20, 21]:

$$DSC = \frac{2 \cdot |A \cap B|}{A + B} \quad 1)$$

Där A är mängden pixlar i modellens segmenteringsmask, B är mängden pixlar i referenssegmenteringen och $|A \cap B|$ är antalet pixlar där de två segmenteringarna överlappar.

2.11 Tolkning av träningskurvor från nnUNet

Under modellträningen genererar nnUNet automatiskt en träningskurva som visualiserar hur olika mått, såsom förlustvärde och DSC, förändras över epoker [7]. Figur 6 visar ett exempel på en träningskurva som skapas under modellens inlärningsprocess. Den vänstra Y-axeln representerar förlustvärdet (loss), där `loss_tr` visar hur stor skillnaden är mellan modellens förutsägelser och de faktiska svaren i träningsdata. Ett högt värde innebär att modellen gör stora fel, medan ett lågt värde visar att modellen börjar lära sig. Förlustvärde på valideringsdata, `loss_val`, mäter samma sak som `loss_tr`, men på en separat datamängd som modellen inte tränas på. Det visar hur bra modellen presterar på nya, osedda data. Om detta värde börjar öka medan `loss_tr` fortsätter minska, kan det tyda på att modellen överanpassar sig till träningsdata (s.k. överträning eller överfitting). Den högra Y-axeln visar DSC, som används för att mäta överlappningen mellan modellens prediktion och referenssegmenteringen. Här presenteras två varianter: pseudo dice, vilket motsvarar det uppmätta värdet per epok, samt pseudo dice (mov. Avg.), som visar ett glidande medelvärde

över epokerna. X-axeln representerar antalet epoker, det vill säga hur många gånger hela träningsdata har passerats genom modellen under träning [7].



Figur 6: Exempel på träningskurva för nnUNet-modellen, som visar hur förlustvärden och DSC utvecklas över epoker. Den vänstra Y-axeln visar förlustvärdet (loss) för träningsdata (loss_tr, blå linje) och valideringsdata (loss_val, röd linje). Den högra Y-axeln visar DSC för valideringsdata, både som uppmätt värde per epok (pseudo dice, grön prickad linje) och som ett glidande medelvärde (pseudo dice (mov. avg.), grön heldragen linje). X-axeln representerar antalet epoker.

3 Metod

Projektet inleddes med att extrahera datortomografibilder ur PACS-systemet vid Karolinska Universitetssjukhuset. Totalt valdes 32 expiratoriska, nativa CT-undersökningar ut från en större databas innehållande patologiska fall, där bilderna representerar relativt frisk lungvävnad. För varje undersökning fanns rekonstruktioner filtrerade med både Br40 och Br44 samt snittjocklekar på 1,0 mm och 0,6 mm tillgängliga i PACS. Vi valde att använda bilder rekonstruerade med Br40 och en snittjocklek på 0,6 mm för att träna modellen. Bilder rekonstruerade med Br44-filtret med snittjocklek 0,6 mm användes enbart för utvärdering av en andra nnUNet modell. Båda modellerna tränades enligt beskrivningen nedan.

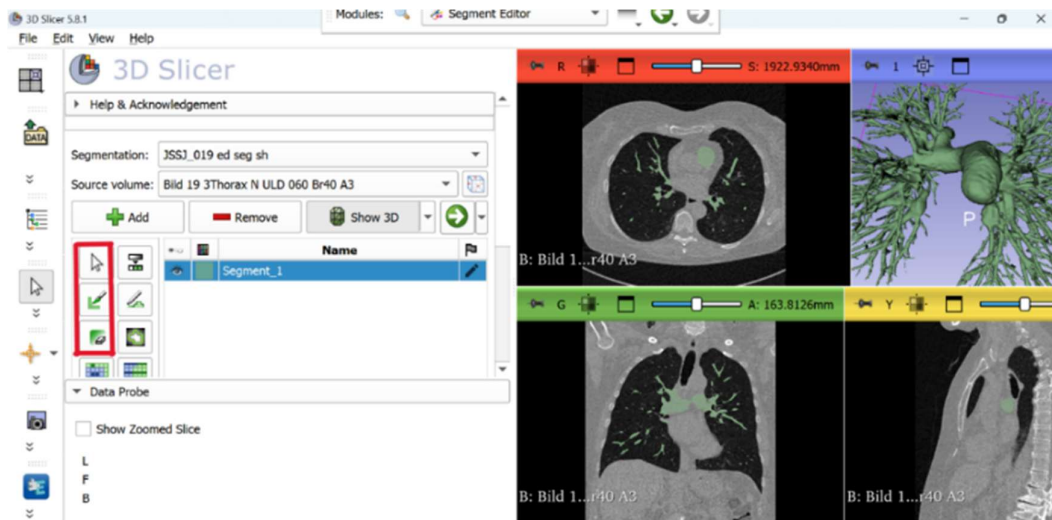
För att genomföra en första segmentering användes Jimmy Yu:s dator, där nnUNet (Division of Medical Image Computing, German Cancer Research Center, Heidelberg, Germany) redan var installerat och färdigt konfigurerat. En mapp, benämnd input folder, skapades för att lagra de fem första CT-bilderna som användes för initial segmentering. Segmenteringen baserades på en tidigare tränad nnUNet modell för inspiratoriska CT-bilder tagna med kontrastmedel. Denna modell var färdigutvecklade av Jimmy Yu och Jonas Sartorius. De initiala segmenteringarna genomfördes genom att i ett terminalfönster följa kommandostrukturen:

```
nnUNetv2_predict -i INPUT_FOLDER -o OUTPUT_FOLDER -d  
DATASET_NAME_OR_ID -c CONFIGURATION --save_probabilities
```

Där INPUT_FOLDER angav sökvägen till mappen med de aktuella bilderna, OUTPUT_FOLDER är mappen där segmenteringsresultaten sparades, DATASET_NAME_OR_ID hänvisade till den modell/dataset som inferenserna skulle baseras på, och CONFIGURATION specificerade önskad modellkonfiguration. För modellen med 2D-konfigurationen angavs CONFIGURATION till 2d och för 3D-konfigurationen angavs den till 3d_fullres. Flaggan --save_probabilities användes för att spara sannolikhetskartor från modellen, se exempel i bilaga 1.

