

Sjálfvirk greining kransæðasjúkdóms út frá CT myndum af hjartanu

RAF620M: Inngangur að vélrænu námi og gervigreind

Unnið af:

Ellert Blær Guðjónsson (ebg35@hi.is)
Róbert Tumi Guðmundsson (rtg5@hi.is)

Dagsetning: 21. apríl 2025

Kennarar: Magnús Örn Úlfarsson
Sverrir Jan Norðfjörð



EFNISYFIRLIT

Ágrip verkefnis	3
1 Inngangur og bakgrunnur	4
2 Núverandi þekkingarástand	6
3 Bráðabirgðaniðurstöður og gögn	7
3.1 CNN módel sem baseline	9
3.2 InceptionV3	10
3.3 Kynslóðaskipt CNN	11
3.4 ResNet50	12
3.5 Kynslóðaskipt dýpkandi CNN	13
3.6 Samantekt	15
4 Rannsóknaáætlun og aðferðafræði	16
5 Stjórnun, samstarf og samvinna	17
6 Niðurstaða	18
Heimildir	19

ÁGRIP VERKEFNIS

Kransæðasjúkdómur (CAD) er helsta dánarorsökin á heimsvísu og stafar af uppsöfnuðum efnum (t.d. fitu og steinefnum) í kransæðum, sem takmarkar blóðflæði til hjartans. Ein algengasta aðferðin til að meta og greina CAD kallast "Coronary Computed Tomography Angiography" (CCTA) og er tölvusneiðmyndataka á hjartanu þar sem skuggaefni er notað til að draga fram æðar á myndunum. Til eru margar myndgreiningaraðferðir fyrir CAD eins og t.d. segulómun (MRI) og býður hún upp á hærri myndgæði, en CCTA er þó oft valið vegna hagkvæmni og er því að verða sífellt algengari við greiningu CAD.

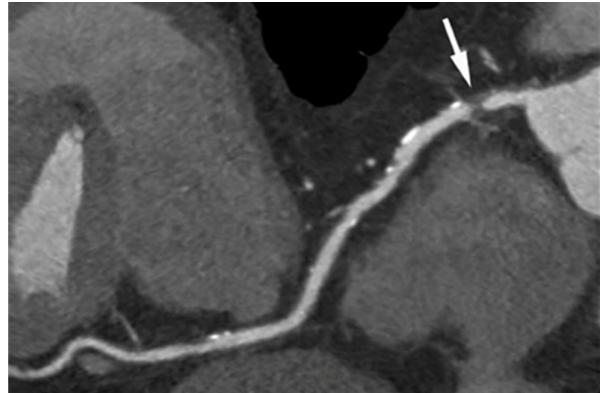
Enn í dag er algengt að yfirferð CCTA mynda fari handvirkta fram og getur það verið virkilega tímafrekt. Það er t.d. vegna slæmra hreyfingarmyndgalla sem eru algengir á CCTA myndum. Nýlegar framfarir í gervigreind og vélrænu námi við greiningu á CAD út frá CCTA myndum lofa góðu, en þær lausnir sem búið er að þróa er ekki enn búið að innleiða á stórum skala vegna áhyggja um áreiðanleika. Markmið þessarar rannsóknar er að þróa nákvæmt og skilríkt gervigreindarmódel til að greina og meta CAD úr CCTA-myndum sem yrði síðar hægt að aðlaga að þörfum heilbrigðisstofnana til að auka innleiðingarmöguleika. Niðurstöður sýna að módel okkar gefa gott AUC og greiningarnákvæmni, eða allt að 0.96 og 96%, sem benda til að módelið gæti straumlínulagað verklag hjartalækna og röntgentækna með því að minnka álag og í leiðinni aukið greiningarnákvæmni CAD með CCTA myndgreiningu.

1 INNGANGUR OG BAKGRUNNUR

Hjarta- og æðasjúkdómar hafa í fjölmörg ár verið helsta dánarorsökin um allan heim og fjöldi dauðsfalla og greininga fer vaxandi með hverju ári. Kransæðasjúkdómur (e. Coronary Artery Disease, CAD) er algengasti hjartasjúkdómurinn og er talið að einn af hverjum sjö látist vegna sjúkdómsins. Fyrir COVID-19 faraldurinn árið 2019 hafði CAD verið helsta dánarorsökin í a.m.k. 30 ár en nú er talið að um 250 milljónir lifi með CAD [1].

CAD er æðaþrengsli í kransæðunum sem bera súrefni til hjartans. Á löngum tíma safnast upp fita í kransæðunum sem getur kalkast og verður til þess að blóð flæðir ekki eins greitt til hjartans og getur það valdið hjartadrep. Uppsöfnun fitu í æðum kallast "atherosclerosis" og fitusstellurnar "atheroma". Algeng einkenni CAD eru m.a. ógleði, brjóstverkir, og öndunarerfiðleikar, en margir finna ekki fyrir neinum einkennum þrátt fyrir talsverða stíflu í kransæðum [2].

Ein algengasta og öflugasta myndgreiningaraðferðin sem er notuð til að greina CAD kallast á ensku "Coronary Computed Tomography Angiography" en þar er skuggaefni sprautað í blóðrás einstaklingsins sem á að mynda og CT skanni snýst um hjartað til að taka röntgenmyndir frá mörgum mismunandi sjónarhornum [4]. Á myndunum sjást æðarnar greinilega sem ljósgráar greinar og uppsöfnuð fita má sjá sem dökkgráir flekkir á veggjum æðanna.



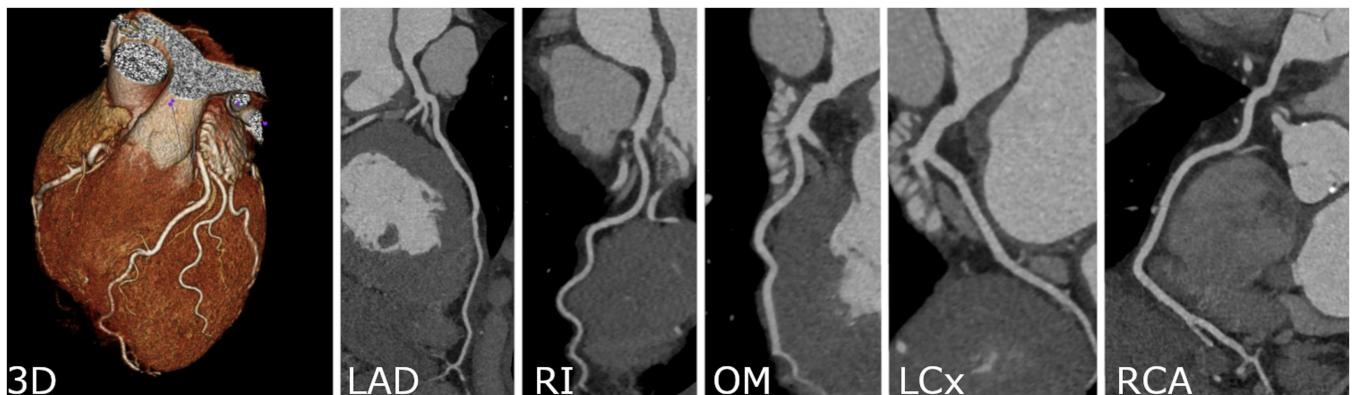
Mynd 1: CCTA mynd sem sýnir mikil æðaþrengsli [3]

Fituskellur sem hafa kalkast sjást mjög vel á CT myndum sem bjartir ljósblossar og er algengt að fara í CT myndatöku án skuggaefnis til að skoða kölkun sérstaklega ef enginn grunur er um CAD. Úr svoleiðis myndatöku fæst svokallað kalkskor (e. coronary calcium score, CCS) sem gefur hugmynd um líkur á kransæðastíflu en er ekki hægt að nota til að greina CAD vegna þess að blóðflæði og æðaþrengsli eru mjög óskýr. Hins vegar hafa verið þróaðar aðferðir til að meta CCS út frá CCTA myndum. CCS er skipt í áhættustig, t.d. ef CCS er 1-10 er mjög lítil hætta á CAD, og ef það er >400 er mjög mikil hætta á CAD [5].

Myndgreiningaraðferðin sem hefur lengi verið sögð best fyrir CAD greiningu kallast "invasive coronary angiography" (ICA) og er mjög svipað CCTA, nema ICA þarfnað aðaleggs til að sprauta skuggaefni í blóðrásina og er því ekki inngrípalaust eins og CCTA. Þegar CAD er greint með ICA er hægt

að setja upp stent í kransæðunum í sömu aðgerð til að auka aftur blóðflæðið. Það sem CCTA hefur fram yfir ICA er hagkvæmni og einfaldleiki. Skv. rannsókn frá 2020 var meðalkostnaður CAD greiningar fyrir ICA um 2.7 sinnum dýrar heldur en greining með CCTA (\$836 vs. \$2284). Einnig hafa nýlegar rannsóknar sýnt fram á að notkun CCTA sem fyrsta greiningaraðferð minnki frekar líkur á t.d. hjartaáföllum og hjartadrepi heldur en ICA [6].

