Anvendelse af kunstig intelligens til estimering af tumorcellekernefraktion i vævssnit af malignt melanom

> Kandidatafhandling Klinisk Videnskab og Teknologi, Aalborg Universitet



Vejleder: Lasse Riis Østergaard og bi-vejleder: Patricia Switten Nielsen Den 01. juni 2021



Titel:

Anvendelse af kunstig intelligens til estimering af tumorcellekernefraktionen i vævssnit af malignt melanom.

Projekt:

4. semester Klinisk Videnskab og Teknologi

Projektperiode:

01. februar 2021 – 01. juni 2021

Projektgruppe: 21gr10507

Deltager: Jeanette Bæhr Georgsen

Vejleder: Lasse Riis Østergaard

Bi-vejleder: Patricia Switten Nielsen

Sidetal: 60 Antal bilag: 2 Afsluttet den: 30. maj 2021

Resume

Baggrund: I Danmark kan der tilbydes flere forskellige former for targeteret kræftbehandling. En af disse kan være BRAFinhibitorbehandling, målrettet BRAF-mutationen, som kan findes ved modermærkekræft. Før behandlingen kan tilbydes, skal mutationen være identificeret ved den enkelte patient, med en godkendt analyse. Både sensitiviteten, såvel som tolkningen af denne analyse, er afhængig af en nøjagtig vurderet tumorcellekernefraktion.

Tumorcellekernefraktionen vurderes i dag ved en subjektive vurdering, og en fejlvurdering af denne kan i yderste konsekvens betyde, at patienten ikke tilbydes en potentiel gavnlig behandling.

Formål: Udvikle og evaluere en applikation baseret på kunstig intelligens, som kan estimere tumorcellekernefraktionen i digitaliserede HE-farvede vævssnit af modermærkekræft.

Metode: Der inkluderes vævssnit af modermærkekræft, hvor de inkluderede vævssnit fordeles i 60% til træning, 10% til validering og 30% til test. Vævssnit til træningen annoteres med en *handcrafted* algortime, og U-NET trænes til segmentering og klassificering af tumorcellekerner og normale cellekerner, for efterfølgende at kunne beregne en nøjagtig tumorcellekernefraktion.

Først evalueres netværkets evne til at segmentere og klassificere korrekt ved en analyse af de labels netværket tildeler og efterfølgende en beregning af sensitiviteten, specificiteten og nøjagtigheden. Efterfølgende visualiseres overensstemmelsen mellem applikationens tumorcellekernefraktion og *gold standard* ved Bland Altman plots og sammenlignes efterfølgende med overensstemmelse mellem *eye balling* og gold standard.

Resultater: Netværkets evne til korrekt segmentering og klassificering afspejles i en sensitivitet på 45% og en specificitet på 66,4%, hvilket giver en nøjagtighed i segmentering og klassificering på 52,8%.

Overensstemmelsen mellem applikationen og gold standard viser, at applikationen konsekvent underestimerer tumorcellekernefraktionen. Overensstemmelsen mellem eye balling og gold standard viser ligeledes, at denne heller ikke er nøjagtig.

Konklusion: Applikationen, som er udviklet med kunstig intelligens kan, ikke for nuværende, estimere en mere nøjagtig tumorcellekernefraktion end den subjektive vurdering, som anvendes i diagnostikken.

Rapportens indhold er frit tilgængeligt, men offentliggørelse (med kildeangivelse) må kun ske efter aftale med forfatteren.



STUDENT REPORT

Title:

Quantification of tumor burden by artificial intelligence in tissue sections of melanoma.

Project:

4th semester Clinical Science and Technology

Project period: February 1st 2021 – June 1st 2021

Project group: 21gr10507

Participant: Jeanette Bæhr Georgsen

Supervisor: Lasse Riis Østergaard

Co-supervisor: Patricia Switten Nielsen

Number of pages: 60 Appendices: 2 Date of completion: May 30th 2021

Abstract

Background: Several different forms of targeted oncology treatments are available in Denmark. An example is the BRAF-inhibitor therapy targeting the BRAF mutation in melanoma. The therapy depends on a positive BRAF mutation identified by an approved assay for each patient. Both the sensitivity as well as the interpretation of this assay depends on an accurately assessed tumor cell nucleus fraction. The tumor cell nucleus fraction is currently assessed subjectively and a misjudgment of this can, as an extreme consequence, lead to the patient not being offered a potential beneficial treatment.

Aim: Develop and evaluate an application based on artificial intelligens that is able to calculate the tumor cell nucleus fraction in whole slide images of HE-stained melanomas.

Method: Tissue sections of melanom are included divided into 60% for training, 10% for validation and 30% for testing. Tissue sections for the training are annotated with a handcrafted algorithm. U-NET is trained to perform segmentation and classification of tumor cell nuclei and normal cell nuclei in order to calculate an exact tumor cell nucleus fraction.

First, the algorithm's segmentation and classification are evaluated by an analysis of the labels assigned by the algorithm. The results will be used to calculate sensitivity, specificity and accuracy. Next, the agreement between the algorithm's tumor cell nucleus fraction and the gold standard is evaluated by a Bland Altman plot. Finally, the algorithm's segmentation and classification will be compared with the agreement between a subjective estimate and the gold standard.

Results: The ability of the algorithm to conduct proper segmentation and classification shows a sensitivity of 45% and a specificity of 66.4%, giving an accuracy in segmentation and classification of 52.8%.

The agreement between the application and the gold standard shows that the application consequently underestimates the tumor cell nucleus fraction and, furthermore, shows the agreement between subjective estimation and the gold standard is also not accurate.

Conclusion: The application developed with artificial intelligence can currently not estimate a more accurate tumor fraction than the subjective assessment used in the diagnosis.

The content of this report is free accessible, but publication (with references) may only be made in agreement with the author.

Dette projekt er udarbejdet af gruppe 21gr10507 som kandidatspeciale på Klinisk Videnskab og Teknologi på Aalborg Universitet. Projektgruppen består af én bioanalytiker med flere års klinisk erhvervserfaring i det pato-anatomiske speciale.

Formålet med dette projekt er at udvikle en applikation til bestemmelse af tumorcellekernefraktionen, i vævssnit af malignt melanom. Applikationen udvikles ved træning af U-NET til korrekt segmentering og klassificering af celler. I forbindelse med projektet udvikles en metode til annotering af forskellige cellekerner til træning af U-NET. Applikationen skal i fremtiden implementeres som et vigtigt støtteredskab til speciallæger i patologi, som varetager diagnostikken af maligne melanomer.

I projektet er der udført en struktureret litteratursøgning, med formålet at afdække, hvorvidt emnet tidligere er belyst i samme omfang, som er hensigten med dette projekt. Identificeret litteratur er anvendt i diskussionen. Søgeprotokollen er vedlagt som Bilag A.

Projektet henvender sig til personale i det pato-anatomiske speciale, som har interesse i udvikling applikationer ved anvendelse af kunstig intelligens, som et redskab til beslutningsstøtte i diagnostikken. Dette projekt kan bidrage med en videnskabelig dokumentation af metoder til udviklingen af en applikation.

Illustrationer i projektet er originale og er udarbejdet af projektgruppen, dog er enkelte illustrationer fundet online. Disse er markeret med *creative commons licens*. Ved de anvendte illustrationer, som er markeret med *creative commons licens*, er der refereret til kilden igennem referencesystemet.

Stor tak til vejleder Lasse Riis Østergaard, for kompetent og konstruktivt vejledning og sparring i udarbejdelsen af dette projekt. Ligeledes en stor tak til bi-vejleder Patricia Switten Nielsen for kompetent bidrag til udvikling af applikation samt for konstruktiv vejledning og sparring igennem projektet. Derudover skal der rettes en tak til professor ved Patologi, AUH, Torben Steiniche for præsentation af problemstillingen i dette projekt samt for bidraget med relevant viden, tid og arbejdskraft. Yderligere skal der rettes en tak til bioanalytiker Julie Duelund Sørensen fra Molekylærlaboratoriet, Patologi AUH, for hendes bidrag af viden og hjælp indsamling af materialet. Endeligt en tak til Daniel Ramsing Lund for udvikling af analysealgoritme til vurdering af applikationens segmentering og klassificering.

Unate

Jeanette Bæhr Georgsen

Dette projekt er inddelt i kapitler, som yderligere er inddelt i underafsnit. Kapitler og afsnit er opstillet i naturlig orden, hvorfor projektet bør læses kronologisk for at sikre optimal forståelse. Hvert kapitel indledes med en kort beskrivelse i form af en metatekst.

Der er i dette projekt anvendt Vancouver referencesystem til håndtering af eksterne kilder. Her angives referencer som tal i parentes. Angives kilden før punktum refereres der udelukkende til den pågældende sætning, hvor kildeangivelse efter punktum referer til forudgående sætninger eller afsnit.

Tal fra et til ti vil blive udskrevet, hvor tal over ti vil angives med tal for at øge læsevenligheden. Dog vil procenter, beløb samt år skrives med tal.

Der vil i projektet anvendes gængse forkortelser, såsom f.eks. og kr., såfremt det bidrager til læsevenligheden. Forkortelser for fagtermer vil først blive udskrevet og forkortelsen angives i parentes, hvorefter denne forkortelse anvendes efterfølgende i projektet. **BRAF-mutation**: BRAF-mutationen findes i flere forskellige varianter, hvor der i dette projekt henvises til V600E eller V600K varianterne, da det er disse mutationer som rutinemæssigt undersøges, og hvortil der findes målrettet behandling.

Malignt melanom: I dette projekt henvises der til kutane melanomer af typen "superficielt spredende malignt melanom", da det er den hyppigste variant og ligeledes den type, hvor BRAF mutationen oftest identificeres.

Tumorcellekernefraktion: tumorcellekerne fraktionen anvendes i dette projekt om begrebet tumorprocent. Denne beregnes ud fra andelen af tumorceller af det samlede antal celler, der er fundet.

Digital patologi: Begrebet digital patologi henviser i dette projekt til den automatiserede billedanalyse, som anvendes til analyse af histologiske vævssnit.

Applikation: Applikation anvendes i dette projekt som et begreb for den færdigudviklede algoritme med efterfølgende programmering, så denne kan anvendes til det egentlige formål.

Eye balling: I dette projekt anvendes begrebet *eye balling* om den subjektive vurdering en patolog foretager ved mikroskopi af et vævssnit.

Indholdsfortegnelse

1 Introduktion	1
2 Problemanalyse	2
2.1 Cancer epidemiologi og økonomi	2
2.2 Malignt melanom	2
2.2.1 Pato-fysiologi og ætiologi ved malignt melanon	3
2.2.2 Diagnostik af malignt melanom	4
2.2.3 Behandling af modermærkekræft	7
2.2.4 Patient perspektiv	9
2.3 Digital patologi	10
2.3.1 Kunstig intelligens	11
2.3.2 Digital patologi som beslutningsstøtte	12
3 Problemformulering	13
3.1 Problemafgræsning	13
3.2 Problemformulering	14
4 Metode	15
4.1 Løsningsstrategi	15
4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk	15 15
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 	15 15 17
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 4.2 Anvendt materiale 	15 15 17 19
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 	15 15 17 19 19
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering 	15 17 19 19 21
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering 4.3 Udvikling af applikation 	15 15 17 19 19 21 24
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET. 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering. 4.3 Udvikling af applikation 4.3.1 Annotering af cellekerner 	15 15 17 19 21 24 24
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering 4.3 Udvikling af applikation 4.3.1 Annotering af cellekerner 4.4 Evaluering af applikation 	15 17 19 21 24 24 28
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET. 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering. 4.3 Udvikling af applikation 4.3.1 Annotering af cellekerner 4.4 Evaluering af applikation 4.4.1 Evaluering af segmentering og klassificering. 	15 17 19 19 21 24 24 28 28
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering 4.3 Udvikling af applikation 4.3.1 Annotering af cellekerner 4.4 Evaluering af applikation 4.4.1 Evaluering af segmentering og klassificering 4.4.2 Evaluering af tumorcellekernefraktionen 	15 17 19 21 24 24 28 28 32
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET. 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering. 4.3 Udvikling af applikation 4.3.1 Annotering af cellekerner 4.4 Evaluering af applikation 4.4.1 Evaluering af segmentering og klassificering 4.4.2 Evaluering af tumorcellekernefraktionen 5 Resultater. 	15 17 19 21 24 28 28 28 32 34
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering 4.3 Udvikling af applikation 4.3.1 Annotering af cellekerner 4.4 Evaluering af applikation 4.4.1 Evaluering af segmentering og klassificering 4.4.2 Evaluering af tumorcellekernefraktionen 5 Resultater 5.1 Evaluering af segmentering og klassificering 	15 17 19 21 24 24 28 28 32 32 34
 4.1 Løsningsstrategi 4.1.1 Konvolutionelle neuralt netværk 4.1.2 U-NET. 4.2 Anvendt materiale 4.2.1 Etiske overvejelser 4.2.2 Dataindsamling og præparering. 4.3 Udvikling af applikation 4.3.1 Annotering af cellekerner 4.4 Evaluering af applikation 4.4.1 Evaluering af segmentering og klassificering. 4.4.2 Evaluering af tumorcellekernefraktionen 5 Resultater. 5.1 Evaluering af segmentering og klassificering 5.1.1 Evaluering af segmentering. 	15 17 19 21 24 28 28 32 32 34 34 35

5.2 Evaluering af tumorcellekernefraktionen	40
5.2.1 Applikationens nøjagtighed	40
5.2.2 Eye ballings nøjagtighed	45
6 Diskussion	46
6.1 Metode til annotering af træningsdata	46
6.2 Applikationens evne til segmentering og klassificering	48
6.3 Vurdering af tumorcellekernefraktionen	50
7 Konklusion	52
8 Perspektivering	53
Søgeprotokol - Bilag A	61
Dataopsamlingsskema- Bilag B	63

Introduktion 1

Kræft opstår på grund af mutationer i vores DNA. Disse mutationer kan påvises ved en række nye molekylærbiologiske metoder, f.eks. gensekventering. Påvisning af mutationerne får stigende betydning for diagnostikken af kræft, fordi der i dag tilbydes flere forskellige behandlinger, som er målrettet mod en del af disse mutationer (1). Et eksempel er behandlingen målrettet BRAF-mutationen ved modermærkekræft. I forbindelse med diagnostik af metastaserende modermærkekræft undersøges vævet for BRAF-mutationen; ca. 50% har denne mutation, og op imod 70% af de behandlede vil opleve betydelig gavn af behandlingen (2).