Efter att den första segmenteringen genomförts justerades de resulterande fem inferenserna manuellt med hjälp av programvaran 3D Slicer, version 5.2.2 (Brigham and Women's Hospital, <https://www.slicer.org/>). Dessa manuellt korrigerade segmenteringar utgjorde sedan grunden för vidare modellträning och utvärdering.

Den manuella justeringen av nnUNet:s segmentering gjordes genom att bilden och den tillhörande inferensen importerades till 3D Slicer. Där valdes bilden som volym och inferensen som segmentering. Figur 8 illustrerar hur detta såg ut i programmet. För att korrigera segmenteringen användes tre huvudsakliga verktyg i 3D Slicer, som syns i den röda rutan i figur 8. Pilen användes för att navigera och zooma in/ut i olika plan, exempelvis axiellt eller sagittalt vy. Sedan användes pennan för att lägga till kärlstrukturer och suddgummit för att ta bort felaktiga segment.



Figur 8: Gränssnitt i programvaran 3D Slicer under segmentering av lungstrukturer. Till vänster visas modulen där användaren kan skapa och redigera segment. Med hjälp av verktygsknapparna till vänster (markerade i röd ruta) kan olika redigeringsfunktioner väljas, exempelvis pilen, penselverktyg, suddverktyg, vilka används för att manuellt justera segmentets omfattning i bilden. Till höger visas tre tvärsnittsbilder i tre anatomiska plan: axial (röd övre vänster), koronal (grön nedre vänster) och sagittal (gul nedre höger). Det fjärde fönstret (övre höger) visar en 3D-rendering av det segmenterade området, där den gröna strukturen representerar segmentet. CT-bilderna är tagna ur PACS och har godkänts för användning av Jimmy Yu.

När de manuellt korrigerade segmenteringarna var färdigställda organiserades datamaterialet i enlighet med nnUNet:s standardstruktur. En datasetmapp skapades med tre undermappar, imagesTr som innehöll CT-bilderna som skulle användas för träning, labelsTr hade motsvarande segmenteringsfiler och imagesTs var avsedd för testbilder, men denna mapp lämnades tom i detta projekt. Utöver dessa mappar skapades en fil kallad dataset.json som innehåller metadata och vägvisning för hur datasetet är strukturerat, inklusive information om bildmodalitet, etiketter samt filvägar för träning och eventuell testning. Denna fil skrevs i Visual Studio Code version 1.100.2 (Microsoft Corporation, <https://code.visualstudio.com/>) och utformades enligt nnUNet:s specifikationer, se bilaga 2 för kod.

Därefter genomfördes en förbearbetning som genererade planer för datakonvertering, normalisering och indelning i patchar, baserat på informationen i den angivna filen dataset.json. Förbearbetningen utfördes genom att ange följande kommando i terminalfönstret:

```
nnUNetv2_plan_and_preprocess -d DATASET_ID -c CONFIGURATION --
verify_dataset_integrity
```

Här användes flaggan --verify_dataset_integrity som säkerställde att datasetet är korrekt organiserat och läsbart för nnUNet. Två förbearbetning utfördes en med 2d och en med 3d_fullres konfiguration. Planerings- och preprocessingfiler sparades sedan i mappen nnUNet_preprocessed.

När förbearbetning hade slutförts utan felmeddelanden inleddes en träning med modelkonfigurationen 2d och en annan med 3d_fullres. Detta gjordes genom att i terminalen skriva kommando som följande:

```
nnUNetv2_train -d DATASET_ID -c CONFIGURATION -f Fold
```

Där -f FOLD anger vilken av de fem korsvalideringsmapparna (0–4) som ska användas. Vi tränade på alla folds därför skrev vi -f all. Se bilaga 1 för exakta kommandon som användes. Efter att första träningsrundan avslutades genomfördes inferens på ytterligare 8 CT-bilder. De resulterande segmenteringarna granskades och justerades manuellt, varefter de adderades till det befintliga datasetets bilder och segmenteringsfiler. Detta utökade dataset användes sedan som grund för en andra träningsrunda med samma modellkonfiguration. En tredje och avslutande träningsrunda genomfördes enligt samma princip, där ytterligare 17 CT-bilder med tillhörande manuellt korrigerade segmenteringar inkluderades. Samtliga träningsrundor avslutades inom intervallet 200–300 epoker. Översikt över träningsrundor och bildtillägg presenteras i Tabell 1.

Tabell 1: Sammanställning av antal träningsbilder och tillägg per träningsrunda.

Träningsrunda	Antal träningsbilder	Nytt tillägg
1	5 bilder	-
2	8 bilder	+ 3 nya
3	25 bilder	+ 17 nya

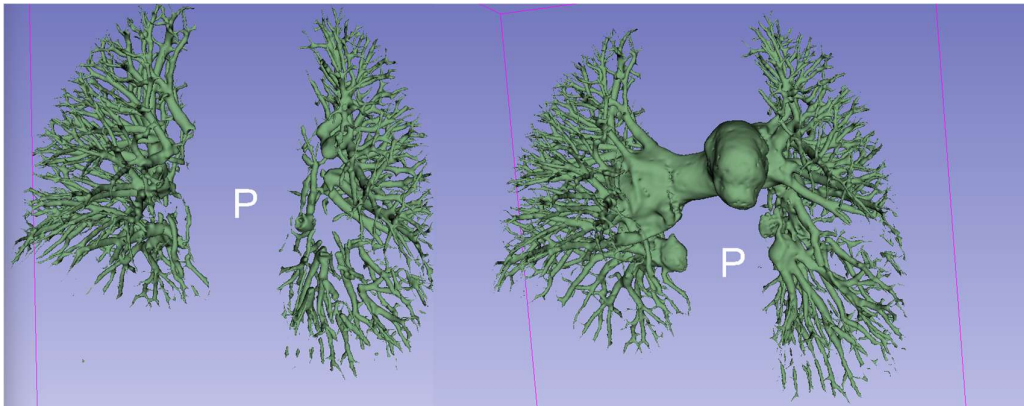
För att utvärdera modellens segmenteringsnoggrannhet användes DSC som kvantitativt mått. Två binära segmenteringsbilder användes, en från modellen och en som vi själva skapade som referens. I båda bilderna representerades kärl av värdet 1 och bakgrund av värdet 0. En manuell segmentering gjordes av 7 utvalda bilder i 3D Slicer. En jämförelse med de automatiskt genererade segmenteringarna från nnUNet efter den första, andra och tredje träningsrundan utfördes med hjälp av MATLAB R2023 (MathWorks, Natick, Massachusetts, USA). DSC-värden beräknades genom att ekvation 1 implementerades i ett MATLAB-skript, se bilaga 3.