Helsti galli CCTA greiningar er að úrlestur myndanna fer almennt fram handvirkta og er frekar flókinn. Úrlesturinn þarfnað mikillar sérhæfingar þar sem á myndunum eru gjarnan miklir myndgallar vegna hreyfinga hjartans og lélegra birtuskila röntgenmyndanna. Greiningarferlið getur því verið mjög tímafrekt og dregur því verulega úr skilvirkni heilbrigðisstarfsmanna [7].



Mynd 2: CCTA myndir af kransæðum settar saman til að mynda 3D módel af hjartanu [8]

Par sem áhuginn á CCTA greiningu er mikið að aukast [4] er þess virði að endurskoða úrlestraraðferðirnar sem eru nú í notkun. Markmið þessa rannsóknarverkefnis er að þróa nákvæmt og skilríkt gervigreindarmódel sem getur sjálfvirkt greint kransæðasjúkdóm út frá CCTA myndum og væri hægt að innleiða í núverandi verkferla heilbrigðisstofnana.

Rannsóknarspurningin sem verkefnið leitast við að svara er: "Hversu nákvæmt og áreiðanlegt getur gervigreindarlíkan verið við greiningu á CAD út frá CCTA myndum, og hvernig má best innleiða slíkt líkan í heilbrigðisstofnum?" Par sem CCTA greining er hvað mest notuð til þess að útiloka CAD hjá einstaklingum sem hafa CAD einkenni en eru ekki í miklum áhættuhóp [9], þá teljum við að sé hægt að þróa einfalt líkan sem tekur inn CCTA myndir af kransæðum og skilar annað hvort jákvæðu eða neikvæðu svari. Til að kanna hvernig best væri að innleiða líkanið í heilbrigðisstofnum um yrði líkanið síðar þróað frekar til að aðlaga að þörfum heilbrigðisstofnana, og yrðu rannsóknir framkvæmdar til að kanna innleiðingarmöguleika og traust á módelið. Það er mikil þörf á áreiðanlegu og einföldu CAD greiningarmódeli og gæti það haft gríðarleg áhrif m.a. á greiningarnákvæmni algengustu dánarorsakar í heimi og verkferla heilbrigðisstofnana um allan heim.

2 NÚVERANDI PEKKINGARÁSTAND

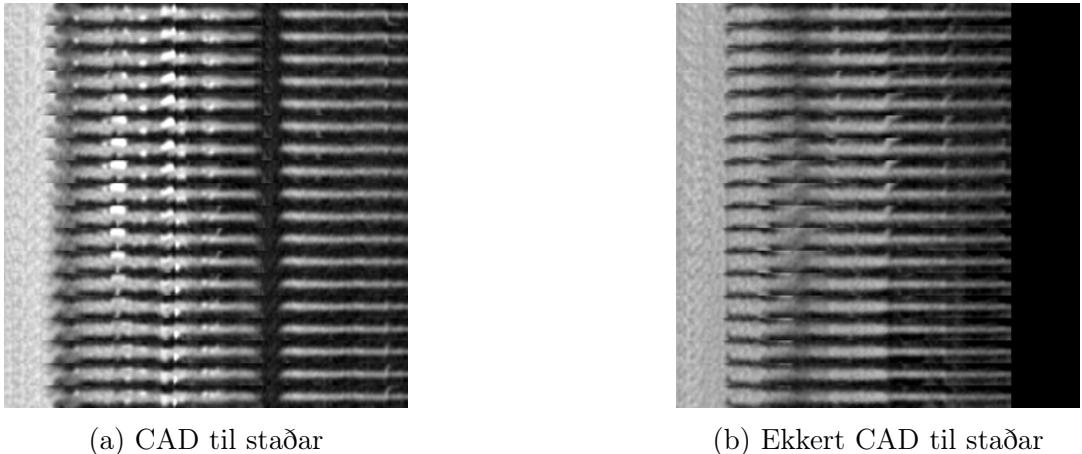
Innleiðing gervigreindar í CAD greiningu með CCTA myndum er mjög vinsælt viðfangsefni og hefur mikið verið rannsakað síðastliðin ár. Það eru fjölmörg dæmi um rannsóknarverkefni sem hafa þróað lausnir á þessu verkefni og margar þeirra lofa mjög góðu. Til dæmis hefur ný lausn frá 2025 sem metur æðaþrengsli í kransæðum sýnt 0.91 area under curve (AUC) gildi á einstaklingsgreiningu CAD á móti AUC 0.77 fyrir reynda úrlesara [7]. Niðurstöður rannsókna eru alls konar og hafa t.d. sýnt AUC 0.782, 0.897, og 0.95 [9–11] en það er algeng leið til að meta frammistöðu líkana í slíkum rannsóknum. Rannsóknir hafa líka náð mjög góðri almennri greiningarnákvæmni, t.d. 80%, 89%, og 90% með specificity allt að 98% og sensitivity að 91% [12]. Einnig hefur verið sýnt fram á að notkun gervigreindar í greiningu CAD með CCTA myndum getur minnkað geislunarskammtinn sem þarf til að mynda sjúklinga, sem er mjög mikilvægt vegna þess að CT notar jónandi geislun til að taka myndir og getur haft alvarlegar afleiðingar eftir margar myndatökur [13]. Í þessu verkefni höfum við það að markmiði að þróa lausn sem mun dýpka skilning og traust á notkun gervigreindar til greiningar á CAD, og stuðla að þeirri hröðu þróun sem er að eiga sér stað stað í CCTA myndgreiningum.

Prátt fyrir miklar rannsóknir á þessu sviði og efnilegar niðurstöður hefur úrlestur CCTA mynda haldist handvirkur á flestum stöðum. Þetta er að miklu leyti vegna þess að núverandi lausnir standast ekki klínískar kröfur og er þörf á frekari þróun og rannsóknum á sjálfvirkum greiningartólum. Til að geta innleitt svona lausn í heilbrigðisstofnun verður að tryggja trúverðugleika t.d. með því að nota gögn frá mismunandi stöðum, og hafa lausnina einfalda þannig að hugbúnaður sem er í notkun geti keyrt hana [9]. Lausnin okkar mun leggja áherslu á einfaldleika og nákvæmni en við notum til þess földunarnet (e. convolutional neural network, CNN). Við viljum prófa margar mismunandi nálganir á verkefninu til að finna hvaða aðferð skilar bestu niðurstöðunum, t.d. nota forþjálfuð módel eins og InceptionV3 og ResNet, eða búa til CNN frá grunni til að gera hana sem einfaldasta.

CNN er tegund af djúpum tauganetum sem geta greint myndræn gögn á mjög skilvirkan hátt og henta því læknisfræðilegri myndgreiningu mjög vel. Í grunninn virkar CNN þannig að myndir eru settar í gegnum margar síur sem gera módelinu kleift að læra og greina hver mikilvægustu mynstur myndanna eru, fyrst er lært á einföld einkenni eins og t.d. brúnir og form, en svo stigmagnast flækjustig lærðómsins með fjölda laga og loks getur módelið greint t.d. hvar kransæðastífla er til staðar á CCTA mynd af hjartanu. Sýnt hefur verið fram á gæði CNN módela á fjölda sviða læknisfræðilegrar myndgreiningar og hefur verið notað t.d. til að greina brjóstakrabbaein, lungnasjúkdóma, og heilablöðfall [14].

3 BRÁÐABIRGÐANIÐURSTÖÐUR OG GÖGN

Gögnin sem við notum til að þjálfa CNN módelin okkar eru úr gagnasetti sem heitir: *Image dataset for a CNN algorithm development to detect coronary atherosclerosis in coronary CT angiography* [15]. Hver mynd sýnir eina kransæð sem hefur verið einangruð úr CCTA mynd og strekkt þannig að hún nái þvert yfir myndina. CT tæki snúast í kringum þá sem myndir eru teknar af, og taka því myndir af sama hlutnum frá mörgum mismunandi sjónarhornum. Hver mynd í þessu gagnasetti sýnir eina æð frá 18 mismunandi sjónarhornum, þ.e. $\frac{180^\circ}{18} = 10^\circ$ fresti, þar sem hverju sjónarhorni er staflað á hvort annað, sú framsetning kallast "mosaic projection view" (MPV). Til að auka fjölbreytileika í gögnunum er röð sjónarhornanna á stórum hluta myndanna breytt á mismunandi hátt, þannig að myndafjöldi verður um 5000. Ekki eru allar myndirnar 'augmentaðar' á þennan hátt, t.d. er engin breyting á validation myndunum. Hér fyrir neðan má sjá dæmi af tveimur myndum úr gagnasettinu, eitt heilbrigtrt tilfelli og annað með CAD.