Der er i Danmark godkendt to forskellige lægemidler med hæmmere mod BRAF-mutationen (2). Behandling med disse BRAF-hæmmere initieres kun ved en positiv mutationsanalyse og gives ikke til BRAF udtrykt uden mutation, BRAF-vildtype cancer, da dette kan føre til en negativ behandlingseffekt. Desuden spildes unødig tid for patienten, og medfører unødige udgifter for samfundet, fordi der benyttes uvirksom medicin. (3,4) Det er tidligere diskuteret i litteraturen, hvorvidt behandlingseffekten med BRAF-hæmmere er afhængig af omfanget af inter- eller intratumor heterogenitet af BRAF-mutationen. Derfor er oplysningen om denne heterogenitet også betydningsfuld i forhold til det indledende valg af behandling (4).

Ved BRAF-mutationsanalyse udvælges et tumorrepræsentativt område (region of interest; ROI), som efterfølgende sekventeres ved *Next Generation Sequencing* (NGS). Analysens sensitivitet stiller dog krav til mængden af tumorceller i det analyserede væv. Ved mikroskopi udvælger og afmærker en patolog derfor ROI, og samtidig laves en subjektiv vurdering af fraktionen af tumorceller af alle cellerne i området. Studiet af Smits et al. (5) har vist, at den subjektive vurdering af tumorcellekernefraktionen er upræcis og unøjagtig, hvilket kan kompromittere sensitiviteten af BRAF-mutationsanalysen. Da svaret på BRAF-analysen har så stor betydning for valg af behandling, er det vigtigt, at den vurderede tumorprocent er så nøjagtigt som muligt, da det ellers kan betyde et forkert valg af behandling.

Tidligere forskning har vist, at kunstig intelligens kan anvendes til segmentering af forskellige celletyper i histologiske vævssnit samt til beregning af både tumor- og arealfraktion (6). Ovenstående har udledt en initierende undren, om hvorvidt kunstig intelligens kan anvendes til en mere præcis og nøjagtig vurdering af tumorcellekernefraktionen ved modermærkekræft forud for BRAF-analysen.

I dette kapitel præsenteres projektets problemanalyse. Indledningsvis præsenteres cancer epidemiologi og økonomi generelt. Efterfølgende belyses relevant viden om malignt melanom, herunder pato-fysiologi og ætiologi, nuværende udredning, diagnostik og behandling. Afslutningsvis vil begrebet kunstig intelligens til billedanalyse blive belyst.

2.1 Cancer epidemiologi og økonomi

På verdensplan blev der i 2018 diagnosticeret ca. 18,1 mio. nye tilfælde af cancer, og det estimeres at cancer også var årsagen til ca. 9,6 mio. dødsfald det år. Den beregnede risiko for at få cancer inden 75-årsalderen er 20%, og risikoen for at dø af canceren er 10%. (7) Antallet af nye tilfælde af cancer forventes at være på 29 mio. i 2040 både pga. den længere levetid samt den generelle stigning i populationen (8).

I Danmark blev der i 2018 diagnosticeret ca. 40.000 nye tilfælde af cancer, hvilket svarer til en risiko på 32,9% for udvikling af cancer inden 75-årsalderen (9). Dette er en betydelig øget risiko for udvikling af cancer i forhold til den samlede risiko på verdensplan. Dog er der i forbindelse med Globocan 2018 rapporten belyst, at der er langt bedre registrering i bl.a. de europæiske lande, end det er tilfældet i lande med begrænsede ressourcer. Ligeledes dækkede rapporten fra 2010 kun omkring 24% af verdensbefolkningen, (10), hvilket derfor kan være årsagen til den øgede incidens i Danmark.

Cancer har store økonomiske konsekvenser for samfundet. I 2010 er der på verdensplan estimeret en omkostning på 1,16 billioner amerikanske dollars, hvilket svarer til 6,3 billioner danske kroner. Denne omkostning inkluderer behandlingsomkostninger, tabt arbejdsfortjeneste for patienter og pårørende samt tabt produktivitet pga. pension eller død. (11) Sundhedsudgifterne i forbindelse med cancer er i Europa i 2014 estimeret til 83,2 mia. euro, hvoraf 1,2 mia. euro, svarende til 6,5 mia. danske kroner, er brugt på sundhedsudgifter i Danmark. Af sundhedsudgifterne i Danmark i 2014 er 1,49 mia. danske kroner, brugt på lægemidler mod cancer. (12)

2.2 Malignt melanom

I 2018 blev der på verdensplan diagnosticeret 287.000 nye tilfælde af malignt melanom, svarende til 1,6% af alle diagnosticerede cancere, og malignt melanom er derved den 19. hyppigste af alle cancertyper. Desuden skyldes ca. 60.000 af de cancerrelaterede dødsfald i 2018 malignt melanom, og sygdommen ligger dermed på 22. pladsen blandt alle cancerrelaterede dødsfald. På verdensplan er den estimerede risiko for at udvikle malignt

melanom inden 75-årsalderen på 0,35% og en risiko for at dø af canceren på 0,07%, hvor risikoen for befolkningen i Nordeuropa er estimeret til hhv. 1,82% og 0,21%. (13)

Der blev, i Danmark, diagnosticeret ca. 3000 nye tilfælde af malignt melanom i 2018, hvilket svarer til en hyppighed på 6,3% af de diagnosticerede cancere i Danmark (14). Ud af disse diagnosticeres ca. 350 årligt med metastaserende malignt melanom (15). I Danmark er den estimerede risiko for udvikling af malignt melanom 2,81% og med en risiko på 0,27% for at dø af canceren (9). Danmark er det land i verden med fjerde flest tilfælde af malignt melanom, når der tages højde for indbyggertal og aldersfordeling af befolkningen, og der bruges årligt 250 mio. kroner på kræft i huden i det danske sundhedsvæsen (15).

Den generelle 5-års overlevelse for patienter diagnosticeret med malignt melanom, uanset stadie, er 92%. Såfremt sygdommen diagnosticeres i et tidligt stadie, ses en 5-års overlevelse på 97 %, hvor hvis sygdommen først diagnosticeres i stadie IV er 5-års overlevelsen på 25%. (16) Fremskridt i den targeterede behandling af malignt melanom har dog vist en øget overlevelse blandt patienter diagnosticeret med avanceret cancer (17). Overlevelsen i alle stadier er afhængig af forskellige faktorer, såsom tumor tykkelse, histologisk og genetisk subtype samt involvering af lymfeknuder (16).

2.2.1 Pato-fysiologi og ætiologi ved malignt melanon

Huden beklæder hele kroppen, og er dermed også kroppens største organ. Huden har mange forskellige funktioner, og består derfor af flere forskellige vævstyper, der overordnet set er fordelt i 2 lag; epidermis og dermis. En af hudens funktioner er at beskytte mod ultraviolette (UV) stråler fra solen. UV-strålerne giver anledning til DNA skader i hudens celler, som efterfølgende kan medvirke til udviklingen af cancer i huden (18). Melanocytterne, som findes i det basale lag af epidermis, producerer og udskiller farvestoffet melanin, når huden udsættes for f.eks. sollys, hvor det efterfølgende optages af keratinocytterne. Melaninen lokaliserer sig over cellekernen, og absorberer en del af de skadelige UV-stråler (18,19). På figur 1 ses hvordan melaninen lægger sig over cellekernen for at beskytte mod de skadelige UV-stråler.



Figur 1 – 40x Mikroskopisk billede af et HE-farvet vævssnit af et malignt melanom, hvor melanin ses som en bræmme over cellekernen.

Til trods for at melanocytternes primære funktion er at beskytte hudens celler mod UV-stråler, viser det sig, at melanocytterne er i lige så stor risiko for DNA skader forårsaget af UV-strålerne

(18,20). Malignt melanom er den type af hudkræft, som udvikles i melanocytterne i epidermis, og kan opstå både i et eksisterende modermærke samt i normal hud. (18)

Den mest almindelige risikofaktor for udvikling af malignt melanom er eksponering af ultraviolette stråler, enten fra solen eller fra solarier. Ligeledes ses det, at omfanget af solskoldninger, specielt i børneårene, har en betydning for udvikling af malignt melanom. (18,20) Risikoen for udvikling af malignt melanom hænger derfor sammen med den enkeltes hudtype, fordi hudtypen er årsagen til risikoen for solskoldninger. Derfor ses det også at incidensen for malignt melanom er langt større i populationer med lys hud. (20)

2.2.2 Diagnostik af malignt melanom

For at sikre en ensartet høj kvalitet i udredningen af cancer, uanset hvor i Danmark patienterne bor, har Sundhedsstyrelsen udarbejdet standardiserede kræftpakker for flere forskellige former for cancer. Kræftpakkerne indeholder en forløbsbeskrivelse, som dækker fra mistanke om sygdom frem til afslutning af behandling. (21) For hvert enkelt kræftpakkeforløb er der udarbejdet vejledende forløbstider, som det skal tilstræbes at overholde for at sikre en både en hurtig udredning samt behandlingsstart for patienten. Disse forløbstider bliver kvartalsvis monitoreret på regions- og landsplan (22), ligeledes bruger de enkelte afdelinger disse forløbstider til egenmonitorering. De anbefalede forløbstider for malignt melanom er offentligt tilgængelige på Sundhedsstyrelsens hjemmeside (23). Ud over etablering af kræftpakkerne er der under Sundhedsstyrelsen også nedsat "Danske Multidiciplinære Cancer Grupper" (DMCG), som er en sammenslutning af de 24 forskellige cancergrupper, der er etableret i Danmark. DMCG'ernes formål er drift af de kliniske databaser samt udarbejdelse af kliniske retningslinjer til ensartet og evidensbaseret diagnostik og behandling ud fra dataindsamling, forskning og vidensdeling. (24) Gruppen for malignt melanom er "Dansk Melanom gruppe", som har udarbejdet forskellige kliniske retningslinjer indenfor udredning, diagnosticering, behandling og opfølgning for malignt melanom (25).

Ved mistænke om malignt melanom anbefales det at fjerne hele tumoren efter gældende retningslinjer for at sikre bedst mulige forudsætninger for en korrekt diagnose (26). Diagnosen stilles af en patolog, ud fra en mikroskopisk undersøgelse af tumoren. Her vurderer patologen melanomets undertype ud fra vækstmønster og ved hjælp af immunhistokemiske markører, som identificerer relevante proteiner, der f.eks. kan relateres til cellens udgangspunkt eller funktion. Vækstmønsteret i et primært malignt melanom kan f.eks. være irregulær vækst igennem basalmembranen, variation af kernestørrelser samt stor infiltration af lymfocytære celler. Figur 2 illustrerer vækstmønster ved henholdsvis et benignt nævi og et malignt melanom. Ved påvisning af et malignt melanom skal patologen også vurdere, hvorvidt tumoren er fjernet i sundt væv med frie resektionsrande, da dette har en betydning for, om der skal re-exciseres, og hvor meget der i så fald skal fjernes i forbindelse med re-excisionen. Yderligere skal forskellige prognostiske parametre vurderes, bl.a. skal Breslow's tykkelse angives, hvor melanomet måles i mm på det vertikalt dybeste sted og ligeledes skal tilstedeværelsen af ulceration vurderes. (27) Malignt melanom kan metastasere til mange forskellige organer i kroppen, og kan være svær, selv for en trænet patolog, at karakterisere i

den histopatologiske diagnostik, fordi tumorcellerne ofte er lavt differentierede, og kan efterligne cellerne i det metastatiske organ (28).