4 Resultat

I detta avsnitt presenteras resultatet från tre på varandra följande träningsrundor med nnUNet. För varje runda redovisas både kvalitativa resultat i form av exempelbilder på segmenteringar, samt kvantitativa utvärderingar baserade på DSC. Dessutom presenteras grafer som illustrerar modellens tränings utveckling över tid samt en jämförelse mellan de tre rundorna. Om inte annat anges avser resultaten modellen tränad på bilder rekonstruerade med Br40-filtret.

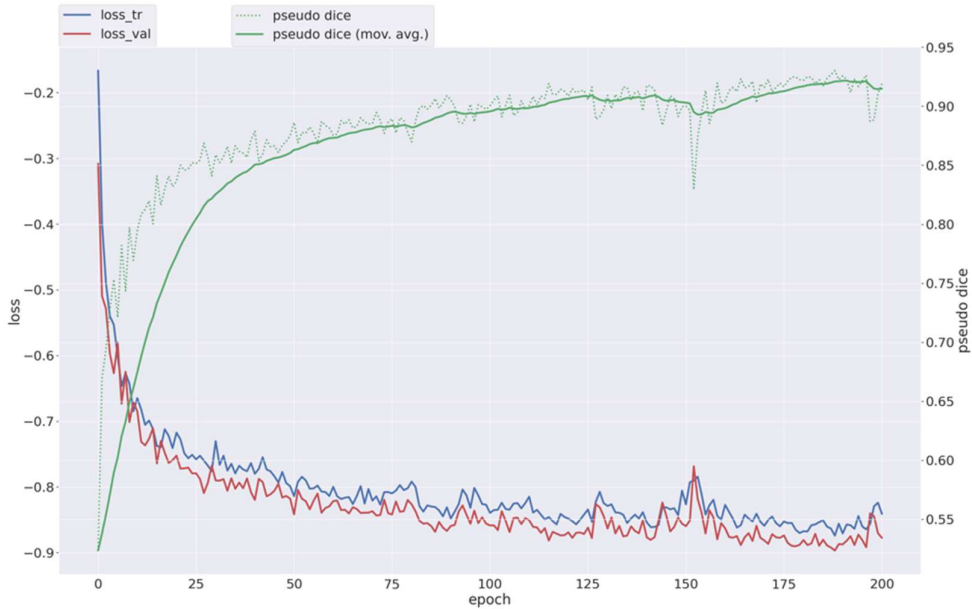
4.1 Träningsrundan 1 - Första modellträningen

Efter första träningsrunda bestående av fem segmenterade volymer genomfördes inferens baserade på denna träning. Figur 11 visar exempel på segmentering före och efter denna första träningsrunda. Innan träning (höger) är segmenteringen baserad på en modell som är förtränad på inspiratoriska CT-bilder tagna med kontrastmedel. Efter träning (vänster bild) framträder en mer sammanhängande segmentering.



Figur 11. Segmentering av bild 009 före och efter första träningsrundan. Vänster: Segmentering från modellen före träning. Höger: Segmentering genererad av modellen efter första träning. Modellen är orienterad i anatomisk riktning, där bokstaven "P" indikerar att strukturen betraktas bakifrån, posterior vy.

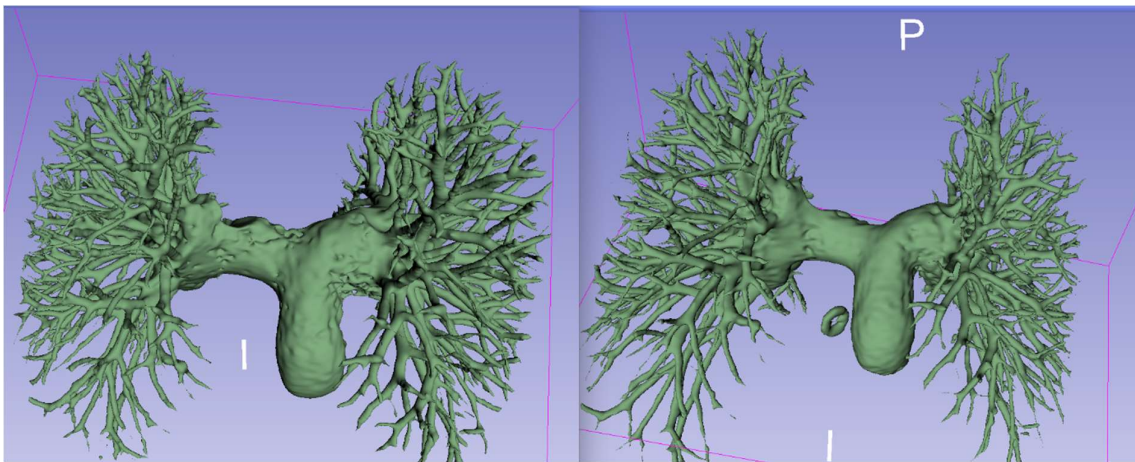
Vidare illustrerar figur 12 diagrammet som visar träningsförloppet för en nnUNet modell. Pseudo dice ökade snabbt under de inledande epokerna och stabiliserades därefter kring ett medelvärde på cirka 0,91. Samtidigt sjönk både tränings- och validationsförlusten (loss_tr och loss_val) snabbt i början.