Mynd 3: Tvær myndir úr gagnasettinu sem er notað til að þjálfa módelin okkar.

Myndin til vinstri sýnir mjög greinilega kölkun með hvítum ljósblossum og stíflu með dökkri rönd sem takmarkar blóðflæðið út æðina. Til hægri sést engin greinileg stífla eða kölkun, en takið eftir að á þessari mynd er æðin ekki nógu löng til að ná yfir alla myndina, þannig að fyllt er í eyðuna með svörtu.

Myndirnar í gagnasettinu eru 299×299 px á .png formatti. Gagnasettið inniheldur myndir af 500 einstaklingum, þar sem helmingur þeirra er greindur með CAD. Bíuð er að skipta einstaklingum í þjálfun/validation/prófun á þennan máta: 300/100/100. Haldið er 50% hlutfalli heilbrigðgra og óheilbrigðra einstaklinga fyrir hvert tilfelli. Hins vegar er framkvæmd meiri 'augmentation' á heilbrigðu einstaklingana í prófunargögnunum og þannig fást 1066 heilbrigðar myndir og 125 óheilbrigðar, og CAD hlutfallið verður þá $\frac{125}{1191} = 10.5\%$, en þetta er nærlaunverulegri skiptingu en 50%. Þeir sem eru taldir vera líklegri til að vera með CAD fara frekar í ICA þar sem hægt er að setja upp stent í leiðinni [6].

Í þessu gagnasetti eru myndirnar aðeins flokkaðar í jávkætt eða neikvætt og engin frekari úrgreining gerð. Í öðrum rannsóknum er hins vegar oft greindur alvarleiki stíflunnar með því að meta hve mikið þvermál æðarinnar hefur minnkað vegna uppsafnaðrar fitu, t.d. gæti alvarlegt CAD verið $\geq 50\%$ þvermálsbreyting. Einnig eru engar upplýsingar um kalkskor eða þéttleika fitunnar. Því er okkur erfitt að þjálfa módel sem segja til um nokkrar aðrar upplýsingar en það eitt hvort á myndinni sé að sem er jákvæð eða neikvæð fyrir CAD og er það ekki tilgangurinn með þessari rannsókn. Að auki höfum við ekki fundið önnur eins gagnasett; svoleiðis vill til að það er þekktur skortur á opnum og stórum CCTA gagnasettum til greiningar á CAD á netinu, enda er það stór ástæða þess að gervigreindarmódel hafa ekki verið innleidd á stórum skala [16].

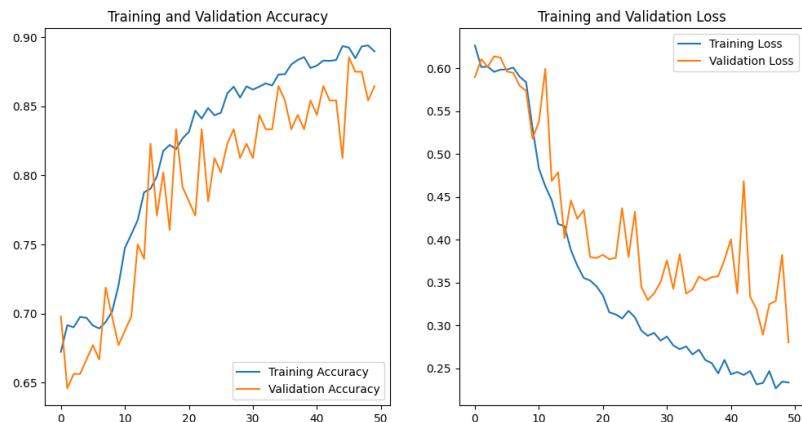
Önnur gangasett eru samt til og gætu verið notuð til þess að útvíkka frekari model til nánari greiningar á CAD. CADICA [17] er safn sem vert er að minnast á; það er flott gagnasafn sem inniheldur margar ICA myndir, en við viljum hins vegar notast við CCTA, svo það nýtist ekki í þessa rannsókn. Þegar niðurstöður CCTA eru ekki afgerandi eða greinilegar er ICA notað; við viljum einblína á CCTA sem er notað við fyrstu greiningu. Einnig er vert að minnast á gagnasafnið ImageCAS [18], sem lofar afar góðu og er fáanlegt á Kaggle, en inniheldur þó aðeins upplýsingar um "segmentation" kransæða út frá CCTA myndum en ekki upplýsingar um CAD. Þetta er gagnasafn með 1000 3D CCTA myndum af hjartanu og er 90GB að stærð. Safnið er ætlað til segmentation vinnu og því gefur að skilja að úrtak af þessum gögnum getur verið flóknara en það tvíundargreinda sem við höfum núna, þó það sé ónýtilegt, þar eð engin merking er gerð á kransæðastíflu. Þar sem þessi gagnasett henta ekki okkar verkefni ætlum við að nota upprunalegu gögnin okkar og einbeita okkur að tvíundargreiningu CAD og þar af leiðandi athugun á skilvirkni MPV framsetningu gagnanna.

Einhverjar rannsóknir hafa nýtt sér sama gagnasett og við og sýnt ágætar niðurstöður. Til dæmis notuðu höfundar gagnasettsins nokkrar mismunandi aðferðir til að þjálfa módel á gögnunum og fengu mest 0.93 AUC. Sú rannsókn sýndi að MPV aðferðin skilar almennt betri niðurstöðum heldur en t.d. "aggregate projection view" (APV) en þá var öllum sjónarhornum af hverri að blandað saman í eina mynd [19]. Seinni rannsókn eftir sömu höfunda notaði aftur MPV framsetningu og að hluta til sömu gögn og fyrri greinin, en þá fékkst m.a. 0.96 AUC og 91.9% greiningarnákvæmni með 85.6% recall, 92.68% specificity, og 58% precision [20].

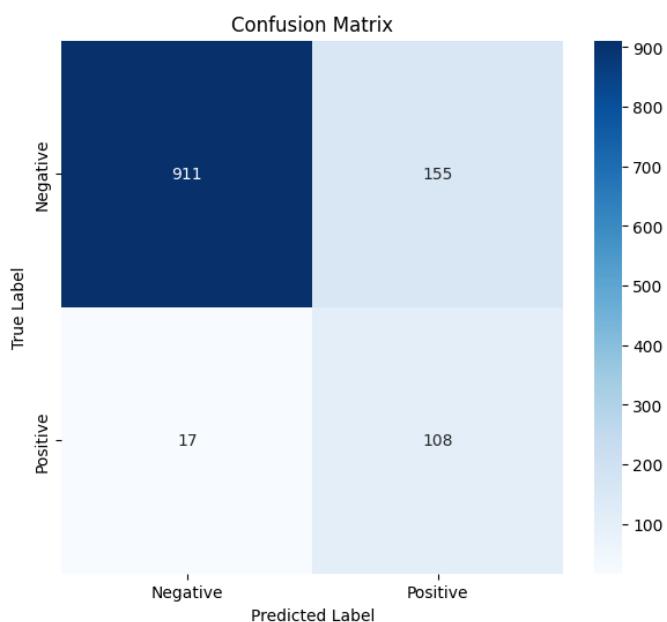
Í okkar verkefni vildum við skoða mismunandi aðferðir til að þjálfa módel á gagnasettinu okkar, bera saman niðurstöðurnar og meta hvaða módel hefði bestu möguleika á frekari þróun þannig að væri hægt að nýta í raunverulega greiningu í heilbrigðisstofnunum. Til að byrja með var einfalt CNN byggt algjörlega frá grunni og það notað sem "baseline" módel fyrir aðrar útfærslur. Við unnum í Google Colab og gátum því nýtt okkur öflug GPU, t.d. A100, til að þjálfa módelin. Allan kóða sem við notuðum má finna á GitHub [21] og módelin sem við þjálfuðum eru á Hugging Face [22].