Figur 2 – 20x histologiske billeder af HE-farvede vævssnit. A: benignt nævi med afgrænset epidermis og dermis med ensartede cellekerner. B: malignt melanom med irregulær epidermis og varierende cellekerner både i epidermis og i dermis. C: nævi mod det subkutane fedtvæv, hvor cellekernerne stadig er ensartede og uden infiltration af lymfocytter. D: malignt melanom mod det subkutane fedtvæv, med stor infiltration af lymfocytter

Udbredelsen af sygdommen inddeles i stadier fra I-IV. Ved bestemmelse af stadiet angives udbredelsen af sygdommen ifølge Tumor, Node og Metastase klassifikation (TNM) efter internationale retningslinjer fra "Union for International Cancer Controls" (UICC) (29). TNM beskriver den anatomiske udbredelse af sygdommen, hvor T0-4 beskriver tumorens tykkelse og ulceration, N0-3 beskriver omfanget af lymfeknude involvering og M0-1 beskriver tilstedeværelsen af fjernmetastaser. Ved malignt melanom anvendes både en klinisk TNM og en patologisk TNM, hvor den kliniske TNM er baseret på kliniske fund på diagnosetidspunktet og angives cTNM, og den patologiske TNM, som angives pTNM, er baseret på den postoperative histologiske vurdering. Stadiet af sygdommen anvendes både som prognostisk redskab og som beslutningsgrundlag for behandling. (29)

Cancercellers egenskaber ved malignt melanom

I 2011 præsenteres *Hallmarks of cancer* som fundament for den biologiske forståelse af cancer. *Hallmark of cancer* definerer forskellige egenskaber, som cancerceller skal have, for at kunne overleve, proliferere og disseminere, se figur 3. Her italesættes det, at såfremt én af disse *hallmarks* har en betydning for den enkelte cancers progression, vil en targeteret inhibition af denne hæmme cancerens mulighed for at overleve og disseminere. Ligeledes vil en sådan targeteret behandling give begrænsede bivirkninger, fordi det netop kun vil have en virkning på de celler, som er påvirket. (30)



Figur 3 - Hallmarks of cancer illustrerer de forskellige egenskaber som kan findes i cancerceller (31).

Ved malignt melanom er der identificeret 2 forskellige egenskaber; PD-L1 og BRAF, som begge ligeledes er mål for targeteret behandling af sygdommen. Cancercellernes udtryk af PD-L1 styrker deres evne til at undgå kroppens eget immunforsvar. PD-L1 undersøges ved en immunhistokemisk farvning af PD-L1, hvor en tumor karakteriseres som positiv ved blot 1 % positive celler i det undersøgte vævssnit. (17) BRAF er en del af MAP-kinase signalvejen i cellen, hvor BRAF-vildtype genet bidrager til normal regulering af celledeling og celleoverlevelse. Figur 4 illustrerer MAP-kinase signalvejen med både vildtype funktion samt den muterede funktion af BRAF. BRAF-mutationen er en onkogen *driver* mutation, fordi en muteret BRAF medvirker til konstant aktivering af signalvejen, og medfører derfor øget proliferation og overlevelse og derved øget tumorvækst (17).



Figur 4 - Illustration af MAP-kinase signalvejen. Tv. ses signalvejens vildtype funktion og th. ses signalvejen med en muteret BRAF.

Der findes stadig modstridende evidens for, hvorvidt BRAF-mutationen er homogen eller heterogen. Nogle studier viser, at mutationen er homogen, fordi den er stabil over tid, og at den findes i alle tumorcellerne, også uanset primær tumor eller metastase (32,33). Hvor andre studier viser en heterogen mutation, som kan forekomme både som intratumor kloner, og kan variere fra primær tumor til metastase (4,34). Heterogeniteten kan være årsagen til behandlingsresistente BRAF-muterede maligne melanomer, fordi behandling kun vil have effekt på de tumorceller, som er muterede, og derved kun en begrænset del af tumoren (4,34). BRAF-muteret malignt melanom er ofte karakteriseret ved et mere aggressivt sygdomsforløb med større tendens til cerebral metastasering, og uden behandling ses en generel kortere overlevelse. Ligeledes rammer BRAF-muteret malignt melanom oftere yngre mennesker sammenlignet med BRAF-vildtype malignt melanom. (4)

BRAF-mutations status undersøges ved en molekylærpatologisk analyse, hvor der er flere forskellige analyser godkendt til denne (35). Molekylærpatologiske analysers sensitivitet er afhængig af andelen af tumorceller i relation til normale celler i det analyserede vævssnit, fordi de normale cellers DNA vil fortynde tumor DNA, hvilket kan medføre en risiko for falsk negative resultater (5,35,36). Molekylærpatologiske analyser kræver derfor en vis tumorcellekernefraktion for at sikre sensitiviteten af analysen. Ud over sikring af analysens sensitivitet er tumorcellekernefraktion ligeledes essentiel for tolkningen af analysens resultater (36). Vurderingen af tumorcellekernefraktionen foretages i dag af en patolog ved simpel *eye balling*, som er en subjektivt vurdering, og den inkluderer derved ikke en egentlig celletælling. Flere studier kritiserer *eye balling* for manglende præcision og nøjagtighed (5,36), hvorfor det er relevant at finde alternative metoder til denne vurdering.

På Patologi, Aarhus Universitetshospital (AUH) analyseres BRAF ved NGS, hvor der undersøges for alle mutationer i BRAF-genet, som er mulige mål for behandling. Som udgangspunkt foretages BRAF-analysen altid på den sidst udtagne biopsi af tumoren, hvad end det er fra den primære tumor eller fra en metastatisk proces. BRAF-analysen ved NGS kræver en tumorcellekernefraktion på mindst 20% for at sikre analysens sensitivitet. Svaret på analysen indeholder allelfrekvensen for muterede BRAF-alleler, hvor denne kan sammenholdes med tumorcellekernefraktionen, og derved kan bidrage til en vurdering af heterogeniteten af mutationen. BRAF-mutationer kan alternativt undersøges ved andre molekylærpatologiske analyser med forskellig sensitivitet, og yderligere kan den hyppigste mutation i BRAF påvises med en immunhistokemisk farvning (35).

2.2.3 Behandling af modermærkekræft

Behandlingen af malignt melanom er afhængig af hvilket stadie sygdommen er i. Ved tidlige stadier er den primære behandling operativ fjernelse af tumoren med eventuel adjuverende behandling, og ved senere stadier anvendes forskellige onkologiske behandlinger (37).

Kirurgisk behandling

Behandlingen af malignt melanom er en komplet excision af den primære tumor, dvs. efter diagnostikken udføres der en re-excision af området omkring det fjernede melanom. Formålet er at mindske risikoen for lokalt recidiv af tumoren (38). Breslow's tykkelse er afgørende for størrelsen af excisionen. Både ulceration og Breslow's tykkelse er afgørende for, hvorvidt der skal laves en *sentinel node* undersøgelse. Ved *sentinel node* identificeres de nærmeste drænerende lymfeknuder, så det er muligt kun at fjerne de lymfeknuder, hvor der er størst risiko for metastasering og derved undgå en lymfeknuderømning, som kan give

komplikationer og forringet livskvalitet for patienten. Ved ulceration eller en tykkelse på over 0,8 mm injiceres et radioaktivt stof ved det fjernede melanom, hvormed de nærmeste drænerende lymfeknuder kan lokaliseres, og efterfølgende operativt fjernes. Såfremt der ved en patologisk undersøgelse påvises metastaser i disse lymfeknuder, vil patienten tilbydes et intenst opfølgningsprogram. (39,40) Yderligere kan metastatiske processer identificeres ved PET-CT skanning af patienten, hvorefter det vurderes, hvorvidt processen kan fjernes som excisionsbiopsi, eller om der blot skal foretages en grovnålsbiopsi til diagnostik. Excisionsbiopsi er en komplet fjernelse af processen, og er muligt ved f.eks. subkutane processer og lymfeknuder, som ikke er identificeret som en drænerende lymfeknude. Grovnålsbioptering anvendes til diagnostik i de tilfælde, hvor den metastatiske proces identificeres i organer, hvor excision ikke umiddelbart er muligt, f.eks. i leveren. (23)

Medicinsk behandling

De fleste patienter diagnosticeres i et tidligt stadie, hvorfor de kan betragtes som raske efter operation. Dog diagnosticeres ca. 350 patienter årligt med metastaserende melanom, således inoperabelt melanom, og derfor tilbydes disse patienter udelukkende medicinsk behandling. De medicinske behandlingsmuligheder for malignt melanom omfatter immunterapi, targeteret behandling og kemoterapi. (17)

I de kliniske retningslinjer for behandling af metastaserende malignt melanom uden cerebrale metastaser, er der udarbejdet en behandlingsalgoritme som grundlag for valg af behandling, se figur 5. Ved hver af de fire undergrupper er der defineret flere linjer af behandling, som følges kronologisk, dog afhængigt af patientens co-morbiditet. (17)



Figur 5 -Algoritme til valg af behandling af metastaserende malignt melanom, hvor der ved hver af de sidste fire grupper er defineret forskellige linjer af behandling.

Immunterapi

Immunterapi er en mulighed for at styrke immunsystemets evne til at genkende og angribe cancerceller eller begrænse cancercellernes evne til at forsvare sig (41). Der findes forskellige typer af immunterapi til behandling af malignt melanom, både generel immunterapi og targeteret immunterapi. Som generel immunterapi kan der anvendes 2 forskellige antistoffer, som aktiverer immunsystemet på forskellige måder, hvor targeteret immunterapi er rettet mod en specifik receptor på cancercellernes overflade. (17)

Immunterapi anvendes som udgangspunkt til første linje behandling ved malign melanom, dog i forskellig kombination alt efter udfaldet af behandlingsalgoritmen (17).

Targeteret behandling

Ca. 50 % af tilfældene af malignt melanom har en mutation i BRAF-genet.

BRAF-mutationen behandles med en inhibitor, som hæmmer den konstante aktivering af MAP-kinase signalvejen, som mutationen har medført, og derved bremser væksten og spredning af canceren (37). Hvis BRAF-vildtype cancer behandles med en BRAF-inhibitor, kan der opstå øget aktivitet i MAP-kinase signalvejen, og derved kan der ses progression af sygdommen (4).

Targeteret behandling gives som anden linje behandling, såfremt target er identificeret (17). Der er dog i litteraturen tvetydige resultater om omfanget af inter- og intratumor heterogenitet, og hvorvidt dette har en betydning for valget af behandling (4).

Kemoterapi

Kemoterapi binder sig til de celler som prolifererer, hvor DNA'et skades, og således undergår cellen apoptose. Dog vil denne behandling også ramme normale celler, hvorfor anvendelsen af traditionel kemoterapi ofte er forbundet med bivirkninger for patienten. (2)

Behandling af malignt melanom med traditionel kemoterapi har ikke vist at øge overlevelsen, og kan derved kun betragtes som en lindrende behandling. Ligeledes tilrådes denne behandling kun, hvis patienten er i god almen tilstand. Derfor behandles malignt melanom kun med traditionel kemoterapi, når ingen af ovenstående behandlinger har effekt. (2,17)

2.2.4 Patient perspektiv

Forskning viser, at mutation i BRAF-genet, i forbindelse med malignt melanom, har en dårlig prognose og dermed en kortere overlevelse, hvis behandlingen ikke er rettet mod mutationen (42). Data fra et klinisk studie, hvor langtidsoverlevelsen er undersøgt, viser en forbedret sygdomsfri overlevelse ved behandling af malignt melanom med den nuværende kombinationsbehandling med BRAF-inhibitor. Ved denne behandling er 54% af de inkluderede patienter i live og uden tilbagefald af sygdom efter 4 år, hvor der i placebogruppen ses en 4-års sygdomsfri overlevelse på 38%. Ligeledes viser studiets dataanalyse, at risikoen for udvikling af fjernmetastaser eller død er reduceret med 47% ved behandling med BRAF-/MEK-inhibitor. (43) Der er ved alle patienterne, som er inkluderet i dette projekt, identificeret en mutation i BRAF, hvorfor der er taget højde for, at prognosen og overlevelsen generelt er dårligere ved BRAF-muteret malignt melanom sammenlignet med BRAF-vildtype malignt melanom. I forbindelse med det kliniske studie er der yderligere publiceret data på den patient relaterede oplevelse af sygdom under behandlingen. Gældende for både behandlingsgruppen samt placebogruppen blev der ikke identificeret ændringer i livskvaliteten for den enkelte. Dermed har valg af behandling ikke haft indflydelse på patientens livskvalitet, hverken under

behandling eller i den efterfølgende opfølgningsperiode, hvorfor behandlingen med fordel kan anvendes. (44) Korrekt identificering af BRAF-mutationen i malignt melanom er derfor vigtigt, for at kunne tilbyde den rette behandling og dermed forbedre den enkelte patients prognose.

2.3 Digital patologi

Begrebet digital patologi er vidt brugt, og dækker mange aspekter af den digitale billedanalyse. Indledningsvis blev begrebet anvendt om digitaliseringen af histologiske vævsnit til nu også at omfatte avanceret billedanalyse (45). Ved digitaliseringen af histologiske vævssnit kan vævet nu, med den rette software, mikroskoperes på en computer ligesom ved konventionel lysmikroskopi. Denne digitalisering har medført fordele i form af telepatologi, hvor det er muligt at dele mikroskopien med andre f.eks. ved multidisciplinære konferencer, ved undervisning samt ved revision af materiale fra andre afdelinger. Ligeledes har arkivering af vævssnittene tidligere krævet fysisk plads, og yderligere er der risiko for at materialet bortkommer eller arkiveres forkert, hvorefter det ikke er muligt at finde det igen. Det digitaliserede vævssnit kan arkiveres på en server, og kan findes frem ved en simpel søgning. (46) Den mere avancerede digitale billedanalyse har, i forskningsmæssige sammenhænge, tidligere vist sig anvendelig på flere områder indenfor den histologiske diagnostik, både som forbedring af den diagnostiske nøjagtighed og til identificering af specifikke elementer i vævet (47).