Figur 12. nnUNet träningsförlopp. Diagram som visar träningsförloppet för en nnUNet-modell. Den gröna prickade linjen representerar det beräknade pseudo DSC-värdet efter varje epok, medan den gröna heldragna linjen visar ett glidande medelvärde för dessa. Den blå respektive röda linjen visar förlustvärdet (loss) på tränings- och validationsdata.

4.2 Träningsrunda 2 - Utökad träning med 8 bilder

Efter den första träningsrundan utökades träningsunderlaget till totalt åtta bilder. Resultatet visas i figur 13 här presenteras initial segmentering på bilder med Br40, där en skillnad i strukturens fyllning visas mellan segmenteringen före andra träningen och efter.



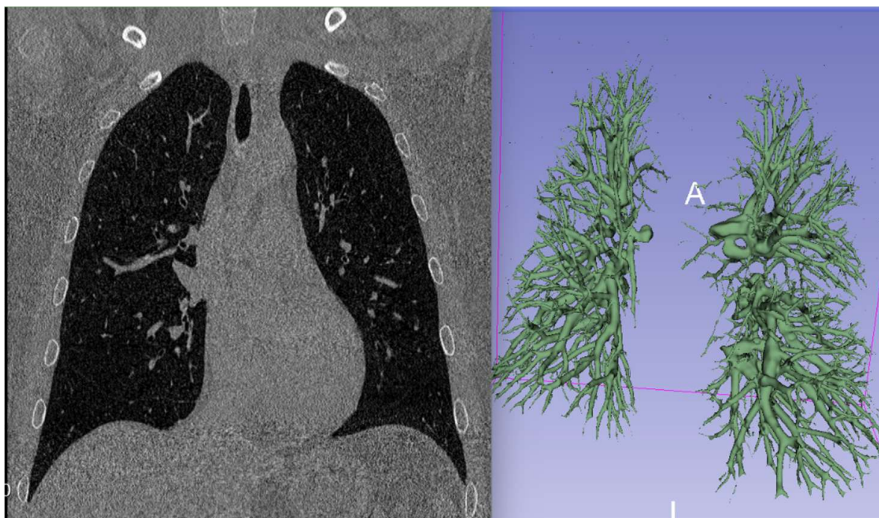
Figur 13: Jämförelse mellan nnUNets segmenteringsresultat efter första träningsrunda 001 (vänster) och andra träningsrundan 002 (höger). Efter den andra rundan ses en tydligare och mer sammanhängande segmentering av det pulmonella kärlträdet, med förbättrad detaljrikedom i perifera kärl. I den högra bilden syns dock en översegmentering i form av en cirkelformad struktur, som inte tillhör lungans anatomiska kärlträd. Bokstaven "P" i den högre bilden markerar posterior (baksidan av kroppen) och "I" markerar inferior (kroppens nedre del).

I figur 14 visas träningsförloppet för nnUNets andra träningsrunda över cirka 270 epoker. Båda kurvorna sjunker snabbt under de första epokerna. Efter cirka 50 epoker planar loss_tr och loss_val kurvorna ut och stabiliseras kring ett lägre värde. Pseudo Dice-koefficienten efter varje epok ökar snabbt i början och stabiliseras efterhand runt 0,85–0,87.



Figur 14: Träningsförlopp för nnUNets andra träningsrunda. Den gröna prickade linjen visar det uppskattade pseudo DSC-värdet per epok, medan den gröna heldragna linjen motsvarar ett glidande medelvärde. Den blå respektive röda kurvan representerar förlustvärdet (loss) för tränings- och validationsdata.

Resultatet av den automatiska segmenteringen efter den andra träningsrundan, på en bild rekonstruerad med filtret Br44, visas i figur 15. Segmenteringen fångar många av de perifera bronkiala strukturerna, men de centrala delarna saknas.



Figur 15. Automatisk segmentering från nnUNet baserad på andra träningsrunda. Vänster bild visar en axial CT-bild på lungan med filterinställningen Br44. Höger bild visar motsvarande 3D-renderade

segmentering av lungkärnen som genererats av modellen. Bokstaven "I" i 3D-renderingen står för inferior, kroppens nedre del, och "A" står för anterior, kroppens framsida. CT-bilden godkänd för användning av Jimmy Yu.

4.3 Träningsrunda 3 - Utökad modellträning med 25 bilder

I den tredje träningsrundan utökades träningsmaterialet genom att ytterligare 17 bilder segmenterades och inkluderades. Tillsammans med de åtta bilderna från träningsrunda 2 omfattade det nya datasetet nu totalt 25 bilder. En jämförelse mellan en segmentering som genererades av nnUNet efter andra träningsrundan och en efter den tredje visas i figur 16. Visuellt kan inga stora skillnader observeras mellan resultaten.



Figur 16: Visuell jämförelse av segmentering före och efter tredje träningsrundan. Vänster bild visar segmentering efter andra träningsrundan, medan höger bild visar segmentering efter den tredje. I båda bilderna syns en översegmentering i form av en cirkelformad struktur längst ned i bilden (inferiort), som inte motsvarar någon anatomisk del av lungorna. Bokstaven "I" markerar kroppens nedre riktning i 3D-renderingen från 3D Slicer.

Figur 17 visar träningsförloppet för den tredje träningsmodellen med nnUNet. Precis som i den första träningsrundan planar pseudo Dice-värdet ut kring 0,91–0,93. Förlustvärdena ($loss_{tr}$ och $loss_{val}$) minskar snabbt under de inledande epokerna och stabiliseras därefter på cirka -0,85.



Figur 17: Träningförloppet för modellens tredje träningsrunda. Den gröna prickade linjen representerar pseudo DSC-värdet efter varje epok och den gröna heldragna linjen visar glidande medelvärde. Den blå och röda linjen visar förlustvärdet på tränings- respektive valdiationsdata.