3.1 CNN MÓDEL SEM BASELINE

Fyrsta CNN módelið var byggt frá grunni. Það hafði fjögur földunararlög sem nýttu ReLu virkniföll, fjögur MaxPooling lög, og í lokin myndaði það fulltengt lag með 512 taugungum og skilaði niðurstöðum sem líkindi á bilinu 0-1 með sigmoid fallinu. Kostnaraðfall tvíundagreiningar var notað og Adam bestunaraðferð með upphafs 'learning rate' 0.001. Til að gera módelin skilvirkari þá 'augmentation' við þjálfunargögnin okkar t.d. með því að snúa þeim, þysja inn/út, færa upp/niður o.fl. Niðurstöður þessa módelið eru eftirfarandi:



Mynd 4: Keyrsla baseline módelins

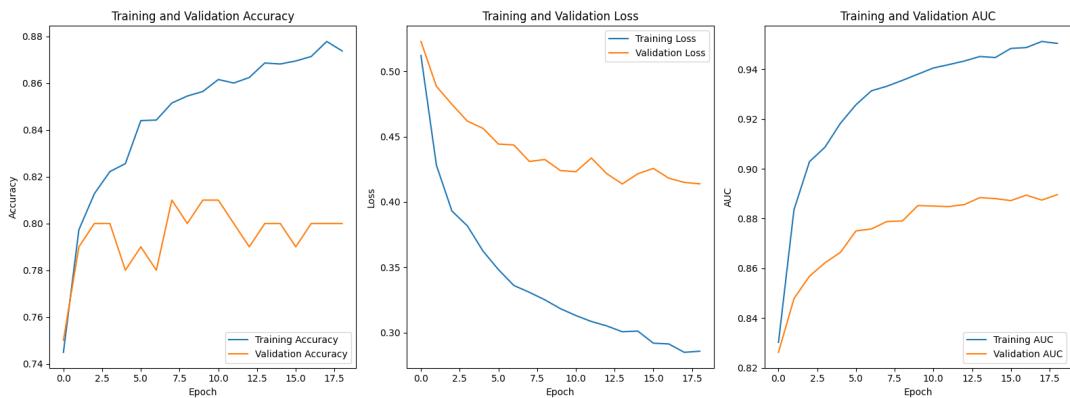


Mynd 5: Confusionfylki

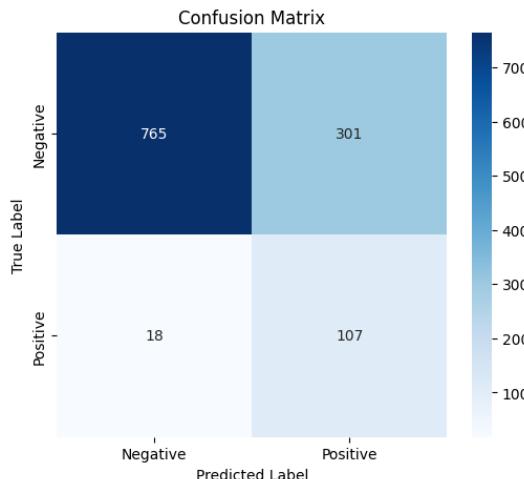
Prófunarnákvæmni var u.p.b. 86% og ROC-AUC 0.79, PR-AUC 0.56, Recall 86%, Precision 41%, F1 0.56. Þetta eru ekki slæmar niðurstöður fyrir fyrstu tilraun en þetta má bæta að miklu leyti, til dæmis með því að auka prófunarnákvæmni í a.m.k. 90% og auka precision þar sem hér er ofgreining.

3.2 INCEPTIONV3

Næst var könnuð frammistaða forþjálfaða módelins InceptionV3. Það módel er 48 lög á dýpt og við notuðum vigtir sem módelið lærði af ImageNet gagnasettinu. Þessi útgáfa módelins er mjög öflug í greiningu einkenna á myndum og er gjarnan notað í læknisfræðilegri myndgreiningu. InceptionV3 var notaður sem grunnur en lögum var bætt við þannig að niðurstöður pössuðu við verkefni okkar, t.d. var sigmoid fall notað á úttak eins og áður.



Mynd 6: Keyrsla InceptionV3 módelins



Mynd 7: Confusion fylki fyrir InceptionV3 módelið

Test Accuracy: 0.7305

Recall: 86%

PR-AUC: 0.6178

Precision: 26%

ROC-AUC: 0.8799

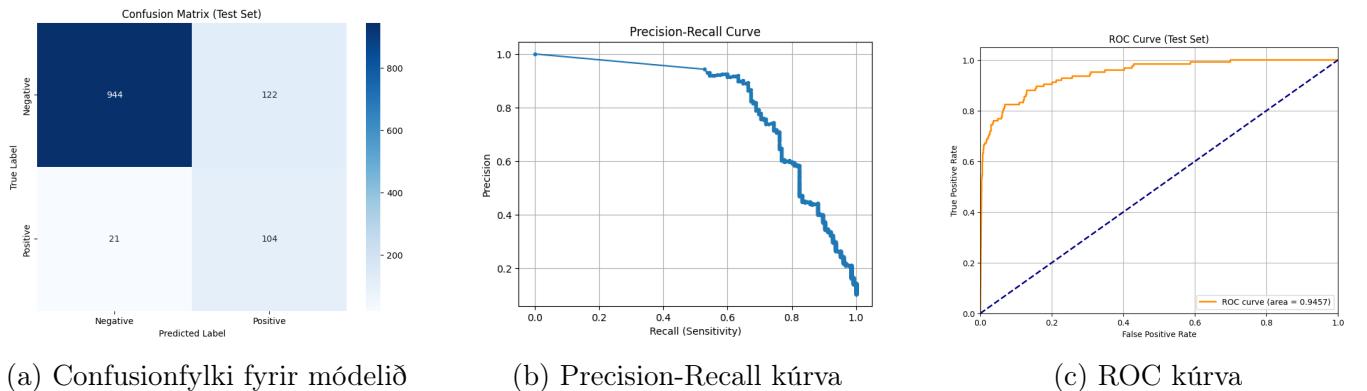
F1: 0.40

Aðeins var þjálfað á meðan 'validation loss' fór minnkandi. Petta er niðurstaða okkar eftir nokkrar tilraunir við að 'tweaka' módelið, t.d. með því að minnka ofþjálfun með 'dropout'. Eins og sést á mynd 6 var samt sem áður mikil ofþjálfun þar sem 'validation' nákvæmni var stöðugt lægri en þjálfunarnákvæmnin. Prófunarnákvæmni var 73% sem er mun lægra en við ætluðumst til.

3.3 KYNSLÓÐASKIPT CNN

Hér datt okkur í hug að útfæra nokkurslags kynslóðaskipt algrím til að þjálfa módelið okkar, svoleiðis að í þessu tilviki voru 5 epoch gerð á 5 ólíkum einstaklingum kynslóðar, sá besti valinn (elítismi) og 10 kynslóðir þjálfaðar. Hér yfirsást okkur hins vegar að láta módelið dýpka með hverri kynslóð, svoleiðis að þetta módel gerir það ekki. Það olli því hins vegar að þessi módel voru fljótari að þjálfast en dýpkandi módelið, enda grynnri. Seinna endurbættum við algrímið, en greint er frá því seinna.

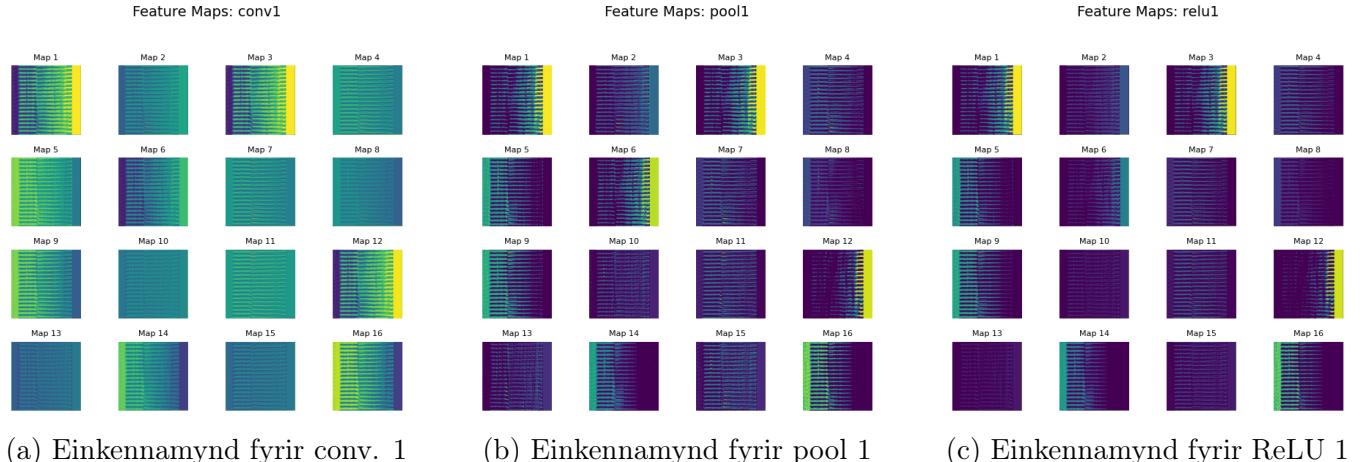
Í þessari útfærslu notuðum við kynslóðaalgrím á CNN sem hafði svipaða uppyggingu og samanburðaraðferðin okkar, nema hér var bætt við dropout og batch-normun til að koma í veg fyrir ofþjálfun. Niðurstöður eru eftirfarandi, en nákvæmari tölur má sjá seinna í töflu 1, ásamt því hvernig metlar breytast með þróskuldi sem hámarkar F1-skor:



(a) Confusionfylki fyrir módelið

(b) Precision-Recall kúrva

(c) ROC kúrva



(a) Einkennamynd fyrir conv. 1

(b) Einkennamynd fyrir pool 1

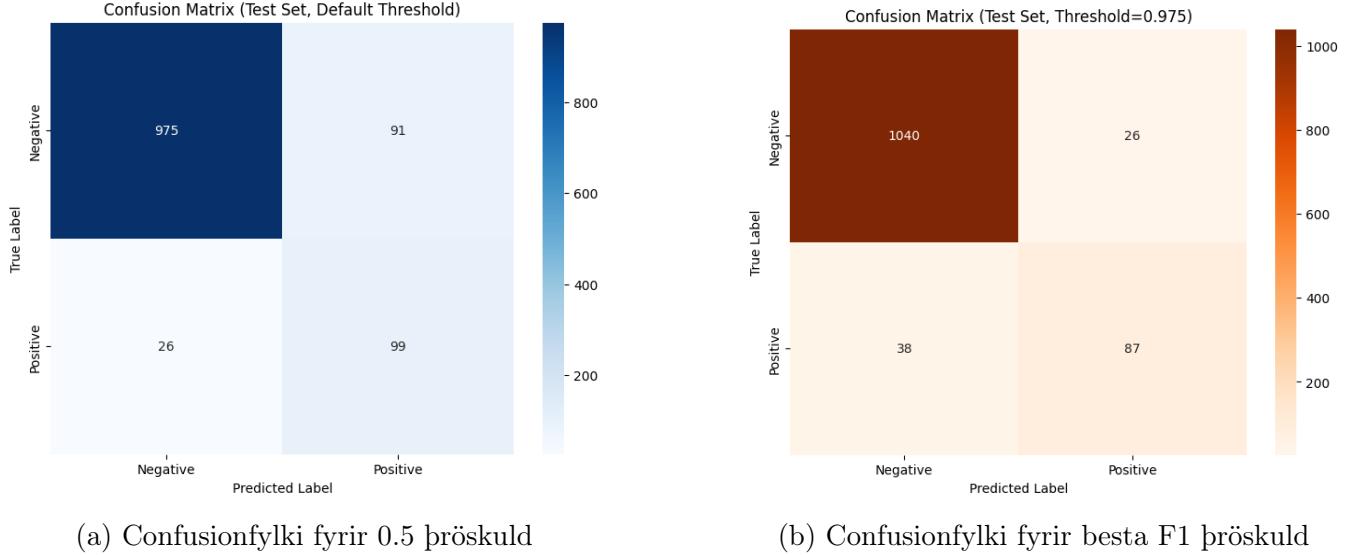
(c) Einkennamynd fyrir ReLU 1

Í þessari útfærslu ákváðum við að gera error analysis og kanna hvernig myndirnar líta út sem eru flokkaðar vitlaust. Ekki sást neitt augljóst sem þessar myndir áttu sameiginlegt þannig að við ákváðum að athuga einkennamyndir til að sjá hverju módelið er að leita að. Nokkrar einkennamyndir má sjá hér að ofan, en aftur var ekkert augljóst sem var hægt að greina úr þeim myndum. Hins

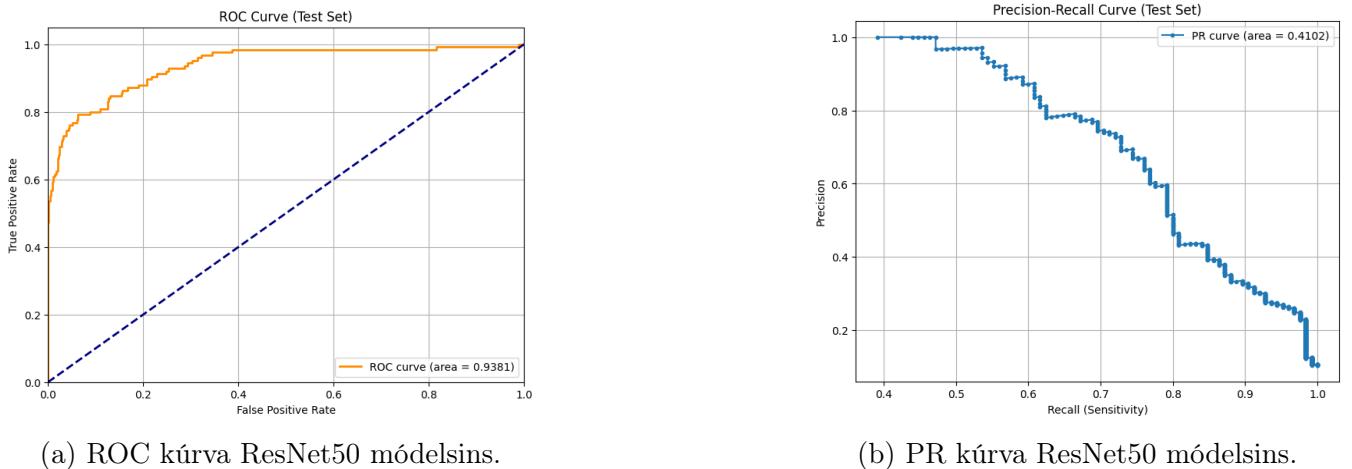
vegar vildum við í næstu útfærslum athuga hvort frammistaða myndi batna ef við klipptum út auðu svæðin sem mynduðust oft lengst til hægri hjá myndum af styttri æðum.

3.4 RESNET50

Hér notuðum við módelið ResNet50 (50 laga djúpt) forþjálfað á ImageNet gagnasettinu með generati-onal aðferðinni okkar. Aftur athuguðum við upprunalega þröskuldinn 0.5 og besta þröskuldinn upp á F1-skor, en það var hér 0.975.



Mynd 10: Confusionfylkin fyrir kynslóðaskipta ResNet50 módelið.



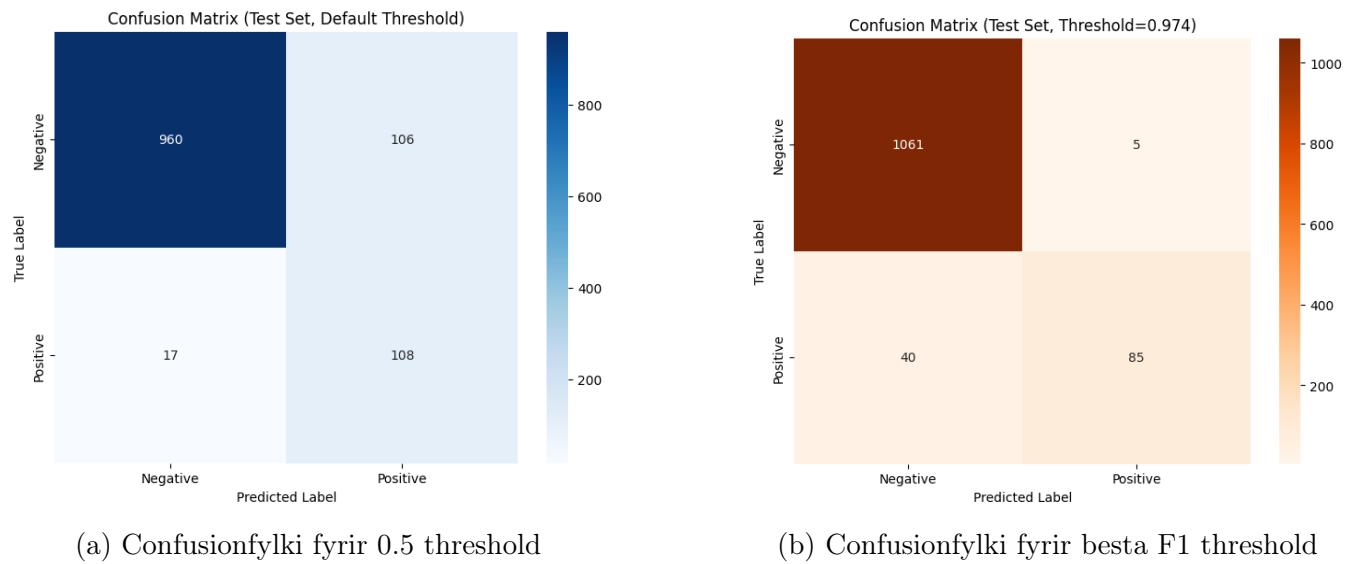
Mynd 11: ROC og PR kúrvur Resnet50 módelins.