Ved skanning af vævssnit digitaliseres det, og kan derefter bruges til automatiseret billedanalyse. Det digitaliserede vævssnit består nu af pixels, hvor hver pixel har en position i billedet samt en farveværdi sammensat af værdien af rød/grøn/blå (RGB værdi). Denne værdi kan nu bruges til udvikling af applikationer til analysen. Billedbehandlingen består oftest først af en præprocessering, hvor bestemte strukturer eller elementer i vævet fremhæves, og eventuel støj undertrykkes. Efterfølgende analyseres billedet, hvor målet segmenteres og kvantificeres. Når først en applikation er udviklet, er det muligt at udføre en standardiseret analyse af mange vævssnit med begrænset interaktion af brugeren. (48)

Applikationer til digital patologi kan udvikles på forskellige måder. I dag er anvendelsen af kunstig intelligens meget udbredt, hvor der tidligere blev anvendt applikationer med en manuelt konstrueret (*handcrafted*) algoritme til digital billedanalyse. Ved en *handcrafted* metode opstiller brugeren algoritmer til segmentering af specifikke elementer i vævet ved f.eks. *"thresholding"*, *"*Bayes klassificering" eller *"K-means clustering"*, hvorefter de identificerede celler tælles, eller arealet kan beregnes (47). Tidligere studier har anvendt denne metode på immunhistokemiske farvninger. Her skal applikationen identificere og kvantificere forskellige farvede celler, f.eks. rødt cytoplasma og brune cellekerner hver for sig, eller brune cellekerner som er omgivet af rødt cytoplasma (49).

2.3.1 Kunstig intelligens

Kunstig intelligens blev præsenteret første gang i 1950'erne, som et begreb for den del af computervidenskaben hvor en computer anvendes til at tolke information og træffe beslutninger, som et menneske ville have gjort med sit intellekt.

Kunstig intelligens kan anvendes i flere niveauer; enten *machine learning* eller *deep learning*. Se figur 6 for illustration over niveauer i kunstig intelligens.



Machine learning's applikationer er velegnede til data med veldefinerede karakteristika. Brugeren udvælger de egenskaber og filtre, der bedst fremhæver de karakteristika, som skal defineres i billedet. Efterfølgende annoterer brugeren de elementer i data, som skal klassificeres. Applikationen trænes ud fra de tidligere valgte egenskaber, og heraf opstilles en algoritme ud fra avancerede statistiske metoder. En trænet *machine learning's* applikation kan nu analysere nyt data og træffe en beslutning ud fra en sandsynliggørelse med algoritmen. (47,50) Hvis det data, som skal analyseres, har mere komplicerede karakteristika, såsom den biologiske eller analytiske variation, som ses i vævssnit eller tilfælde hvor kontrasterne i billedet er subtile, kan det være en udfordring at træne en traditionel *machine learning's* applikation tilstrækkeligt til at den kan generalisere på det rette grundlag. Derfor kan der med fordel anvendes *deep learning* (DL) til træning og analysering i de tilfælde, hvor de enkelte karakteristika er varierende eller omfattende. (47)

Ved DL trænes applikationen på samme måde som ved *machine learning*, dog er det ikke nødvendigt, at brugeren definerer karakteristika og softwaren opstiller algoritmen gennem et neuralt netværk. Igennem træningen identificerer netværket selv forskellige egenskaber i billedet, som adskiller de forskellige klasser. Ved et repræsentativt træningssæt, vil applikationen have evnen til at tage højde for den biologiske og analytiske variation, samt at kunne definere de subtile kontraster, som kan findes i et vævssnit. Træningen gennemføres

typisk som superviseret træning, men kan superviseres på forskellige niveauer alt efter målet. Enten kan hele billedet klassificeres, hvor softwaren tillærer sig forskellige egenskaber i det hele billede eller forskellige elementer i billedet kan annoteres af brugeren, hvor softwaren identificerer forskellige egenskaber til adskillelse af de enkelte elementer. (47)

2.3.2 Digital patologi som beslutningsstøtte

Den patologiske diagnostik i dag er præget af subjektivitet, forskel i den visuelle opfattelse og forskel i beslutningen patologerne i mellem, og kan derved medføre en diskrepans i diagnosen. (45) Til trods for at den digitale patologi har vist at kunne bidrage til en mere nøjagtig diagnostik, er implementeringen af digital patologi i diagnostikken dog stadig begrænset (51).

Superviseret træning af en applikation til digital patologi er dybt afhængig af, at elementerne i vævet er annoteret korrekt, hvorfor det er vigtigt, at træningen udføres i tæt samarbejde med en speciallæge i patologi (45). Det er dog vist, at det kan være omfattende og tidskrævende at få annoteret tilstrækkeligt data til træning blandt andet for at imødekomme både den biologiske variation i vævet, såvel som den analytiske variation, i form af varierende snittykkelse og farveintensitet (51). Hvorfor det kan være relevant at finde alternative metoder til annotering af træningsdata.

Flere studier italesætter vigtigheden i, at en teknologi skal implementeres på et videnskabelig grundlag, førend patologer vil inddrage det i den daglige rutine (47,51). Dog er netop anvendelsen af kunstig intelligens i den patologiske diagnostik kritiseret, fordi grundlaget for beslutningerne kan være vanskelige at dokumentere, og derved ses manglende transparens (45,47). I Europa skal apparatur samt software være CE-godkendt, før det må markedsføres som et diagnostisk redskab. CE-mærkningen viser, at den gældende lovgivning for medicinsk udstyr er overholdt. (52) Der findes i dag flere CE-godkendte applikationer til digital patologi på markedet (53). Dog kan *in-house* applikationer i digital patologi anvendes som beslutningsstøtte, uden dette kræver en CE-godkendelse, eksempelvis kan digital patologi anvendes som en *second opinion* på en diagnose stillet af en patolog, såfremt applikationen er trænet til det (47).

I dette afsnit opsummeres problemanalysen i problemafgrænsningen og efterfølgende præsenteres projektets problemformulering samt forskningsspørgsmål, som anvendes til belysning af problemformuleringen.

3.1 Problemafgræsning

Targeteret behandling mod BRAF-mutationen ved malignt melanom har vist god effekt for patienter, hvor der ikke tidligere har været et effektivt behandlingstilbud. Behandlingen kræver dog, at BRAF-mutationen er identificeret ved den enkelte patient, førend den kan Analysen til identificering af denne mutation stiller et krav tilbydes. til tumorcellekernefraktionen bl.a. for at sikre analysens sensitivitet, dvs. at hvis tumorcellekernefraktion er under 20%, er der risiko for et falsk negativt svar. Yderligere anvendes tumorcellekernefraktionen i tolkningen af analysesvaret til vurderingen af omfanget af mutationen i tumoren, således at hvis der er stor forskel mellem antal af fundne BRAFmuterede alleler og tumorcellekernefraktionen, kan det resultere i, at behandlingen fravælges. Det er derfor væsentlig, at den vurderede tumorcellekernefraktion er så nøjagtig, som muligt. På Patologi, AUH foretages vurderingen af tumorcellekernefraktionen i dag subjektivt af en patolog, en metode som er vist, i flere studier, at være upræcis og unøjagtig. Derfor er det ønskeligt at finde en mere valid og pålidelig metode, som dermed kan sikre både sensitiviteten såvel som tolkningen af BRAF-analysen. Forskning har vist, at kunstig intelligens til digital patologi, med fordel, kan anvendes til en mere objektiv vurdering af vævssnit, såfremt denne er trænet tilstrækkeligt. Det kan dog være vanskeligt at sikre tilstrækkeligt repræsentativt data til træningen, hvorfor det er relevant at udvikle en metode til annotering af træningsmateriale, således at udgangspunktet for træningen af netværket optimeres.

Formålet med dette projekt er derfor at træne et kunstigt neuralt netværk til korrekt klassificering af celler, og efterfølgende opsættes det trænede netværk som en applikation til bestemmelse tumorcellekernefraktionen i vævssnit af malignt melanom. I forbindelse med projektet udvikles en metode til annotering af forskellige cellekerner til træning af det kunstige neurale netværk. Applikationen skal i fremtiden implementeres som et vigtigt støtteredskab til speciallæger i patologi, som varetager diagnostikken af maligne melanomer. På baggrund af dette opstilledes følgende problemformulering.

3.2 Problemformulering

Kunstig intelligens kan anvendes til udvikling af en applikation, som kan estimere tumorcellekernefraktionen i et digitaliseret HE-farvet vævssnit fra både primær tumor og metastaser, såvel som fra excisionsbiopsier og grovnålsbiopsier, med større nøjagtighed end den subjektive vurdering, som foretages i forbindelse med diagnostikken.

For at kunne belyse problemformuleringen vil jeg:

- Udvikle en handcrafted algoritme til annotering af cellekerner i vævssnit.
- Træne et kunstigt neuralt netværk til segmentering og klassificering tumorcellekerner og andre cellekerner i et HE-farvet vævssnit.
- Undersøge applikationens evne til at kunne segmentere og klassificere korrekt.
- Undersøge henholdsvis *eye balling* og applikationens nøjagtighed i vurderingen af tumorcellekernefraktion i forhold til *gold standard*.

I dette kapitel præsenteres indledningsvist løsningstrategien for projektet, herunder en beskrivelse af relevant teori. Herefter beskrives det anvendte materiale i projektet, hvor der herunder beskrives de etiske overvejelser, der er gjort i forbindelse med indsamling af materialet, efterfulgt af beskrivelsen af identificering og præparering af materialet. Efterfølgende beskrives de anvendte metoder til udvikling af applikationen. Afslutningsvis beskrives anvendte metoder til evaluering af applikationen.

4.1 Løsningsstrategi

Formålet med dette projekt er at udvikle en applikation ved hjælp af kunstig intelligens, til estimering af tumorcellekernefraktionen i et digitaliseret HE-farvet vævssnit. Dette formål vil opnås ved at træne det konvolutionelle neurale netværk (*convolutional neural network,* CNN) U-NET til at segmentere og klassificere henholdsvis tumorcellekerner og andre cellekerner i et digitaliseret HE-farvet vævssnit og efterfølgende beregne fraktionen af tumorcellekerner ud af alle cellekerner.

U-NET er tilgængeligt i Visiopharm Integrater System (VIS; Visiopharm A/S, Hørsholm, DK), som er en software udviklet til digital patologi. VIS anvendes i dag på Patologi, AUH til forskellige forskningsformål inden for den digitale patologi, f.eks. til identificering og beregning af en celle-proliferationsrate eller immunscore. I VIS findes flere moduler med forskellige funktioner. *Viewer* modulet anvendes til den digitale mikroskopi, hvor der nemt kan navigeres og zoomes i digitaliserede vævssnit. Modullet *Tissuealign* anvendes til digital sammensætning af flere vævssnit, så det er muligt at identificere et område fra et vævssnit på et andet vævssnit. *Image analysis* modulet er udviklet specielt til det histopatologiske speciale, og kan anvendes både til opstilling af *handcrafted* algoritmer, såvel som træning af et neuralt netværk, herunder *deep learning. Stereology* modulet anvendes bl.a. til tælling af profiler i et vævssnit. (54)

4.1.1 Konvolutionelle neuralt netværk

Ved anvendelsen af kunstig intelligens opstilles algoritmen ved træning af et kunstigt neuralt netværk. Der findes flere forskellige typer af kunstige neurale netværk, som defineres af opbygningen af netværket. En type af neurale netværk er et konvolutionelt neuralt netværk, som anvendes i dette projekt. Netværket består af et input lag, hvor et billede præsenteres for netværket, skjulte lag hvor forskellige egenskaber i billedet defineres og et output lag, hvor inputtet er blevet tildelt forskellige klasser. (55)

I input laget opdeles billedet i mindre billeder, f.eks. 512x512 pixels. I de skjulte lag, hvor forskellige egenskaber defineres, udføres tre forskellige operationer:

- konvolution, en matematisk funktion hvor filtre anvendes i en bestemt størrelsesmatrix, f.eks. en 3x3 matrix. Filtrene udtrækker bestemte egenskaber i billedet, og der dannes et *featuremap*.
- Aktiveringsfunktion, hvor værdierne fra *featuremap* omregnes efter en given funktion. F.eks. en ensretning (*rectified linear unit; ReLu*), en matematisk funktion hvor negative værdier sættes til nul, og derved er det kun er positive værdier som overføres til næste lag.
- Pooling, en funktion hvor værdierne i *featuremap* opsamles og reduceres. F.eks. en max pooling, hvor en ny matrix med den største værdi føres videre til næste lag.

Processen i de skjulte lag gentages, indtil et tilstrækkeligt antal egenskaber er identificeret. Dette er afhængig af hvilket CNN, der anvendes. Efterfølgende bliver billedet klassificeret i output laget. (56)

Der er forskellige tilgængelige CNN, hvor U-NET er udviklet til billeder inden for det biomedicinske speciale, hvorfor dette vil være anvendeligt til digital patologi (57).

Træning af neurale netværk

Træningen af netværket udføres som superviseret træning, hvor de klasser, som netværket skal genkende, annoteres af brugeren. I træningen præsenteres netværket nu for billeder som input og med de definerede klasser som output.