4.4 DSC Resultat

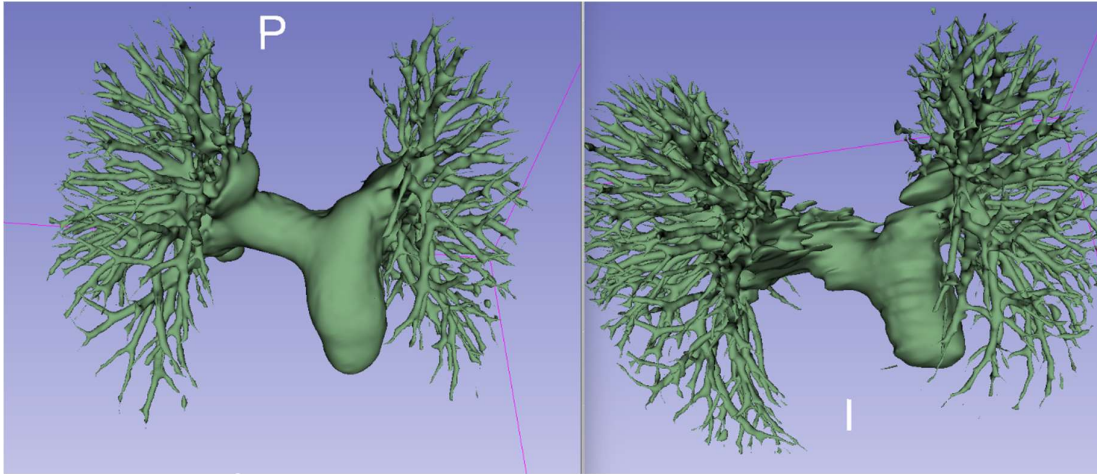
Tabell 2 visar DSC mellan manuellt korrigerade segmenteringar och modellens prediktioner för sju utvalda CT-bilder (bildnummer: 26, 32, 33, 36, 37, 38, 40). Jämförelserna gjordes efter varje träningsrunda (001 = första rundan, 002 = andra rundan, 003 = tredje rundan). Ett medelvärde av DSC har beräknats över de sju testbilderna för varje träningsrunda. Ett medelvärde har beräknats för varje träningsrunda baserat på de sju testbilderna. Resultaten visar att DSC-värdet ökar med varje träningsrunda.

Tabell 2: Resultatet av DSC beräkning för nnUNets segmentering efter träningsrundorna 001, 002 och 003.

Bildnr	26	32	33	36	37	38	40	Medelvärde
Runda 001	0.905	0.909	0.909	0.910	0.910	0.903	0.869	0.902
Runda 002	0.936	0.934	0.934	0.934	0.935	0.929	0.913	0.931
Runda 003	0.996	0.996	0.996	0.993	0.985	0.993	0.977	0.991

4.5 Resultat av de olika konfigurationsvalen.

Resultatet av de olika konfigurationsvalen visas i figur 18. Till vänster ses resultatet av en automatisk segmentering efter träning med 3D-konfigurationen, 3d_fullres, och efter träning med 2D-konfiguration, 2d, visas till höger.



Figur 18: Jämförelse mellan segmenteringsresultat med två olika konfigurationer. Vänster: resultat efter träning med 3d_fullres-konfiguration. Höger: resultat efter träning med 2d-konfiguration. Samma CT-bild användes i båda fallen. Bokstaven "P" i den här 3D-renderinge från 3D Slicer står för "Posterior", kroppens baksida. Bokstaven "I" står för "Inferior", kroppens nedre del.

5 Diskussion

I detta kapitel diskuteras studiens resultat i relation till tidigare forskning och den teoretiska bakgrunden kring nnUNets konfiguration, CT-teknik bildkvalitet. Dessutom diskuteras användning av AI för automatisk segmentering av pulmonella blodkärl. Syftet är att sätta resultaten i sammanhang och belysa deras relevans ur ett tekniskt och metodologiskt perspektiv.

5.1 Valet av modellens konfiguration

Ett viktigt metodval i detta projekt var att segmentera lungkärlen med en 3D-konfiguration snarare än med en 2D-konfiguration. Vårt beslut grundade sig på kärlets anatomiska egenskaper, de utgör komplexa strukturer som sträcker sig genom flera snitt i en CT-volym [25]. 3D-konfiguration gjorde att modellen följde kärlets riktning och utbredning mellan snitt som nämndes i rapportens bakgrundsdel.

Våra resultat bekräftar att detta val var välgrundat. Som det illustreras i Figur 18, där samma CT-bild har segmenterats med både 3D- och en 2D-konfiguration, är skillnaderna tydliga. Bilden till vänster (3D) uppvisar ett mer komplett kärlträd, där strukturerna följer en naturlig linje genom volymen. I kontrast syns i 2D-segmenteringen (höger) flera avbrott i kärstrukturen, särskilt i centrala delar av lungan där kärlet löper genom flera snitt. Detta anser vi antyder på att 2D-modellen, till skillnad från 3D-modellen, haft svårigheter att korrekt sammanfoga snitten.

Vår erfarenhet i detta projekt stämmer även överens med tidigare studie som visat att 3D-segmentering ofta är mer lämpade för att segmentera komplexa strukturer i medicinska bilder [26]. Även om denna specifika studie fokuserar på segmentering av hjärnstrukturer i magnetresonans (MR) bilder [26], bedömer vi att resultaten är överförbara även till vår tillämpning. Precis som hjärnstrukturer är lungkärl komplexa och tredimensionella, och därmed beroende av att modellen kan utnyttja den rumsliga kontexten mellan skivor [25]. Studien visar att det med 3d_fullres generellt uppnås högre noggrannhet än både vid 2d-konfigurationen och 3d_lowres, särskilt vid begränsad tillgång på träningsdata [26].