Pess ber að geta að PR kúrvan hér að ofan er ekki fullkomin þar sem vantar hluta hennar þar sem Recall<0.4, og því flatarmálið lægra en það ætti að vera. Okkur tókst því miður ekki að leiðréttu þessa villu en sjáum að kúrvan er talsvert fyrir ofan skálínuna yfir miðju grafsins, þannig að flatarmálið er a.m.k. stærra en 0.5.

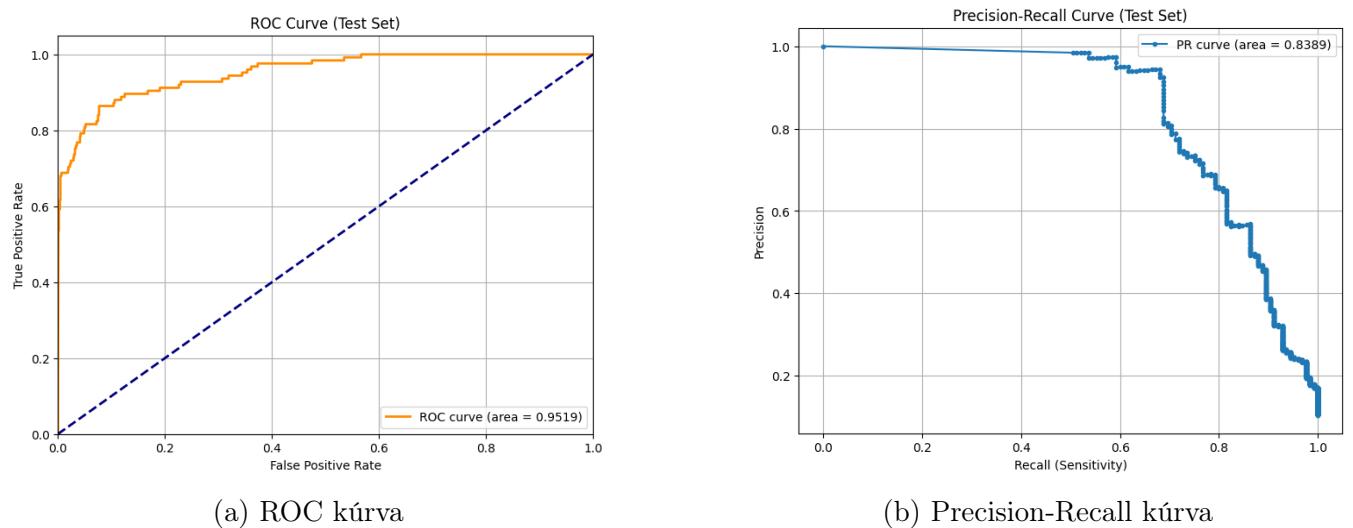
3.5 KYNSLÓÐASKIPT DÝPKANDI CNN

Að lokum prófuðum við að þjálfa módel sem dýpkar í þjálfuninni, eins og upprunalega ætlunin var. Hugsunin með þessu er að besta módelið sé valið eftir epoch þjálfunina, en að svo sé bætt aftan á það lögum og þau þjálfuð út frá þessu besta módeli. Við prófuðum þetta með lagfæringum á kóðunum sem við höfum fyrr notað, þannig að checkpoint systemið var með. Við hófum þjálfunina á því að í hverri kynslóð bættust við 3 lög, en minnkuðum niður í 2 aukaleg lög eftir fyrstu kynslóðina. Þess vegna eru 21 aukalög ($3 + 2 \cdot 9$) í módelinu.

Þetta módel var fljótara að þjálfast samanborið við fyrri módel og gaf okkur auk þess góðar niðurstöður sem eru hér að neðan:

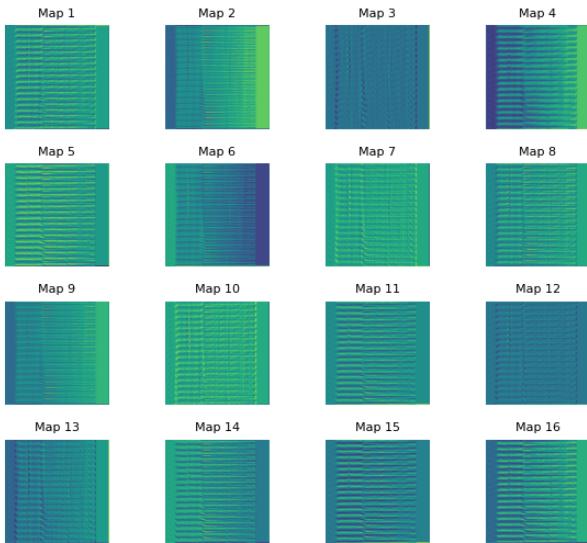


Mynd 12: Confusionfylkin fyrir kynslóðaskipta, dýpkandi CNN módelið.



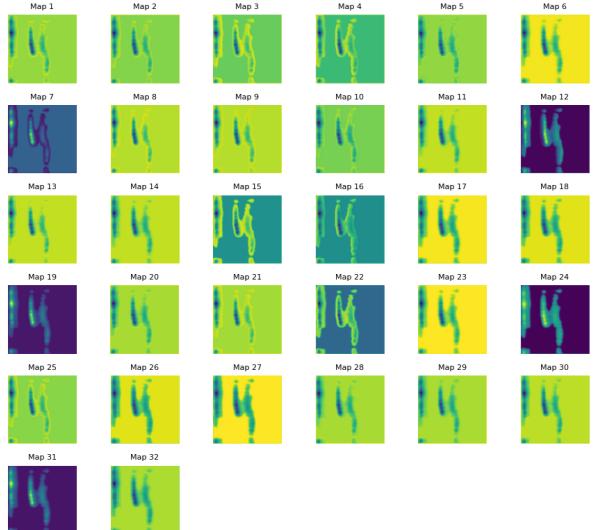
Mynd 13: Grafísk framsetning á ROC og Precision-recall

Feature Maps: conv1



(a) Einkennamynd fyrir conv. 1

Feature Maps: extra_conv_layers[40] (Last Conv)



(b) Einkennamynd síðasta conv. lags

Mynd 14: Nokkrar einkennamyndir módel eins

Sjá má á mynd 14b) að mögulega hefur þetta líkan greint þau svæði myndarinnar sem mest gefa til kynna um kransæðastíflu, þ.e. svokölluð "high-importance areas". Hins vegar þykir nokkuð óeðlilegt að allar einkennamyndir lagsins merkja nákvæmlega sama svæði myndanna. Við erum því óvissir með sannindi þeirrar niðurstöðu sem myndin bendir til. Ef þetta reynist hinsvegar rétt mætti minnka datasettið því t.d. sést að sjónarhornin sem eru tæplega neðst vega lítið til upplýsingar. Svona breytingar á gagnasettinu þyrfti að gera í samráði við heilbrigðismenntaða sérfræðinga.

3.6 SAMANTEKT

Hér að neðan má sjá yfirlit yfir þau módel sem við þróuðum í rannsóknarvinnu okkar ásamt tveimur rannsóknum sem áður hefur verið greint frá.

Módel	Acc	Rec	Pre	F1	ROC-AUC	PR-AUC
Baseline	86%	86%	41%	0.56	0.79	0.56
InceptionV3	73%	86%	26%	0.40	0.88	0.62
CNN Gen (0.5)	88%	83%	46%	0.59	0.95	0.81
CNN Gen (0.994)	96%	66%	89%	0.76	0.95	0.81
ResNet50 Gen (0.5)	90%	79%	52%	0.63	0.94	>0.5
ResNet50 Gen (0.975)	95%	70%	77%	0.73	0.94	>0.5
Dýpkandi Gen (0.5)	90%	86%	50%	0.64	0.95	0.84
Dýpkandi Gen (0.974)	96%	68%	94%	0.79	0.95	0.84
Rannsókn [20]	92%	86%	58%	0.69	0.96	N/A
Rannsókn [11]	91%	78%	65%	0.71	0.93	N/A

Tafla 1: Samanburður módelna (Acc = Accuracy, Rec = Recall, Pre = Precision)

Ekkert módel sem við þróuðum er augljóslega betra en hin. Hins vegar ná flestar útfærslur mjög háu AUC fyrir bæði ROC og PR kúrvurnar. Þar sem mikill halli er á neikvæð sýni í prófunargögnum okkar er viðmiðunargildi fyrir PR-AUC í kringum 0.105, og fáum við stöðugt mjög há gildi miðað við það. Greiningarnákvæmni er hærri hjá módelunum sem nýttu generational algrím okkar, en sá metill getur verið villandi vegna hallans í prófunargögnunum. Þegar mismunandi þröskuldar voru skoðaðir sýndi hærri þröskuldurinn stöðugt lægra recall og hærra precision. Til að koma í veg fyrir vitlausar jákvæðar greiningar CAD þá væri betra að nota 0.5 þröskuldinn, en þá verður aftur á móti meiri ofgreining á þeim sem eru heilbrigðir.