Netværkets præstation er afhængig af forskellige træningsparametre, hvor disse skal defineres forud for træning, og kan justeres løbende igennem træningen ved behov, således at netværket lærer optimalt af træningsmaterialet. Disse parametre kan f.eks. være læringsraten, tabsfunktionen og batchstørrelsen, ligeledes kan antallet af epoker varieres. (58)

- Læringsraten bestemmer, hvor meget vægtene skal justeres i løbet af træningen, og det anbefales at indstille denne så højt, som det er muligt uden at det forårsager uoverensstemmelser. Typisk vil denne værdi være mindre end 1 og større end 10⁻⁶, hvor en værdi på 1 angiver et balanceret datasæt uden uoverensstemmelser.
- Tabsfunktionen er den matematiske funktion, som skal beskrive det netværket skal lære. Dvs. den formel, som anvendes i beregningen af uoverensstemmelsen mellem netværkets output og den label som er angivet af brugeren.
- Batchstørrelsen er det antal vævssnit som passerer igennem netværket uden justering af vægtene. Dvs. jo større batch, jo hurtigere træning, hvilket vil spare tid i træningen. Dog kan netværkets generaliserbarhed falde ved et større batch.
- Antallet af epoker er det antal gange hele træningssættet (alle input-output par) er passeret gennem netværket, og bør øges, indtil der ses en lille fejlrate. (58,59)

Træningen foregår i iterationer, hvor et input og et output propagerer gennem netværket. Første input propagerer gennem netværket i den naturlige retning, med vilkårlige værdier i hver vægt. Netværkets output sammenlignes nu med de prædefinerede klasser og ud fra dette beregnes en fejlrate, herefter modpropagere brugerens output. Afvigelsen mellem de to modsatrettede propagationer benyttes nu til indstilling af vægtene i de enkelte lag. For efterfølgende iterationer vil denne afvigelse mellem propagationerne blive mindre og fejlraten vil ligeledes blive mindre. Fejlraten bør løbende monitoreres af brugeren, og træningen stoppes, når fejlraten er tilpas lille, fordi det må formodes at netværkets evne til at generalisere er bedst ved en lav fejlrate. (56)

Netværkets evne til at generalisere valideres løbende gennem træningen, hvilket gøres med et nyt datasæt, som ikke tidligere er præsenteret for netværket. En dårlig generalisering skyldes oftest en af to ting; at netværket er *underfitted* eller *overfitted*. *Underfitting* sker når netværket ikke har lært nok eller ikke er i stand til at lære tilstrækkeligt, f.eks. hvis annotationer er forkerte eller uspecifikke. *Overfitting* sker bl.a. hvis det anvendte træningssæt er for lille, således at evnen til at generalisere på et nyt sæt er begrænset. (56) Overfitting kan til dels imødekommes ved data augmentation, hvor træningsbillederne flippes, drejes eller spejlvendes, så billederne vil fremstå som et nyt input.

Såfremt valideringen godkendes, og netværket derved kan generalisere tilstrækkeligt, kan det nu anvendes på nye og ukendte billeder.

4.1.2 U-NET

U-NET består af flere konvolutionelle lag, som er opdelt i to dele, hvilket giver den u-formede struktur. Figur 7 er en skitsering af U-NETs arkitektur. Den første del er opbygget som et traditionelt konvolutionelt neuralt netværk, og består af to 3x3 konvolutioner, som hver er fulgt af en ensretning og en 2x2 max pooling. I denne del udtrækkes kun information om specifikke egenskaber i billedet, og den spatiale information reduceres. Den anden del af netværket består af en op-konvolutionerende del, hvor de definerede egenskaber fra første del kombineres med billedets spatiale information igen. Outputtet er nu segmenteret og klassificeres med en softmax funktion, som angiver sandsynligheden for, at det segmenterede tilhører de forskellige klasser. (57)



Træning af U-NET

Der kan ikke defineres et mål for omfanget af tilstrækkeligt data til træning af et neuralt netværk, da der er flere forskellige faktorer, som har indflydelse herpå. Disse faktorer kan f.eks. være den morfologiske og biologiske variation, som ses i forskellige vævstyper i et vævssnit og variationen mellem lavt differentierede tumorceller, hvor cellernes udgangspunkt kan være vanskelig at definere ud fra et HE-farvet vævssnit. Yderligere kan den analytiske variation i form af varierende snittykkelse, farveintensitet og forstørrelsen ved indskanningen have en betydning. (61) Derfor er det en kendt udfordring ved anvendelse af *deep learning*, at der kræves et stort datasæt til træning for at sikre et tilstrækkeligt repræsentativt træningsmateriale og derved en endelig optimal generaliserbarhed (62). Ligeledes er det vigtigt, at de annoteringer, som anvendes i træningen, er af høj kvalitet, og derfor anbefales det, at en speciallæge i patologi udfører denne opgave. Det er dog et tidskrævende og ensformigt arbejde, hvorfor der ofte opstår et flaskehalsproblem ved dette tidlige trin i anvendelsen af *deep learning* til digital billedanalyse. (62)

For at imødekomme de udfordringer der kan opstå, både med et begrænset datasæt til træning og ligeledes afhængigheden af speciallægens kompetencer, som er nødvendige for at sikre kvaliteten af annotationerne, er der i forbindelse med dette projekt udviklet en metode til annotering af forskellige cellekerner i vævssnit af malignt melanom.

I den daglige histopatologiske diagnostik anvendes en traditionel Hæmatoxylin-Eosin farvning (HE farvning) på alle vævssnit. Diagnostikken kan yderligere suppleres med forskellige immunhistokemiske farvninger (IHC), målrettet bestemte proteiner udtrykt på specifikke celler. IHC anvendes som redskab til at kunne skelne forskellige celler fra hinanden, og kan ligeledes anvendes til at definere udgangspunktet for lavt differentierede tumorceller. I diagnostikken af malignt melanom anvendes SOX-10 som en immunhistokemisk markør specifik for melanocytære celler (63). Derfor kan IHC med SOX-10 benyttes i dette projekt til identificering af melanocytære celler og derved muliggøre, at annoteringen af forskellige celler til træningen kan uddelegeres. Det er dog essentielt, at netværket trænes på det HEfarvede vævssnit af melanomet, fordi det rutinemæssigt er det snit som anvendes i analysen af BRAF. Der udvikles derfor en metode til annotering af SOX-10 positive og SOX-10 negative cellekerner, som efterfølgende overføres til det HE-farvede vævssnit til træning af netværket. Ved annoteringsmetoden digitaliseres først et HE-farvet vævssnit af melanomet, som efterfølgende farves igen med SOX-10, så tumorcellekernerne nemt kan identificeres, og vævssnittet digitaliseres igen. De 2 digitaliserede vævssnit, som stammer fra det samme vævssnit kan nu overlejres i en digital dobbeltfarvning vha. en digital aligning. Herved er det muligt på cellekerneniveau at identificere en cellekerne fra det ene vævssnit i det andet vævssnit. Til det digitaliserede SOX-10 farvede vævssnit designes en handcrafted algoritme til annotering af SOX-10 farvede cellekerner samt andre cellekerner, som er negative for SOX-10. SOX-10 positive cellekerner annoteres med en tumorcellekerne label og alle andre cellekerner med normal cellekerne label. Disse labels kan nu overføres til det digitaliserede HE-farvede vævssnit. De overførte labels til det digitaliserede HE-farvede vævssnit anvendes herefter som annoteringer til træning af U-NET i dette projekt. Processen i denne metode er illustreret i figur 8.



Figur 8 illustration af processen i annotering af cellekerner vha. en *handcrafted* algoritme.

Netværkets evne til segmentering og klassificering af de forskellige cellekerner i et vævssnit valideres løbende ved en manuel inspektion af dets resultater i valideringssættet med alignede HE-SOX-10 vævssnit.

Efter en godkendt validering testes netværket på et testsæt, hvor netværkets evne til at klassificere korrekt evalueres op imod en *gold standard*. Dette gøres ved en analyse af, hvorvidt netværkets klassificering i form af tildelte labels er korrekt i forhold til de labels, som er givet på det korresponderende immunfarvede snit ved den *handcraftede* algoritme. Da en sådan *handcrafted* algoritme ikke kan være 100% valid, suppleres resultatet af denne med en manuel inspektion af både HE- og SOX-10 farvningen med eventuelle tilretninger, således der opnås en mere pålidelig *gold standard*. Ligeledes evalueres netværkets nøjagtighed i vurderingen af tumorcellekernefraktionen ud fra en manuel tælling af cellekerner i det immunfarvede snit, som ligeledes kan betragtes som *gold standard*. Efterfølgende vurderes patologens *eye ballings* nøjagtighed også ud fra *gold standard*, hvorefter netværkets og *eye ballings* nøjagtighed sammenlignes til vurdering af den mest optimale metode til vurdering af tumorcellekernefraktionen.

4.2 Anvendt materiale

4.2.1 Etiske overvejelser

I dette projekt inkluderes udelukkende vævssnit af malignt melanom, hvor der i forbindelse med diagnostikken er foretaget en vurdering af tumorcellekernefraktionen, og efterfølgende foretaget en BRAF-analyse med henblik på valg af behandling. Alt det inkluderede materiale er total anonymiseret umiddelbart efter indsamling af materiale og relevante oplysninger om vævsprøven. Dette er gjort for at imødekomme den problemstilling, der vil kunne opstå såfremt den vurderede tumorcellekernefraktion i diagnostikken afviger drastisk fra den talte fraktion, som i dette projekt anvendes som *gold standard*. Problemstillingen vil opstå, fordi tolkningen af BRAF-analysen er afhængig af tumorcellekernefraktionen, og hvis denne er afvigende fra *gold standard*, kunne tolkningen eventuelt være anderledes, hvilket kan have en betydning for valg af behandling af den enkelte patient.

Regulatoriske instanser

Forud for ethvert projekt, som involverer patienter og/eller patientens materiale, skal man forholde sig til de forskellige regulatoriske instanser, der varetager patienternes interesser.

Det "Nationale Videnskabsetisk Komitésystem" (NVK) udspringer af Helsinki-deklarationen, og skal sikre, at et sundhedsvidenskabeligt projekt udføres etisk forsvarligt. Dette projekt er, ifølge NVK's retningslinjer ikke anmeldelsespligtig, fordi det kan betragtes som et kvalitetsudviklingsprojekt. Projektet er et retrospektivt projekt med formålet at udvikle en ny metode inden for et allerede etableret område, og uden at dette vil få en betydning for de inkluderedes patienters behandling. Ligeledes er det inkluderede materiale i projektet fuldstændig anonymiseres, eftersom materialet hverken direkte eller indirekte kan henføres til bestemte personer. (64) NVK anbefaler dog, at der anmodes om en konkret vurdering af anmeldelsespligten for hvert projekt, således det efterfølgende kan dokumenteres, at projektet er vurderet af komitéen (65). Der er i dette projekt anmodet om en vurdering af anmeldelsespligten d. 8. december 2020 ved "De Videnskabsetiske Komitéer for Region Midtjylland", hvorefter komitéen dokumenterer at projektet, i henhold til lovgivningen, ikke er anmeldelsespligtet.

Datatilsynet skal sikre, at patienters oplysninger behandles i henhold til Databeskyttelsesloven og Databeskyttelsesforordningen. Dette skal sikre, at man som patient trygt kan overlade sine data til myndighederne, uden at disse misbruges, og at patientens rettighed over egne data beskyttes. Derfor skal Datatilsynet informeres, hvis der i et forskningsprojekt skal indsamles og behandles oplysninger, som er direkte personhenførbare. (66) Som databehandlende sundhedsmyndighed sker anmeldelsen til Datatilsynet under en paraplyanmeldelse i de enkelte regioner, hvor det i Region Midt er "Intern fortegnelse over forskningsprojekter med Region Midt som dataansvarlig", som varetager denne anmeldelse (67). Til trods for at dette projekt kun behandler fuldstændig anonymiseret materiale, er den interne fortegnelse over forskningsprojekter informeret om, at vi ønsker at indsamle data, men at disse er anonymiseret forud for databehandlingen.

Vævsanvendelsesregistret (VAR) er et register, hvor man som patient til enhver tid kan nedlægge forbud mod at ens materiale kan anvendes til andet end patientens egen primær diagnostik. Projekter som inkluderer væv fra patienter, skal altid foretage opslag i VAR for at sikre, at ingen af de inkluderede patienter har nedlagt forbud mod anvendelse. (68) Der er i dette projekt foretaget opslag i VAR på alle de inkluderede patienter, hvortil der ikke blev identificeret patienter, som har nedlagt forbud mod anvendelse.

4.2.2 Dataindsamling og præparering

Der anvendes, i dette projekt, udelukkende formalinfikseret paraffinindstøbt materiale, oprindeligt fra den diagnostiske biobank. Der inkluderes materiale i tre forskellige datasæt til henholdsvis træning, validering og test. Efter indsamling farves vævssnittene og disse digitaliseres, hvorefter de digitaliserede vævssnit overføres til softwaren til videre behandling. Processen for præparering af data illustreres i figur 9. Materialet (N=99) opdeles i et træningssæt bestående af 60%, et valideringssæt bestående af 10% og et testsæt bestående af 30% vævssnittene.



Figur 9 - Illustration af processen i præpareringen af datamaterialet.

Træningssæt og valideringssæt

Materialet til træningssættet samt til valideringssættet findes i overskydende materiale fra et tidligere forskningsprojekt (69) på maligne melanomer og tilhørende metastaser. De overskydende vævssnit er anonymiseret og gemt i en forskningsbiobank efter afslutningen af det tidligere projekt. Til træningssættes udvælges et vævssnit fra 15 forskellige primære melanomer og 15 forskellige metastaser vilkårligt. Valideringssættet består af 5 primære og 4 metastaser, som ligeledes er vilkårligt udvalgt.