Dessutom bestämde vi oss att använda 3d_fullres i stället för alternativen 3d_lowres eller 3d_cascade_fullres. Valet baserades på önskad noggrannhet, resurskrav och praktisk tillämpning för den aktuella typen av datasetet. Enligt information från Jimmy Yu och Jonas Sartorius anses 3d_lowres generellt vara mindre lämplig, även om den reducerar minneskraven och snabbar upp träningen, men på bekostnad av upplösningen vilket gör det svårt att identifiera mindre strukturer i detalj. 3d_cascade_fullres kräver i sin tur längre körtid, eftersom modellen först körs med 3d_lowres och därefter förfinas med 3d_fullres. Baserat på information från Jimmy Yu och Jonas Sartorius valde vi att använda 3d_fullres, eftersom det enligt deras erfarenhet ger jämförbar noggrannhet med 3d_lowres och 3d_cascade_fullres, men med kortare körtid. Vi genomförde själva inga körningar med 3d_lowres eller cascade-varianterna, utan beslutet grundade sig på deras rekommendation samt vår tillgång till tillräckliga datorkapaciteter för att hantera 3d_fullres effektivt.

Vid träning specificerades även flaggan -f all. Detta valdes för att uppnå en robust modell som generaliserar bättre till nya data, då träning sker på hela datasetet i olika delmängder. En alternativ strategi hade varit att träna endast på enskilda folds (t.ex. -f 0), men detta kan leda

till en modell som är mer känslig för variationer i data.

5.2 Effekter av bildkvalitet på modellens prestanda

Kärl och omgivande vävnad i nativa bilder uppvisar ofta liknande attenuering. Därför var det en utmaning för nnUNet att genomföra en prediktion på dessa. Det var dessutom svårt att visuellt särskilja kärlen från den omgivande vävnaden. Kärlens gränser var otydliga och små mängder brus och artefakter påverkade bildkvaliteten avsevärt. Särskilt i området kring hjärtat försvårades tolkningen ytterligare av rörelseartefakter orsakade av hjärtats aktivitet vilket gjorde det extra svårt att urskilja intilliggande kärl.

Eftersom nnUNet initialt var tränad på bara inspiratoriska CT-bilder IVK, visade inferensen stora brister när den tillämpades på nativbilder. Detta syns tydligt i figur 11 (höger bild), där modellen inte lyckas identifiera strukturer i de centrala delarna av lungan. Dessutom uteblev korrekt segmentering i de perifera områdena, där kärlen är smalare och svårare att separera från omgivande vävnad. Redan efter den första träningsomgången förbättrades modellens prestanda avsevärt. Efter den andra träningen kunde nnUNet börja finsegmentera även i mer perifera delar som det syns i figur 13 där mindre kärl blev mer komplett segmenterade jämfört med resultatet efter den första träningen. Förbättringen i segmenteringskvalitet återspeglas också i träningsförloppet, figur 12, där pseudo Dice-värdet kontinuerligt ökar och når över 0,9 samtidigt som valideringsförlusten hålls låg. Detta tyder på att modellen blivit bättre på att generalisera utan att överanpassas till träningsdata.

Det fanns dock brister i modellens segmentering i de perifera område, en omständighet som kan förklaras av variationer i bildens hårdhet, det vill säga valet av rekonstruktionsfilter. Undersökningarna innehöll bilder rekonstruerade med både det hårdare filtret Br44 och Br40. Vid automatisk segmentering av en bild rekonstruerad med Br44 filter kunde modellen inte identifiera strukturer i de centrala delarna av lungan, och kärlsegmenteringen visade tydliga brister. Kärlen framstod som osammanhängande, med många avbrott och felaktiga segmenteringar, se figur 15. Eftersom detta påverkade nnUNets förmåga att behandla bilderna på ett enhetligt sätt, valde vi att inte använda Br44 bilderna från undersökningarna.

5.3 Datamängdens betydelse

Tidigare studier har visat att datasetets egenskaper, inklusive storlek och variation, har en tydlig inverkan på modellens prestanda vid segmentering [7]. I detta projekt tränades modellen i tre steg, där antalet träningsbilder successivt ökade från 5 till totalt 25 bilder. Resultaten för DSC, som illustreras i tabell 2, visar en tydlig förbättring mellan varje träningsrunda, från ett medelvärde på 0,90 i första rundan till 0,93 i den andra och slutligen 0,99 i den tredje.

Den förbättring vi observerade mellan träningsrundorna kan, enligt vår bedömning, förklaras inte enbart av den ökade mängden träningsdata, utan även den större variationen i bildmaterialet vi använde. När vi successivt la till fler patienter till träningen märkte vi att modellen i högre grad exponerades för olika typer av kärlstrukturer och anatomiska skillnader. Det vi såg i våra resultat stämmer väl överens med vad tidigare forskning har visat, att både träningsvolym och variation i datasetet är viktiga faktorer för att uppnå hög prestanda vid medicinsk bildsegmentering [7].

En tydlig begränsning i projektet är att det dataset vi använde enbart innehöll CT-bilder från

patienter utan synliga patologiska förändringar. Det innebär alltså att nnUNet inte har exponerats för kärlstrukturer som påverkas av sjukdom, exempelvis förgreningar vid lungfibros. I efterhand bedömer vi att om vi haft tillgång till ett större och mer varierat dataset, där även patologiska fall ingick, hade detta sannolikt förbättrat nnUNets förmåga att generalisera till nya data. Detta styrks också av tidigare forskning som visat att datamängdens sammansättning, särskilt inkluderingen av patologiska fall, har en stor påverkan på modellens prestanda och förmågan att generalisera [27]. En sådan förbättring hade också kunnat öka modellens kliniska användbarhet, särskilt i diagnostiska sammanhang där det enligt Jimmy Yu är avgörande att kunna skilja mellan friska och sjuka kärlstrukturer.

5.4 Begränsningar med DSC

Även om DSC är ett etablerat mått för att kvantifiera överlappningen mellan två segmenteringsmasker [21], har vi i detta projekt identifierat flera begränsningar som gör att vi anser att DSC är otillräckligt som ett ensamt mått för att fullt ut kunna bedöma segmenteringskvalitet för lungkärl. Under projektets gång observerade vi att DSC inte alltid ger en heltäckande bild av modellens faktiska prestanda. Trots höga DSC-värden i den tredje träningsrundan, se tabell 2, observerades att tunna perifera kärl ofta missades. Detta innebär att modellen kunde ha stora strukturella brister, exempelvis avbrott i kärlträdet, utan att detta i större utsträckning påverkade DSC-värdet.