Miðað við aðrar rannsóknir sem hefur verið greint frá sýna módelin okkar yfirleitt sambærileg eða jafnvel betri ROC-AUC og náum við góðri greiningarnákvæmni, eða allt að 96%. Hins vegar er F1-skor oft frekar lágt og erfitt að ná góðu jafnvægi milli precision og recall.

4 RANNSÓKNAÁÆTLUN OG AÐFERDAFRÆÐI

Við ætlum að nota CNN (Convolutional Neural Network) til að greina kransæðamyndir í jákvæða eða neikvæða niðurstöðu, svoleiðis að úrlausn mynda einfalist inni á spítala, og hraði við þessa vinnu aukist. Við ætlum einnig að halda áfram þróuninni á kynslóðaskipta algríminu, og sjá hvort ekki megi tvinna þá þjálfun saman við forþjálfað módel, svo við blöndum saman kostum forþjálfaðs módels og dýpkandi kynslóðaalgrímsins.

Við viljum gera þá tilraun þegar við verðum komnir með gott og áreiðanlegt módel að nota gagnasafnið "ImageCAS" svoleiðis að við myndum vinna það yfir á læsilegt form fyrir líkanið okkar, láta líkanið flokka gagnasafnið, og þjálfa svo nýtt líkan á vélflokkad safninu. Við þetta er hætta á því að bias og gallar þess módels sem flokkar erfist yfir í nýja gagnasafnið og að öll módel sem þjálfuð verði á því reynist vera léleg. Við myndum hins vegar nota upprunalega gagnasafnið í heild sinni sem validation- og testing gögn, og þessi mannflokkuðu gögn myndu vonandi ljóstra upp um hvort módelið hafi raunverulega þjálfast rétt eður ei. Petta myndi að nokkru leyti leysa gagnasafnsvandann sem við glímum við í þessu verkefni, þar til betri gagnasöfn mætti finna.

Við framvæmd þessa verkefnis var ljóst að hvorki var hægt að nota gagnasett okkar eða módel til að kalk-greina, enda er það frekar hugsað sem útvíkkun á afurð þessa verkefnis. Hér er einungis gengið út frá jákvæðri eða neikvæðri greiningu CAD en til þess að hægt væri að meta t.d. stíflunarhlutfall eða kalkgreiningu þyrfti að merkja gögnin á viðeigandi hátt og við teljum okkur ekki til þess hæfa að gera það. Af sömu ástæðu getum við ekki einangrað staðsetningu stíflunnar. Reynist hinsvegar feature mappið á dýpkandi gagnasettinu rétt mætti mögulega reyna að einangra staðsetninguna á einhvern hátt út frá því.

Markmiðið er að læra á gervigreinarvinnu og einnig að þróa grunnmódel sem er í framtíðinni hægt að breyta og bæta með betri gagnasettum og aukinni þekkingu höfunda, svoleiðis að á endanum sé vonandi hægt að greina frekari þætti, eins og t.d. staðsetningu og stífluhlutfall. Stóra markmið þessarar rannsóknar má þó segja að sé að gera módel sem er nægilega áreiðanlegt til að innleiða í heilbrigðiskerfið.

5 STJÓRNUN, SAMSTARF OG SAMVINNA

Hugmynd þessa verkefnis fengum við að miklu leyti frá starfandi hjartalækni á Sjúkrahúsínu á Akureyri sem sérhæfði sig nýlega í úrlestri CCTA mynda til greiningar á CAD. Hann hefur sjálfur fundið fyrir mikilli tímapressu og á lagi við úrlestur þessara mynda og fylgist vel með þróun og framförum gervigreindarlausna á þessu sviði. Að sögn hans eyðir hann reglulega um 30 mínútum í úrlestur CCTA mynda á hvern sjúkling, en það er talsvert meira álag en búist er við fyrir hans starfshlutfall í úrlestri. Það eru mjög fáir á landinu sem hafa sérhæft sig í úrlestri CCTA mynda og nauðsynlegt að þróa aðgengilegt, sjálfvirkt greiningartól til að auðvelda úrlestur þessara mynda—enda er CCTA að verða vinsælla með ári hverju og CAD greiningar að aukast.

Hluti kóðans sem var notaður í að þróa gervigreindarlausnirnar okkar fékkst frá gervigreindar-spjallmennum eins og t.d. ChatGPT, Gemini og Deepseek. Kóði frá þessum spjallmennum var sameinaður og borinn saman til að gefa sem bestar niðurstöður, og alltaf reynt að beita gagnrýnni hugsun þar sem við gerum okkur vel grein fyrir því að lausnir þeirra geta verið vafasamar, en þær er á sama tíma erfitt að villugreina. Spjallmenni gáfu einnig góðar hugmyndir um byrjunarpunkta og áfanga verkefnisins.

Við hjálpuðumst að við þjálfun líkananna og Róbert skrifaði suma kafla hér en Ellert aðra. Róbert fann gagnasettin og ræddi við hjartalækninn en við fundum nokkuð jafnt greinar. Við þjálfuðum báðir CNN módel, en Róbert gerði InceptionV3 og ResNet módelin, en Ellerti datt kynslóðahugmyndin í hug og þjálfarinn kynslóðamódelið sem dýpkaði.

Einnig fékkst góð aðstoð frá kennurum þessa námskeiðs, þar sem þeir gáfu góð ráð og uppbyggilega gagnrýni, og leiðbeiningu.

6 NIÐURSTAÐA

Við höfum þróað fjölda gervigreindarmódel sem geta greint CAD út frá CCTA myndum með háu AUC gildi og mikilli greiningarnákvæmni. Við útfærðum skilvirktn kynslóðaskipt algrím sem finnur jafnóðum bestu vigtir tauganets meðan á þjálfun stendur. Nota mætti þessi módel t.d. í þróun mixture-of-experts módel sem segir fyrst til um hvort einstaklingur sé með CAD, notar svo annað módel sem þjálfa mætti á flóknara gagnasetti (t.d. ImageCAS) sem staðsetur kvillann, og þriðja módel sem greinir úrtak þessa módel og gefur kalkskor.

Það eru ýmis atriði í verkefni okkar sem við viljum útvíkka og rannsaka frekar. Meðal annars myndum við vilja prófa frekari aðferðir sem ekki gafst tími til að skoða í þetta sinn til að sjá hvernig þær spjara sig. Við viljum prófa random forest sem er oft notað í læknisfræðilegri myndgreiningu, support vector machine, multilayer perceptron—fyrst við erum að tvíundargreina—, og logistic regression. Þetta gæfi gott yfirlit yfir hæfni uppbyggingu líkananna okkar miðað við arkitektúr annara módelna sem reynst hafa vel í þessu fagi. Eitt vandamál er þó óleyst: gagnasett í þessum málum er hrikalega erfitt að finna. Því væri áhugavert að prófa að tvinna saman ImageCAS, sem er CAD gagnasett í þeim tilgangi að ná að segmenta æðarnar, og módelin okkar. Á þann hátt væri mögulegt að láta okkar módel myndgreina settið, svo gætum við þjálfað módel á þessu AI-myndgreinda setti, og notað svo allt gagnasettið sem við nýttum í þetta verkefni sem validation og testing sett. Það er hættulegt við þetta að matsskekja gervigreindarinnar erfist yfir á gervigögnin, en með því að nota ground-truth gögnin okkar sem validation og þjálfunargögn gætum við séð hvort líkanið okkar sé nærrí lagi eða ekki. Við þetta myndum við hafa aukið magn þjálfunargagna umtalsvert. Ef svoleiðis reynist að einkennamyndin á dýpkandi módelinu (mynd 14b) segir rétt til um þau svæði myndarinnar sem máli skipta þá mætti einnig prófa að eyða úr myndunum þeim svæðum sem ekki skora hátt á mappinu, svoleiðis að stærð settsins minnnkar. Þetta byggist hinsvegar á því að featuremappið sé áreiðanlegt, en við erum frekar óvissir um það.

Tilgangur þessa verkefnis var að rannsaka skilvirkni gervigreindarmódelna í sjálfvirkri greiningu CAD frá CCTA myndum. Núverandi lausnir á þessu vandamáli eru taldar vera óaðgengilegar eða óareiðanlegar og er þörf á frekari rannsóknum til að geta þróað einfalda lausn sem væri hægt að innleiða á smærri stofnunum sem hafa ekki aðgang að sérhæfðum úrlesurum. Þessi sjálfvirkju greiningartól hafa þann möguleika að straumlínulaga verkferla lækna og auka skilvirkni og nákvæmni greiningar algengasta sjúkdóms heims.