Vævssnittene til træning og validering farves med HE på HE 600 (Roche Diagnostics, Tucson, AZ, USA), hvorefter vævsnittene digitaliseres ved indskanning på NanoZoomer 2.0-HT scanner (Hamamatsu Photonics, Japan) ved 20x forstørrelse. Efterfølgende afmonteres dækglasset, og vævssnittet farves nu med en immunhistokemisk farvning for SOX-10. Se figur 10 for illustration af samme vævssnit farvet med henholdsvis HE og SOX-10. Den immunhistokemiske farvning udføres med den standardprotokol på Ventana Ultra (Roche Diagnostics, Tucson, AZ, USA), som anvendes i diagnostikken, hvor SOX-10 visualiseres med DAB som kromogen. SOX-

10 positive celler vil fremstå med brune cellekerner. Efterfølgende digitaliseres vævssnittet igen.



Figur 10 - Vævssnit af malignt melanom. Tv. ses vævssnittet farvet med HE og th. ses vævssnittet farvet med SOX-10.

Alle digitaliserede vævssnit importeres til VIS. I modulet *Tissuealign* kan de indskannede vævssnit nu alignes med nøjagtighed. Ved aligning overlejres de to skanninger fra samme vævssnit digitalt således, at de to billeder kan visualiseres simultant på celleniveau. Dette er muligt, fordi det er det samme vævssnit, der er indskannet to gange, blot med forskellige farver på. Se figur 11 for illustration af aligning af indskannede vævssnit.



Figur 11 - Vævssnit som er alignet på celleniveau. Tv. HE-farvet vævssnit og th. SOX-10 (*fast red*) farvet vævssnit. Der ses yderligere brunt pigmentering i vævssnittet.

Ved træning med det første træningssæt er der dog identificeret en udfordring ved den DABfarvede markering af SOX-10, hvor det er vanskeligt at skelne mellem den brune DAB og det naturlige pigment som findes i varierende omfang i melanocytter og melanofager. IHC kan dog også rutinemæssigt visualiseres med *fast red*, som fremstår rød, og derved er det nemmere at skelne den røde markering af SOX-10 fra pigmentet i huden. Derfor udvælges der yderligere vævssnit fra 15 primære maligne melanomer og 15 metastaser fra samme forskningsbiobank som tidligere. Disse vævssnit følger samme procedure med HE-farvning, digitalisering og efterfølgende immunhistokemisk farvning med SOX-10, dog er SOX-10 denne gang visualiseret med det røde kromogen, *fast red*. Se figur 12 som illustrerer to vævssnit farvet med henholdsvis *fast red* og DAB sammen med det naturlige pigment i melanomet. Disse nye vævssnit digitaliseres og alignes ligeledes, og indgår som en del af træningssættet.



Figur 12 - 2 forskellige vævssnit af malignt melanom. Tv. ses vævssnit farvet med fast red kromogen og th ses vævssnit farvet med DAB kromogen. De gule pile markere pigment i melanomet.

Testsæt

Testsættet består af 30 vævssnit af maligne melanomer eller metastaser derfra.

Vævsmaterialet identificeres ved en søgning på SNOMED-koder i Patobanken. Alle vævsprøver der undersøges på en afdeling for patologi i Danmark, kodes med SNOMED-koder, som beskriver materialet, og registreres efterfølgende automatisk i den landsdækkende database, Patobanken. De enkelte afdelinger kan kun søge på materiale undersøgt på egen afdeling, hvorfor der i dette projekt, kun er søgt på materiale undersøgt på Patologi, AUH. SNOMED-koderne kan identificeres på hjemmesiden Patobank.dk. Søgningen i Patobanken vil bestå af følgende inklusionskriterier:

• Inklusionsperiode 1/1/2020 til 1/5/2021.

• SNOMED malignitetskoder:

- M8720*, trunkeret søgning for malignt melanom, primær, recidiv og metastase.
- M8743*, trunkeret søgning for superficielt spredende malignt melanom, primær, recidiv og metastase.

Disse er valgt for at ekskludere andet end malignt melanom, da BRAF rutinemæssigt også undersøges på andet materiale end malignt melanom.

• SNOMED funktionskode:

• FE13E*, trunkeret søgning for status på BRAF

Denne er valgt for at kunne identificere det materiale som tidligere er undersøgt for BRAF i forbindelse med patientens primære diagnostik.

Der inkluderes udelukkende materiale fra maligne melanomer eller metastaser derfra, som tidligere er fundet positiv for BRAF-mutation. Af det inkluderede materiale vil der, fra den diagnostiske biobank, indsamles det HE-farvede vævssnit, hvor der tidligere er foretaget en vurdering af tumorcellekernefraktionen. Dette er valgt fordi, der er foretaget en vurdering af tumorcellekernefraktionen på det pågældende materiale i forbindelse med den rutinemæssige diagnostik. Dermed sikres det, at projektets undersøgelser laves på samme udgangspunkt, som vurderingen i forbindelse med diagnostikken. Ligeledes undgås den bias som kan opstå, hvis tumorcellekernefraktionen skal vurderes ved eye balling, når patologen kender formålet med vurderingen, og derved kan være mere omhyggelig med vurderingen. De indsamlede vævssnit nummereres med projekt ID fortløbende fra 1-30 og data i form af tumorcellekernefraktion, materialeudgangspunkt (primær eller metastase) samt materialetype (grovnålsbiopsi eller excisionsbiopse). Disse oplysninger registreres i dataopsamlingsskemaet, vedlagt som Bilag B. Yderligere indsamles BRAF-mutations status og allelfrekvensen for fundne BRAF-muterede alleler, som skal bruges til senere analyser. Vævssnittene præpareres på samme måde som trænings- og valideringssættet, hvor vævssnittene digitaliseres og efterfølgende afmonteres dækglasset, hvorefter vævssnittene farves med SOX-10 visualiseret med fast red og digitaliseres igen. Efterfølgende overføres de digitaliserede vævssnit til VIS.

4.3 Udvikling af applikation

En applikation til digital patologi udvikles gennem flere trin; først annoteres træningsmateriale og efterfølgende trænes netværket og valideres løbende. Afslutningsvis evalueres netværkets evne til at generalisere på ikke tidligere anvendt materiale.

4.3.1 Annotering af cellekerner

I forbindelse med dette projekt er der udviklet en metode til præcist, at kunne annotere et stort og repræsentativt omfang af forskellige cellekerner. Dette er gjort for at imødekomme udfordringen med et for lille træningssæt samt afhængigheden af en patolog til at foretage annoteringerne.

I modulet *Image analysis* er der opstillet en *handcrafted* algoritme, til identificering og annotering af SOX-10 positive cellekerner, blå cellekerner, bindevæv og ufarvet baggrund i de immunfarvede digitaliserede vævssnit. Den *handcraftede* algoritme var indledningsvist opsat til detektering af brune cellekerner. Dette er gjort med en præprocessering indeholdende en *deconvolution* af den brune farve, som klassificeres med *thresholding* og tildeles en label. Efter ændring af SOX-10 farvningen til anvendelsen af *fast red* ændres *deconvolution* til den røde farve. Efterfølgende identificeres de blå cellekerner ud fra en rød kromaticitet, hvor de blå kerner vil fremhæves, som igen klassificeres med en *thresholding* og tildeles en label. De enkelte cellekerner er forsøgt adskilt allerede i præprocessering med et polynomial filter (blobs filter), som fremhæver runde strukturer. Figur 13 illustrerer de forskellige præprocesseringer til identificering og adskillelse af cellekerner.



Figur 13 - Illustration af de 3 forskellige præprocesseringer af billedet til segmentering af hhv. de brune og de blå cellekerner.

Efter klassificering af billedet forbedres dette resultat med en postprocessering, hvor bl.a. både tumor og normale cellekerner sorteres ud fra areal og forsøges separeret yderligere ud fra blobs filtret og en *watershed* algoritme på baggrund af cellekernernes diameter. Figur 14 illustrerer det klassificerede billede og det endelige billede efter postprocessering.



Figur 14 - Illustration af klassificeringen. Tv første klassificeret billede og th den endelige klassificering efter postprocessering. Tumorceller markeres med en rød label og normale celler med en grøn label.

I hvert immunfarvet digitaliseret vævssnit udvælges repræsentative områder til annotering, og denne annotering gennemgås efterfølgende, og tilrettes ved eventuelle fejl. Annoteringen kan nu overføres med stor nøjagtighed til det HE-farvede digitaliserede vævssnit, hvor annoteringen fra de SOX-10 farvede cellekerner overføres som en tumorlabel (rød label), og de blå cellekerner overføres som label til alle andre cellekerner (grøn label). Se figur 15 for illustration af nøjagtigt overførte annoteringer.



Figur 15 - Illustration af labels på et sox-10 farvet vævssnit overføres på celleniveau til de korresponderende HE-farvet vævssnit.

Der er med denne metode annoteret ca. 48.500 tumorcellekerner og 48.500 normale cellekerner, fordelt på 60 vævssnit, til træning af det neurale netværk. Heraf er 34 vævssnit fra metastaser med annoteringer af ca. 31.000 tumorcellekerner og 24.500 normale cellekerner. De resterende 26 snit af træningssættet er vævssnit fra primære maligne melanomer med annoteringer af ca. 17.500 tumorcellekerner og 24.000 normale cellekerner.

Træning af U-NET

Kunstig intelligens anvendes ligeledes i modulet *Image analysis,* hvor *deep learning* vælges som klassificeringsmetode. Efterfølgende vælges det neurale netværk, som skal trænes, hvor der i dette projekt anvendes det konvolutionelle neurale netværk U-NET, som er udviklet til det biomedicinske speciale (57).

Forud for træningen indstilles netværket, herunder input og sampling, netværksparametre, træningsparametre og dataaugmentation for at sikre en optimeret træning.

Input og sampling

Input størrelsen anvendes til at kontrollere størrelsen af alle billederne som sendes gennem netværket. Det anbefales, at ved cellekernesegmentering skal input sættes til 256 eller 512,

hvorfor det indstilles til 512, da dette vurderes at være et tilstrækkeligt repræsentativt input. Der vælges RGB-farveskala, fordi inputtet er *brightfield* billeder. (59)

Netværksparametre

Netværkets dybde kan indstilles alt efter hvilke vægte, som skal justeres i løbet af træningen, og netværkets vidde kan ligeledes indstilles efter den ønskede størrelse af det originale U-NET, som der ønskes anvendt. I dette projekt vælges *default* indstillinger, hvor de 2 første blokke i dybden fryses, således at vægtene her ikke justeres gennem træningen, og at der anvendes 100% af netværkets vidde. (59)

Træningsparametre

Træningsparametrene indstilles ud fra tidligere erfaring med træning af U-NET.

Læringsraten indstilles til 10⁻⁶, hvilket indikerer at vægtene justeres med 10⁻⁶ efter hver iteration. En for lav læringsrate vil betyde en langsom træning, fordi det kan kræve et stort antal iterationer førend netværket er trænet, hvor en for høj læringsrate vil gøre det vanskelig at finde den laveste fejlrate. Ved en ideel læringsrate vil fejlraten falde efter hver iteration.

Batchstørrelsen indstilles til 1, så tabsfunktionen evalueres samt vægtene justeres efter hvert billede, som er sendt gennem netværket. Denne batchstørrelse anbefales ligeledes af udviklerne af U-NET (57).

Tabsfunktionen indstilles til den matematiske formel, som beskriver det problem som netværket løser. Her lærer netværket ved minimering af tabsfunktionen, dvs. jo lavere tabsfunktion, jo bedre præsterer netværket. U-NET anvender *cross-entropy*, hvor netværkets evne til at klassificere ud fra en sandsynlighed beregnes. Dette gøres for hver pixel og efterfølgende som gennemsnit for det hele billede. (59)

Tabsvægtning kan med fordel anvendes, hvis de forskellige klasser i datasættet ikke er repræsenteret ligeligt. Ved indstilling af *class constant weighting* inkorporeres de forskellige klassers vægte i tabsfunktionen, således at den lavest repræsenterede klasse tildeles en højere vægt, hvor en dominerende klasse vil tildeles en lavere vægt. (59)

Dataaugmentation

Det kan vælges, hvilken form for dataaugmentation som ønskes i træningen. I VIS er der defineret to forskellige grupper af dataaugmentation; *Basic* og *H*&*E*.

Basic som indeholder rotation, flipping samt afvigelser i skarphed og kontrast og *H&E*, som indeholder afvigelser af farvetone, farvemætning samt specifikke afvigelser, som kan relateres til HE-farvningen. (59)

Træning påbegyndes nu, og fejlraten overvåges løbende. Når fejlraten vurderes tilpas lille stoppes træningen, og algoritmen gemmes som en applikation navngivet med det antal iterationer som træningen består af. Det er derved muligt at kunne anvende en tidligere version af applikationen, såfremt algoritmen ikke har forbedret sin generaliserbarhed ved yderligere træning eller vurderes *overfitted*. Algoritmen valideres på et udvalgt valideringssæt, der ikke tidligere har været præsenteret for netværket, bestående af digitaliserede HE-farvede vævssnit af 9 maligne melanomer og metastaser. Valideringen består af en visuel vurdering af algoritmens evne til at identificere de aktuelle cellekerner, såvel som dennes evne til at kunne adskille tætliggende cellekerner.

Efter en godkendt validering opsættes postprocesseringsalgoritmer, hvor netværkets resultater forbedres. Disse algoritmer bygger blandt andet på de morfologiske operationer dilatation og erosion, samt ændringer af labels i forhold til størrelse og relation til omkringliggende labels. Ændringen af labels i forhold til relation og størrelse er f.eks. når en kraftig farvet nukleole inde i en tumorcellekerne fejlagtigt klassificeres som 'normal', da den kan ligne en lille lymfocyt. Derfor opstilles postprocesseringsalgoritmen; "hvis en 'normal' er omgivet af 100% 'tumor' og er mindre end 5 μ m² skift 'normal' til 'tumor'. Derved forbedres netværkets klassificering.