Dessa observationer överensstämmer med tidigare forskning. Reinke et al. [28], visade att DSC är dåligt på att fånga topologiska fel som exempelvis avbrott i strukturer, och att det därmed kan ge falsk bild av modellens prestanda i segmenteringsuppgifter som kräver strukturell integritet. På liknande sätt argumenterar Karimi och Salcudean [29] för att DSC bör kompletteras med andra mått, som exempelvis Hausdorff-avstånd.

Mot denna bakgrund av dessa begränsningar anser vi att DSC bör tolkas med viss försiktighet och inte användas ensamt. Istället anser vi att det bör kompletteras med utvärderingsmått som är särskilt lämpade för att fånga segmenteringskvalitet i perifera och tunna kärlgrenar.

5.5 Feltyper och felkällor i segmenteringen

Trots att processen för att skapa referenssegmenteringar utgjorde en central styrka i projektet, finns det risker som bör beaktas. En av dessa är risken för systematiska tolkningar av anatomiska strukturer. Det innebär att vissa kärl tolkas på samma sätt varje gång, till exempel att små kärl ofta bedöms höra till eller inte höra till kärlträdet. Detta kan i sin tur påverka hur vissa kärl segmenteras eller utelämnas.

Våra resultat visade exempelvis att modellen ibland översegmenterar kärlstrukturer. I figur 16 syns en översegmentering av specifika kärlområden. Trots att pseudo Dice-värdet kontinuerligt förbättrades och översteg 0.90 efter cirka 150 epoker (figur 14 och figur 17) och trots låg valideringsförlust, kvarstod tydliga segmenteringsfel efter både träningsrunda 2 och 3. Detta visar att höga DSC-värden inte alltid speglar god anatomisk precision, särskilt i mer komplexa strukturer som lungkärl. Om sådana fel förekommer konsekvent finns en risk att modellen lär sig dessa mönster och återger dem i sina förutsägelser. Problemet kan förstärkas av variationer i kvalitet och metodik mellan olika referenssegmenteringar. Modeller som tränas på ett mindre antal fall är särskilt känsliga för denna typ av systematiska fel, eftersom enstaka avvikelser får större genomslag jämfört med träning på ett

större och mer varierat dataset.

6 Slutsats

Automatisk segmentering med hjälp av nnUNet är möjlig även på expiratoriska CT bilder tagna utan kontrastmedel. Studien visar att modellen har god inlärningsförmåga och uppnår förbättrad prestanda redan efter ett fåtal träningsomgångar. Med ytterligare optimering och validering har metoden potential att implementeras i klinisk praxis för att underlätta analys av nativbilder där kontrastmedel inte kan användas.

7 Referenser

- [1] R. S. of N. A. (RSNA) and A. C. of Radiology (ACR), "Chest CT", Radiologyinfo.org. Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://www.radiologyinfo.org/en/info/chestct>
- [2] E. A. Kazerooni, "High-Resolution CT of the Lungs", *Am. J. Roentgenol.*, vol. 177, nr 3, s. 501–519, sep. 2001, doi: 10.2214/ajr.177.3.1770501.
- [3] G. Litjens *m.fl.*, "A survey on deep learning in medical image analysis", *Med. Image Anal.*, vol. 42, s. 60–88, dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [4] P. Manava *m.fl.*, "Improving Contrast Enhancement in Pulmonary CTA: The value of breathing maneuvers", *Eur. J. Radiol. Open*, vol. 7, s. 100280, jan. 2020, doi: 10.1016/j.ejro.2020.100280.
- [5] P. Raczeck *m.fl.*, "Influence of Respiratory Position on Contrast Attenuation in Pulmonary CT Angiography: A Prospective Randomized Clinical Trial", *Am. J. Roentgenol.*, vol. 206, nr 3, s. 481–486, mar. 2016, doi: 10.2214/AJR.15.15176.
- [6] P. Khomduean *m.fl.*, "Segmentation of lung lobes and lesions in chest CT for the classification of COVID-19 severity", *Sci. Rep.*, vol. 13, s. 20899, nov. 2023, doi: 10.1038/s41598-023-47743-z.
- [7] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, och K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation", *Nat. Methods*, vol. 18, nr 2, s. 203–211, feb. 2021, doi: 10.1038/s41592-020-01008-z.
- [8] N. K. Abdulkareem, S. I. Hajee, F. F. Hassan, I. K. Ibrahim, R. E. H. Al-Khalidi, och N. A. Abdulqader, "Investigating the slice thickness effect on noise and diagnostic content of single-source multi-slice computerized axial tomography", *J. Med. Life*, vol. 16, nr 6, s. 862–867, juni 2023, doi: 10.25122/jml-2022-0188.
- [9] L. W. Goldman, "Principles of CT: Radiation Dose and Image Quality", *J. Nucl. Med. Technol.*, vol. 35, nr 4, s. 213–225, dec. 2007, doi: 10.2967/jnmt.106.037846.
- [10] A. Cruickshank, "Kernel (image reconstruction for CT) | Radiology Reference Article | Radiopaedia.org", Radiopaedia. Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://radiopaedia.org/articles/kernel-image-reconstruction-for-ct-1>
- [11] "Ultrahigh-Resolution Photon-Counting CT in Cadaveric Fracture Models: Spatial Frequency Is Not Everything - PMC". Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://www.mdpi.com/2075-4418/13/10/1677>
- [12] "CT Radiographic Techniques | Radiology | SUNY Upstate Medical University". Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://www.upstate.edu/radiology/education/rsna/ct/technique.php>
- [13] "Photoelectric effect | Radiology Reference Article | Radiopaedia.org". Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://radiopaedia.org/articles/photoelectric-effect>
- [14] "AI inom hälso- och sjukvården", Region Stockholm. Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://www.regionstockholm.se/om-region-stockholm/sa-jobbar-vi-med-ai/ai-inom-halso--och-sjukvarden/>
- [15] M. Aslam, "The Role of Artificial Intelligence in Pulmonary Medicine: Transforming Diagnosis, Treatment, and Research", *J. Adv. Lung Health*, vol. 4, nr 3, s. 126–127, sep. 2024, doi: 10.4103/jalh.jalh_38_24.
- [16] "Så fungerar luftvägar och lungor - 1177". Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://www.1177.se/liv--halsa/sa-fungerar-kroppen/luftvagar-och-lungor/>
- [17] "nnUNet/documentation/how_to_use_nnunet.md at master · MIC-DKFZ/nnUNet", GitHub. Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://github.com/MIC->