Bráðabirgðarniðurstöður okkar sýna að hægt er að segja nákvæmlega til um hvort sjúklingur hafi kransæðastíflu, sem getur einfaldað læknum ótrúlega; vel flýtir fyrir að þurfa ekki að skoða myndir sem er allt í lagi með. Hins vegar er áberandi þörf á frekari úrvinnslu og þróun módelna okkar, og myndu framtíðarverk leggja áherslu á að minnka fjölda óheilbrigðra sem fá neikvæða greiningu.

HEIMILDIR

- [1] British Heart Foundation, “Global Heart Circulatory Diseases Factsheet,” 01 2025. [Online]. Available: <https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf>
- [2] NHS, “Coronary heart disease,” 2024. [Online]. Available: <https://www.nhs.uk/conditions/coronary-heart-disease/>
- [3] Applied Radiology, “Top Ten clinical indications for coronary CT angiography,” Apr. 2008. [Online]. Available: <https://appliedradiology.com/articles/top-ten-clinical-indications-for-coronary-ct-angiography>
- [4] B. Baefler, M. Götz, C. Antoniades, J. F. Heidenreich, T. Leiner, and M. Beer, “Artificial intelligence in coronary computed tomography angiography: Demands and solutions from a clinical perspective,” *Front. Cardiovasc. Med.*, 2023. [Online]. Available: <https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2023.1120361/full>
- [5] A. Schuhbaeck, Y. Otaki, S. Achenbach, C. Schneider, P. Slomka, D. S. Berman, and D. Dey, “Coronary calcium scoring from contrast coronary ct angiography using a semiautomated standardized method,” *Journal of Cardiovascular Computed Tomography*, vol. 9, no. 5, pp. 446–453, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1934592515002038>
- [6] P. N. Rudziński, M. Kruk, C. Kępka, U. J. Schoepf, K. Otani, T. J. Leonard, M. Dębski, Z. Dzielińska, J. Pręgowski, A. Witkowski, W. Ruzyłło, and M. Demkow, “Assessing the value of coronary artery computed tomography as the first-line anatomical test for stable patients with indications for invasive angiography due to suspected coronary artery disease. initial cost analysis in the cat-cad randomized trial,” *Journal of Cardiovascular Computed Tomography*, vol. 14, no. 1, pp. 75–79, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1934592519300917>
- [7] R. Bernardo, N. S. Nurmohamed, M. J. Bom, R. Jukema, R. W. de Winter, R. Sprengers, E. S. G. Stroes, J. K. Min, J. Earls, I. Danad, A. D. Choi, and P. Knaapen, “Diagnostic accuracy in coronary CT angiography analysis: artificial intelligence versus human assessment,” *Open Heart*, 2025. [Online]. Available: <https://openheart.bmj.com/content/12/1/e003115>
- [8] M. Robert J. Widme, M. Zachary P. Rosol, M. Subhash Banerjee, M. Yader Sandoval, and M. Jeffrey M. Schussler, “Cardiac Computed Tomography Angiography in the Evaluation of Coronary Artery Disease: An Interventional Perspective,” *Journal of the Society for Cardiovascular Angiography Interventions*, Mar. 2024. [Online]. Available: [https://www.jscai.org/article/S2772-9303\(24\)00008-5/fulltext](https://www.jscai.org/article/S2772-9303(24)00008-5/fulltext)

- [9] A. R. Ihdayhid, A. Sehly, A. He, J. Joyner, J. Flack, J. Konstantopoulos, D. E. Newby, M. C. Williams, B. S. Ko, B. J. Chow, and G. Dwivedi, “Coronary artery stenosis and high-risk plaque assessed with an unsupervised fully automated deep learning technique,” *JACC: Advances*, vol. 3, no. 9_Part_2, p. 100861, 2024. [Online]. Available: <https://www.jacc.org/doi/abs/10.1016/j.jacadv.2024.100861>
- [10] H. Lee, B. G. Kang, J. Jo, H. E. Park, S. Yoon, S.-Y. Choi, and M. J. Kim, “Deep learning-based prediction for significant coronary artery stenosis on coronary computed tomography angiography in asymptomatic populations,” *Frontiers in Cardiovascular Medicine*, vol. Volume 10 - 2023, 2023. [Online]. Available: <https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2023.1167468>
- [11] J. Dundas, J. Leipsic, S. Sellers, P. Blanke, P. Miranda, N. Ng, S. Mullen, D. Meier, M. Akodad, J. Sathananthan, C. Collet, B. Bruyne, O. Muller, and G. Tzimas, “Artificial Intelligence-based Coronary Stenosis Quantification at Coronary CT Angiography versus Quantitative Coronary Angiography,” *Radiology: Cardithoracic Imaging*, vol. 5, 12 2023. [Online]. Available: <https://pubs.rsna.org/doi/full/10.1148/ryct.230124>
- [12] K. R. Branch, “Evolution or revolution?” *JACC: Advances*, vol. 3, no. 9_Part_2, p. 100860, 2024. [Online]. Available: <https://www.jacc.org/doi/abs/10.1016/j.jacadv.2024.100860>
- [13] J. Liao, L. Huang, M. Qu, B. Chen, and G. Wang, “Artificial Intelligence in Coronary CT Angiography: Current Status and Future Prospects,” *Frontiers in Cardiovascular Medicine*, vol. 9, 2022. [Online]. Available: <https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2022.896366>
- [14] C. Chen, N. A. Mat Isa, and X. Liu, “A review of convolutional neural network based methods for medical image classification,” *Computers in Biology and Medicine*, vol. 185, p. 109507, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482524015920>
- [15] M. Demirer, M. B. Vikash Gupta, B. Erdal, L. Prevedello, and R. White, “Image dataset for a cnn algorithm development to detect coronary atherosclerosis in coronary ct angiography,” 11 2019. [Online]. Available: <https://data.mendeley.com/datasets/fk6rysf3h9/1>
- [16] P. P. Sengupta, D. Dey, R. H. Davies, N. Duchateau, and N. Yanamala, “Challenges for augmenting intelligence in cardiac imaging,” *Lancet Digit Health*, 2024. [Online]. Available: [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(24\)00142-0/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(24)00142-0/fulltext)
- [17] A. Jiménez-Partinen, M. A. Molina-Cabello, K. Thurnhofer-Hemsi, E. J. Palomo, J. Rodríguez-Capitán, A. I. Molina-Ramos, and M. Jiménez-Navarro, “CADICA: a new dataset for coronary artery disease detection by using invasive coronary angiography,” Feb. 2024. [Online]. Available: <https://data.mendeley.com/datasets/p9bpx9ctcv/2>

- [18] A. Zeng, C. Wu, G. Lin, W. Xie, J. Hong, M. Huang, J. Zhuang, S. Bi, D. Pan, N. Ullah, K. N. Khan, T. Wang, Y. Shi, X. Li, and X. Xu, “ImageCAS: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images,” *Computerized Medical Imaging and Graphics*, vol. 109, p. 102287, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895611123001052>
- [19] V. Gupta, M. Demirer, M. Bigelow, K. J. Little, S. Candemir, L. M. Prevedello, R. D. White, T. P. O’Donnell, M. Wels, and B. S. Erdal, “Performance of a Deep Neural Network Algorithm Based on a Small Medical Image Dataset: Incremental Impact of 3D-to-2D Reformation Combined with Novel Data Augmentation, Photometric Conversion, or Transfer Learning,” *Journal of Digital Imaging*, vol. 33, no. 2, pp. 431–438, Apr 2020. [Online]. Available: <https://doi.org/10.1007/s10278-019-00267-3>
- [20] R. White, B. Erdal, M. Demirer, V. Gupta, M. Bigelow, E. Dikici, S. Candemir, M. Galizia, J. Carpenter, T. O’Donnell, A. Halabi, and L. Prevedello, “Artificial Intelligence to Assist in Exclusion of Coronary Atherosclerosis during CCTA Evaluation of Chest-Pain in the Emergency Department: Preparing an Application for Real-World Use,” 08 2020. [Online]. Available: <https://link.springer.com/article/10.1007/s10278-021-00441-6>
- [21] Ellert Blær og Róbert Tumi, “Ellertblaer/gervigreind-lokaverkefni,” <https://github.com/Ellertblaer/Gervigreind-Lokaverkefni>, 2025, accessed: Apr. 17, 2025.
- [22] ——, “Ellertblaer/inngangur_ad_mynstursgreiningu,” https://huggingface.co/Ellertblaer/Inngangur_ad_mynstursgreiningu, 2025, accessed: Apr. 17, 2025.