Den færdige applikation er nu klar til at blive testet på det udvalgte testsæt.

4.4 Evaluering af applikation

Applikationen testes på testsættet, som ikke tidligere har været præsenteret for netværket. De digitaliserede HE-farvede vævssnit importeres til VIS og på hvert enkelt vævssnit optegnes det ROI, som tidligere er opmærket af en speciallæge i patologi, som værende det område med størst tumorprocent. Efterfølgende analyseres de forskellige ROI med applikationen.

Applikationen evalueres på dens evne til at segmentere og klassificere korrekt, hvor segmenteringen er evnen til at identificere cellekernerne og klassificeringen er evnen til at tildele den korrekte klasse. Applikationen evalueres ligeledes på, hvorvidt den evner at estimere tumorcellekernefraktionen mere nøjagtigt end den subjektive vurdering, som foretages ved *eye balling* i diagnostikken. Efter en overordnet databehandling opdeles data i undergrupperne; materialeudgangspunkt og materialetype. Formålet med opdelingen er at identificere, hvorvidt der er en bestemt gruppe, hvor netværket præsterer mere eller mindre godt for at kunne målrette fremtidig træning.

Den statistiske databehandling, i forbindelse med dette projekt, foretages i SPSS 27 (IBM, Armonk, NY, USA) og i Microsoft Excel 2016 (Microsoft Corporation, Redmond, WA, USA).

4.4.1 Evaluering af segmentering og klassificering

I dette projekt opstilles en klassificeringsalgoritme i Python 3.8, til sammenligning af labels på henholdsvis de digitaliserede HE-farvede vævssnit og de korresponderende digitaliserede SOX-10 farvede vævssnit. Hermed analyseres, hvorvidt de labels som applikationen tildeler cellekernerne i de digitaliserede HE-farvede vævssnit fra testsættet er korrekte. Labels tildelt af den *handcraftede* algoritme skal inspiceres og tilrettes manuelt, da intervaller for dennes *thresholding* var meget snævre, således elementer i baggrunden ikke ved fejl blev klassificeret som en cellekerne. Derved er der cellekerner som ikke klassificeres, fordi de ligger uden for de satte *thresholds.* På ovenstående figur 15 ses det, at ikke alle cellekerner er tildelt en label, hvorfor disse skal tildeles manuelt. Efter tilretningen kan labels i det korresponderende digitaliserede SOX-10 farvede vævssnit, betragtes som *gold standard*.
Efter applikationens analyse gemmes et tilfældigt billede på 256x256 pixel af de tildelte labels uden baggrund fra hver af de digitaliserede HE-farvede vævssnit. Ligeledes gemmes det tilsvarende billede af de korresponderende labels, fra det digitaliserede SOX-10 farvede vævssnit, som er annoteret af den *handcraftede* algoritme og efterfølgende tilrettet. Figur 16 illustrerer 2 korresponderende billeder, som klassificeringsalgoritmen skal analysere for korrekt segmentering og klassificering.



Figur 16 - Korresponderende billeder af tildelte labels. Tv. ses labels til.delt af applikationen og th. ses labels annoteret af den *handcraftede algoritme* og efterfølgende tilrettet.

Indledningsvist identificerer algoritmen de labels, som er tildelt af applikationen, som røde og grønne labels, hvorefter det vha. den digitale aligning er muligt at identificere, hvorvidt den korresponderende label har den samme farve. Der vil for hvert billedpar angives antallet af sandt positive og sandt negative, og hvor afvigelser i applikationens labels angives som henholdsvis falsk positiv eller falsk negativ. Se figur 17 for algoritmens sammenlægning af resultater.



Figur 17 - Resultat fra klassificeringsalgoritmen.

Blå ring viser korrekt segmenterede og klassificerede cellekerner. Lilla ring viser forkert klassificerede cellekerner. Grå ring viser forkert segmenterede cellekerner Afvigelserne kan bestå af både forkert segmenterede kerner, samt forkert klassificerede kerner. Forkert segmentering er tilfælde, hvor applikationen ikke har segmenteret en kerne, som faktisk er der. Forkert klassificering er de tilfælde, hvor applikationen har klassificeret en kerne forkert samt de tilfælde, hvor der er klassificeret en kerne, som ikke er der. Figur 18 skitserer hvordan resultaterne klassificeres.



Figur 18 - Skitsering af definition af klassificeringsalgoritmens resultater.

De tumorcellekerner som applikationen ikke har evnet at segmentere defineres som de falske negative (R FN), da de per definition burde have været positive. De normale cellekerner som ikke er segmenteret (G FN) defineres som falsk positiv, da de burde være negative.

Ud fra klassificeringsalgoritmens resultater beregnes applikationens sensitivitet, specificitet samt nøjagtigheden for evnen til segmentering og klassificering at tumorcellekerner og normale cellekerner. Beregninger foretages på applikationens samlede evne, hvorfor tumorcellekernerne defineres som positive og normale cellekerner defineres som negative.

Sensitiviteten er et udtryk for testens evne til at identificere de sandt positive, og beregnes ved formlen (70):

$$Sensitivitet = \frac{\text{Sandt positive}}{(\text{sandt positive} + \text{falsk negative})}$$

Specificiteten er omvendt et udtryk for testens evne til at identificere de sandt negative, og beregnes ved formlen (70):

$$Specificitet = \frac{\text{Sandt negative}}{(\text{sandt negative} + \text{falsk positive})}$$

Applikationens nøjagtighed er et udtryk for testens evne til generelt at kunne klassificere korrekt, hvad end det er sandt positivt eller sandt negativt, og beregnes ved formlen (70):

$$Nøjagtighed = \frac{(Sandt positive + sandt negative)}{Alle}$$

Applikationens sensitivitet, specificitet og nøjagtighed beregnes for den overordnede evne til at segmentere og klassificere korrekt. Efterfølgende beregnes sensitiviteten, specificiteten og nøjagtigheden for applikationens evne til henholdsvis at segmentere og klassificere korrekt. Tabel 1 skitserer, hvorledes de positive og negative defineres for den samlede evaluering.

Samlet evaluering				
Sandt positive App har lablet sand tumor (R TP)				
Sandt negative App har lablet sand normal (G TP)				
Falsk positiv	App har lablet en normal som tumor (R FP)			
	Manglende label af normal (G FN)			
Falsk negativ App har lablet en tumor som normal (G FP)				
	Manglende label af tumor (R FN)			

Tabel 1 - skitsering af definitioner af positive og negative for den samlede evaluering

For evnen til at segmentere korrekt, vil de ikke segmenterede cellekerner indgå som de falsk definerede grupper. Denne beregnes for at belyse, hvorvidt applikationen evner at segmentere alle cellekernerne. Sensitiviteten, specificiteten og nøjagtigheden beregnes ligeledes for applikationens overordnede klassificeringsevne, hvor tumorklassificeringen defineres som de positive og normalklassificering defineres som de negative. Se tabel 2 for skitsering af definitionerne af positive og negative.

Klassificering		Segmentering	
	App har lablet sand tumor		App har lablet sand tumor
Sandt positive	(R TP)	Sandt positiv	(R TP)
	App har lablet sand normal		App har lablet sand normal
Sandt negative	(G TP)	Sandt negative	(G TP)
Falsk positiv	App har lablet en normal som	Falsk positiv	App har ikke lablet en normal
klasse	tumor (R FP)	segmentering	(G FN)
Falsk negativ	App har lablet en tumor som	Falsk negativ	App har ikke lablet en tumor
klasse	normal (G FP)	segmentering	(R FN)

Tabel 2 - skitsering af definitioner af positive og negative i de to grupper.

Evaluering af applikationens evne til henholdsvis segmentering og klassificering kan efterfølgende anvendes i en senere optimering af algoritmen, fordi det hermed belyses hvilke udfordringer applikationen har.

Data fra klassificeringsalgoritmen er vedlagt i det samlede dataopsamlingsskema, i Bilag B.

4.4.2 Evaluering af tumorcellekernefraktionen

Både applikationens nøjagtighed i vurdering af tumorcellekernefraktionen samt *eye ballings* nøjagtighed evalueres ved overensstemmelsen med en manuel cellekernetælling, som *gold standard*. Derfor indsamles tumorcellekernefraktionen ved de tre forskellige metoder; applikationen, *eye balling* og *gold standard*. Nøjagtigheden af både applikationen og *eye balling* illustreres i hvert sit Bland Altman plot og sammenlignes efterfølgende. Data fra de tre metoder indsættes i det samlede dataopsamlingsskema, vedlagt som Bilag B.

Applikation

Efter applikationens segmentering og klassificering af vævssnittene beregnes fraktionen af tumorcellekerner ud af alle kernerne, som er klassificeret i de enkelte vævssnit. Den beregnede tumorcellekernefraktion noteres i det samlede dataopsamlingsskema, som applikationens tumorcellekernefraktion.

Eye balling

I forbindelse med den primære diagnostik er der, forud for BRAF-analysen, vurderet en tumorcellekernefraktion af ROI ved *eye balling*. Denne information noteres, i forbindelse med indsamling af materialet, i det samlede dataopsamlingsskema.

Gold standard

I testsættet er det digitaliserede HE-farvede vævssnit farvet med SOX-10 efter digitaliseringen. Der er i dette projekt valgt at tælle på det SOX-10-farvede vævssnit for at sikre korrekt adskillelse af tumorcellekerner og alle andre cellekerner. Det digitaliserede SOX-10-farvede vævssnit overføres til VIS, hvorefter ROI indtegnes. *Stereology* modulet i VIS kan med fordel anvendes til manuel tælling af profiler i et digitaliseret vævssnit, fordi det er muligt at markere talte cellekerner, og softwaren tæller alle markeringer i et snit. Ligeledes sikrer modulet, at tællerammerne placeres struktureret tilfældigt, hvilket er nødvendigt for at opnå et repræsentativt resultat (71). I *stereology* modulet angives nu størrelse samt antal af tællerammer, som skal anvendes til manuel optælling af tumorcellekerner og andre cellekerner. Efter vejledning fra professor i stereologi, Jens Nyengaard, vælges tællerammer af størrelsen 100*60 μm og 15 tællerammer tælles i hvert ROI, for at sikre, at der som minimum tælles 100 profiler. Profilerne er markeret af professor i patologi, Torben Steiniche. Figur 19 illustrere et digitaliseret SOX-10-farvet vævssnit med en tælleramme. Resultaterne fra de manuelle optællinger noteres i det samlede dataopsamlingsskema.



Figur 19 - Tælleramme i et SOX-10-farvet vævssnit fra en malignt melanom metastase. Kerner inden i rammen af røde og grønne linjer tælles. Kerner som rører de grønne linjer tælles, hvor kerner som rører de røde linjer ikke tælles.

A er talt som tumorcellekerner og B er talt som normal cellekerne.

Bland Altman plot

Et Bland Altman plot anvendes til sammenligning af 2 forskellige metoder. Plottet består af et scatter plot, hvor x-aksen angiver gennemsnit af hver enkelt parret måling, og y-aksen angiver forskellen mellem de parrede målinger. Yderligere indtegnes den gennemsnitlige forskel på de to metoder som en vandret streg i plottet. En gennemsnitlig forskel på 0 indikerer en god generel overensstemmelse mellem de to metoder. Såfremt de parrede målingers forskel findes normalfordelt, kan der efterfølgende beregnes et 95% konfidensinterval for gennemsnittet af forskellene samt *Limits Of Agreement* (LOA). 95% konfidensintervallet indtegnes i plottet, og såfremt denne indeholder 0, kan det konkluderes, at den eventuelle forskel mellem metoderne ikke skyldes en systematisk fejl. LOA udgør intervallet, hvor 95% af forskellene mellem målingerne vil ligge. Efter beregning af LOA indtegnes både den øvre og nedre grænse for LOA og afstanden mellem øvre og nedre LOA, giver en visuel vurdering af overensstemmelsen mellem metoderne, hvor en lille afstand indikerer en god overensstemmelse. (72)

LOA beregnes ved formlen:

Limits Of Agreement = mean \pm (SD * 1,96)

I resultatafsnittet præsenteres projektets resultater for henholdsvis evaluering af applikationens segmentering og klassificering samt evalueringen af vurdering af tumorcellekernefraktionen.

5.1 Evaluering af segmentering og klassificering

Den udviklede klassificeringsalgoritme analyserer antallet af sandt positive, sandt negative samt falsk positive og falsk negative for både segmenterings- såvel som klassificeringsafvigelser. Først evalueres applikationens overordnede evne til segmentering og klassificering af tumorcellekerner og normale cellekerner. Efterfølgende evalueres applikationens segmenterings evne og klassificerings evne.

Samlet evaluering		Gold st		
		Positiv	Negativ	l alt
	Positiv	840	357	1197
Аррикаtion	Negativ	1026	706	1732
l alt		1866	1063	2929

Data for det samlede antal positive og negative opstilles i en matrix i tabel 3.

Tabel 3 - Matrix over det samlede antal positive og negative.

Ud fra ovenstående beregnes applikationens overordnede sensitivitet, specificitet samt nøjagtighed for segmenteringen og klassificering af cellekernerne i testsættet. Se tabel 4.