- DKFZ/nnUNet/blob/master/documentation/how_to_use_nnunet.md
- [18] F. Isensee, J. Petersen, S. A. A. Kohl, P. F. Jäger, och K. H. Maier-Hein, "nnUNet: Breaking the Spell on Successful Medical Image Segmentation", arXiv.org. Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://arxiv.org/abs/1904.08128v1>
- [19] *MIC-DKFZ/nnUNet*. (22 maj 2025). Python. MIC-DKFZ. Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://github.com/MIC-DKFZ/nnUNet>
- [20] *MIC-DKFZ/nnUNet*. (22 maj 2025). Python. MIC-DKFZ. Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://github.com/MIC-DKFZ/nnUNet>
- [21] A. A. Taha och A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool", *BMC Med. Imaging*, vol. 15, nr 1, s. 29, aug. 2015, doi: 10.1186/s12880-015-0068-x.
- [22] "Dice score". Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://oecd.ai/en/metrics/dice-score>
- [23] "The Difference Between Dice and Dice Loss - PYCAD - Your Medical Imaging Partner". Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://pycad.co/the-difference-between-dice-and-dice-loss/>
- [24] "dice - Sørensen-Dice similarity coefficient for image segmentation - MATLAB". Åtkomstdatum: 22 maj 2025. [Online]. Tillgänglig vid: <https://se.mathworks.com/help/images/ref/dice.html>
- [25] H. Shikata, G. McLennan, E. A. Hoffman, och M. Sonka, "Segmentation of Pulmonary Vascular Trees from Thoracic 3D CT Images", *Int. J. Biomed. Imaging*, vol. 2009, s. 636240, 2009, doi: 10.1155/2009/636240.
- [26] A. Avesta, S. Hossain, M. Lin, M. Aboian, H. M. Krumholz, och S. Aneja, "Comparing 3D, 2.5D, and 2D Approaches to Brain Image Auto-Segmentation", *Bioengineering*, vol. 10, nr 2, s. 181, feb. 2023, doi: 10.3390/bioengineering10020181.
- [27] T. Ekman, A. Barakat, och E. Heiberg, "Generalizable deep learning framework for 3D medical image segmentation using limited training data", *3D Print. Med.*, vol. 11, nr 1, s. 9, mar. 2025, doi: 10.1186/s41205-025-00254-1.
- [28] A. Reinke *m.fl.*, "Understanding metric-related pitfalls in image analysis validation", *Nat. Methods*, vol. 21, nr 2, s. 182–194, feb. 2024, doi: 10.1038/s41592-023-02150-0.
- [29] D. Karimi och S. E. Salcudean, "Reducing the Hausdorff Distance in Medical Image Segmentation With Convolutional Neural Networks", *IEEE Trans. Med. Imaging*, vol. 39, nr 2, s. 499–513, feb. 2020, doi: 10.1109/TMI.2019.2930068.

Bilaga 1: nnUNet kommando i Terminal

För att aktivera nnUNet:

```
conda activate nnunet-env
```

För att köra förbearbetning (preprocessing):

Med 3d_fullres konfiguration:

```
nnUNetv2_plan_and_preprocess -d 1 -c 3d_fullres --verify_dataset_integrity
```

Med 2d konfiguration:

```
nnUNetv2_plan_and_preprocess -d 1 -c 3d_fullres --verify_dataset_integrity
```

För att köra träningen:

Med 3d_fullres konfiguration:

```
nnUNetv2_train 1 3d_fullres all
```

Med 2d konfiguration:

```
nnUNetv2_train 1 2d all
```

För automatisk segmentering (m.a.o inferens):

Med 3d_fullres konfiguration:

```
nnUNetv2_predict -i nnunet_input -o nnunet_output -c 3d_fullres -f all --save_probabilities
```

Med 2d konfiguration:

```
nnUNetv2_predict -i nnunet_input -o nnunet_output -c 2d -f all --save_probabilities
```


Bilaga 2: Konfigurationsfilen dataset.json för träning i nnUNet

Koden i VSC för att köra träningen:

```
{
  "channel_names": {
    "0": "CT"
  },
  "labels": {
    "background": 0,
    "vessel": 1
  },
  "file_ending": ".nrrd",
  "numTraining": 5,
}
```


Bilaga 3: MATLAB KOD för DSC

```
1  clc; clear all;
2
3  % Filvägarna till NRRD-filerna
4  nnunet_seg_path = 'C:\Users\46790\Desktop\inferens3\case_036.nrrd';
5  user_seg_path = 'C:\Users\46790\Desktop\FullstandigtSeg\case_036_fullständigt.nrrd';
6
7  % Läser filerna
8
9  nnunet_seg = nrrdread(nnunet_seg_path);
10 user_seg = nrrdread(user_seg_path);
11
12 % Omvandla segmenteringarna till binära bilder
13 nnunet_seg_bin = nnunet_seg > 0;
14 user_seg_bin = user_seg > 0;
15
16 % Beräknar Dice-koefficienten
17 intersection = nnunet_seg_bin & user_seg_bin; % skärning mellan de två segmenteringarna
18 dice = 2 * sum(intersection(:)) / (sum(nnunet_seg_bin(:)) + sum(user_seg_bin(:)));
19
20 % Resultat:
21 disp(['DSC: ', num2str(dice)]);
22
23
```