	Sensitivitet	Specificitet	Nøjagtighed
Segmentering	0,450	0,664	0,528

Tabel 4 - Skema over den beregnede sensitivitet, specificitet og nøjagtighed for applikationens overordnede evne til segmentering og klassificering.

I tabel 4 kan det aflæses, at applikationens evne til korrekt at segmentere og klassificere tumorcellekernerne har en sensitivitet på 0,450, hvilket betyder, at 45% af tumorcellekernerne segmenteres og klassificere. Ligeledes ses specificiteten på 0,664, hvilket svarer til, at applikationen korrekt segmenterer og klassificerer 66,4% af de normale cellekerner. Samlet ses en nøjagtighed for applikationens overordnede segmentering og klassificering af både tumorcellekerner og normale cellekerner at være på 52,8%.

5.1.1 Evaluering af segmentering.

Det samlede antal positive og negative, hvor tumor defineres som positiv og normal defineres som negativ, og hvor forkert segmenterede cellekerne defineres som henholdsvis falsk negativ og falsk positiv. Data opstilles i en matrix i tabel 5.

		Gold st		
		Positiv	Negativ	l alt
Applikation	Positiv	840	240	1080
Аррикаціон	Negativ	563	706	1269
l alt		1403	946	2349

Tabel 5 - Matrix over det samlede antal positive og negative i segmenteringen.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for segmenteringen af cellekernerne i testsættet. Se tabel 6.

	Sensitivitet	Specificitet	Nøjagtighed
Segmentering	0,599	0,746	0,656

Tabel 6 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for korrekt segmentering.

I tabel 6 ses det, at applikationens evne til korrekt at segmentere tumorcellekernerne har en sensitivitet på 59,9%. Ligeledes ses specificiteten på 74,6%. Samlet ses en nøjagtighed for applikationens overordnede segmentering af både tumorcellekerner og normale cellekerner at være på 65,6%.

Segmenteringen evalueres ved opdeling af resultaterne i to grupper; materialeudgangspunkt, hvor resultaterne sorteres i primær tumor og metastase, og materialetype, hvor der sorteres i grovnålsbiopsi og excisionsbiopsi. Dette gøres for at vurdere, hvorvidt applikationen kan præstere bedre på en undergruppe frem for en samlet gruppe.

Materialeudgangspunkt

Resultaterne sorteres efter om udgangspunktet for materialet er primær tumor eller metastase. Der er inkluderet 5 primære tumorer og 25 metastaser i testsættet.

Primær		Gold standard		
		Positiv	Negativ	I alt
	Positiv	195	21	216
Аррикаtion	Negativ	97	89	186
I alt		292	110	402

Data fra de primære tumorer opstilles i tabel 7.

 Tabel 7 - Matrix over det samlede antal positive og negative i segmenteringen af primær tumor.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for segmenteringen af cellekernerne i vævssnit fra primær tumor. Se tabel 8.

	Sensitivitet	Specificitet	Nøjagtighed
Segmentering primær	0,668	0,809	0,706

Tabel 8 - Skema over sensitivitet, specificitet og nøjagtighed for applikationens evne til segmentering af primære tumorer.

Derved ses applikationens sensitivitet at være 66,8%, og specificiteten at være 80,9% for korrekt segmentering af henholdsvis tumorcellekerner og normale cellekerner i vævssnit af primære tumor. Yderligere ses en samlet nøjagtighed for korrekt segmentering af både tumorcellekerner og normale cellekerner på 70,6%.

Data fra metastaserne opstilles i tabel 9.

Metastaser		Gold standard		
		Positiv Negativ		I alt
Annlikation	Positiv	645	219	864
Аррикаціон	Negativ	466	617	1083
I alt		1111	836	1947

Tabel 9 - Matrix over det samlede antal positive og negative i segmentering af metastaser.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for segmenteringen af cellekernerne i vævssnit fra metastaser. Se tabel 10.

	Sensitivitet	Specificitet	Nøjagtighed
Segmentering mets	0,581	0,738	0,648

Tabel 10 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for korrekt segmentering af metastaser.

Derved ses applikationens sensitivitet at være 58,1%, og specificiteten at være 73,8% for korrekt segmentering af henholdsvis tumorcellekerner og normale cellekerner i vævssnit af metastaser. Yderligere ses en samlet nøjagtighed for korrekt segmentering af både tumorcellekerner og normale cellekerner på 64,8%.

Materialetype

Resultaterne sorteres efter materialetypen. Der er, i projektet, inkluderet 15 excisionsbiopsier af henholdsvis 6 primære tumor og 9 metastaser, og 15 grovnålsbiopsier kun fra metastaser.

Data for excisionsbiopsierne opstilles i tabel 11.

Excisionsbiopsi		Gold standard		
		Positiv Negativ		l alt
	Positiv	475	108	583
Аррикаціон	Negativ	221	256	477
l alt		696	364	1060

 Tabel 11 - Matrix over det samlede antal positive og negative i segmentering af excisionsbiopsierne.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for segmenteringen af cellekernerne i vævssnit fra excisionsbiopsier. Se tabel 12.

	Sensitivitet	Specificitet	Nøjagtighed
Segmentering	0,682	0,703	0,690
excisionsbiopsi			

Tabel 12 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for korrekt segmentering af excisionsbiopsier.

Derved ses applikationens sensitivitet at være 68,2%, og specificiteten at være 70,3% for korrekt segmentering af henholdsvis tumorcellekerner og normale cellekerner i vævssnit af excisionsbiopsier. Yderligere ses en samlet nøjagtighed for korrekt segmentering af både tumorcellekerner og normale cellekerner på 69%.

Data fra grovnålsbiopsier opstilles i tabel 13.

Grovnålsbiopsi		Gold standard		
		Positiv	Negativ	I alt
Applikation	Positiv	365	132	497
Аррикаціон	Negativ	342	450	792
l alt		707	582	1289

Tabel 13 - Matrix over det samlede antal positive og negative i segmentering af grovnålsbiopsier.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for segmenteringen af cellekernerne i vævssnit af grovnålsbiopsier. Se tabel 14.

	Sensitivitet	Specificitet	Nøjagtighed
Segmentering	0,516	0,773	0,632
grovnålsbiopsi			

Tabel 14 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for korrekt segmentering af grovnålsbiopsier.

Derved ses applikationens sensitivitet at være 51,6%, og specificiteten at være 77,3% for korrekt segmentering af tumorcellekerner og normale cellekerner i vævssnit af grovnålsbiopsier. Yderligere ses en samlet nøjagtighed for korrekt segmentering på 63,2%.

5.1.2 Evaluering af klassificering

Det samlede antal af tumorer og normale cellekerner som applikationen har klassificeret. Data opstilles i en matrix i tabel 15.

		Gold sta		
		Tumor	Normal	l alt
Applikation	Tumor	840	117	957
	Normal	463	706	1169
I alt		1303	823	2126

Tabel 15 - Matrix over det samlede antal positive og negative for applikationens klassificering.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for klassificering af cellekernerne i testsættet. Se tabel 16.

	Sensitivitet	Specificitet	Nøjagtighed
klassificering	0,645	0,858	0,727

Tabel 16 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for klassificering.

Ud fra ovenstående ses det, at applikationens evne til korrekt at klassificere tumorcellekernerne har en sensitivitet på 64,5%. Ligeledes ses specificitet på 74,6%, hvilket betyder at applikationens evne til at klassificere de normale cellekerner korrekt er på 85,8%. Samlet ses en nøjagtighed for applikationens overordnede klassificering af både tumorcellekerner og normale cellekerner at være på 72,7%.

Klassificeringen evalueres ligeledes ved opdeling af resultaterne i de to grupper; materialeudgangspunkt og materialetype.

Materialeudgangspunkt

Resultaterne sorteres efter om udgangspunktet for materialet er primær tumor eller metastase. Der er inkluderet fem primære tumorer og resterende 25 er fra metastaser.

Primære		Gold standard		
		Positiv	Negativ	I alt
Anglihatian	Positiv	195	8	203
Аррикаціон	Negativ	29	89	118
l alt		224	97	321

Data fra de primære tumorer opstilles i en matrix i tabel 17.

 Tabel 17 - Matrix over antal positiv og negativ klassificerede i primære tumorer.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for klassificering af cellekernerne i primære tumorer. Se tabel 18.

	Sensitivitet	Specificitet	Nøjagtighed
Klassificering primære	0,871	0,918	0,885

Tabel 18 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for klassificering af primær tumor.

Derved ses applikationens sensitivitet at være 87,1%, og specificiteten at være 91,8% for korrekt klassificering af tumorcellekerner og normale cellekerner i vævssnit af primære tumorer. Yderligere ses en samlet nøjagtighed for korrekt segmentering på 88,5%.

Data fra metastaser opstilles i en matrix i tabel 19.

Metastaser		Gold standard		
		Positiv	Negativ	I alt
Annelihertiene	Positiv	645	109	754
Аррикацон	Negativ	434	617	1051
l alt		1079	726	1805

Tabel 19 - Matrix over antal positive og negative klassificerede i metastaser.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for klassificering af cellekernerne i metastaser. Se tabel 20.

	Sensitivitet	Specificitet	Nøjagtighed
Klassificering metastaser	0,598	0,850	0,699

Tabel 20 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for klassificering af metastaser.

Derved ses applikationens sensitivitet at være 59,8%, og specificiteten at være 85% for korrekt klassificering af tumorcellekerner og normale cellekerner i vævssnit af metastaser. Yderligere ses en samlet nøjagtighed for korrekt segmentering på 69,9%.

Materialetype

Resultaterne sorteres efter materialetypen. Der er, i projektet, inkluderet 15 excisionsbiopsier af henholdsvis 6 primære tumor og 9 metastaser, og 15 grovnålsbiopsier kun fra metastaser.

Data fra excisionsbiopsier opstilles i en matrix i tabel 21.

Excisionsbiopsi		Gold standard		
		Positiv	Negativ	I alt
	Positiv	475	85	560
Аррикацоп	Negativ	116	256	372
l alt		591	341	932

Tabel 21 - Matrix over positive og negative klassificerede excisionsbiopsier.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for klassificering af cellekernerne i excisionsbiopsierne. Se tabel 22.

	Sensitivitet	Specificitet	Nøjagtighed
Klassificering excisionsbiopsi	0,804	0,751	0,784

Tabel 22 - Skema over applikationens sensitivitet, specificitet og nøjagtighed for klassificering af excisionsbiopsier.

Derved ses applikationens sensitivitet at være 80,4%, og specificiteten at være 75,1% for korrekt klassificering af tumorcellekerner og normale cellekerner i vævssnit af excisionsbiopsier. Yderligere ses en samlet nøjagtighed for korrekt klassificering på 78,4%.

Data fra grovnålsbiopsierne opstilles i en matrix i tabel 23.

Grovnålsbiopsi		Gold standard		
		Positiv	Negativ	I alt
Annlihatian	Positiv	365	32	397
Аррикаціон	Negativ	347	450	797
l alt		712	482	1194

Tabel 23 - Matrix over positive og negative klassificerede grovnålsbiopsier.

Ud fra ovenstående beregnes applikationens sensitivitet, specificitet samt nøjagtighed for klassificering af cellekernerne i excisionsbiopsierne. Se tabel 24.

36	ensitivitet	Specificitet	Nøjagtighed
Klassificering grovnålsbiopsi 0,),513	0,934	0,683

Tabel 24 - Skema over applikationens sensitivitet, specificitet og nøjagtighed i klassificering af grovnålsbiopsier.

Derved ses applikationens sensitivitet at være 51,3%, og specificiteten at være 93,4% for korrekt klassificering af tumorcellekerner og normale cellekerner i vævssnit af grovnålsbiopsier. Yderligere ses en samlet nøjagtighed for korrekt klassificering på 68,3%.

5.2 Evaluering af tumorcellekernefraktionen

Henholdsvis applikationens og *eye ballings* nøjagtighed evalueres med et Bland Altman plot. Forud for Bland Altman plottet testes forskellene for begge overensstemmelser, for normalfordelingen ved en Shapiro-Wilk test.

5.2.1 Applikationens nøjagtighed

Forskellen mellem applikationens vurdering af tumorcellekernefraktionen og *gold standard* findes normalfordelt. Således kan der beregnes både LOA og 95% konfidensinterval for forskellene. Overensstemmelsen mellem applikationens nøjagtighed i beregningen af tumorcellekernefraktionen og *gold standard* illustreres i figur 20.



Bland Altman plot - Applikation

Figur 20 - Bland Altman plot af overensstemmelsen mellem applikationens tumorcellekernefraktion og *gold standard.* Gennemsnittet af forskellen er markeret med en orange linje, hvor 95% konfidensintervallet markeres af grønne stiplede linjer. LOA er markeret med blå linjer.

På figur 20 ses en gennemsnitlig forskel på -40,5%, hvilket indikerer en dårlig overensstemmelse mellem applikationen og *gold standard*. Yderligere ses det, at 95% konfidensintervallet for den gennemsnitlige forskel ligger mellem -29,5% og -51,5%, og dermed indeholder konfidensintervallet ikke 0, hvilket indikerer, at der foreligger en systematiske fejl. Yderligere kan det udledes, at applikationen konsekvent underestimerer tumorcellekernefraktionen. LOA ligger fra 17,4% til -98,4, og svarer til ± 58 i variation fra den gennemsnitlige forskel, hvilket må vurderes at være en bred afstand, og dermed ligeledes indikerer en dårlig overensstemmelse.

Overensstemmelsen evalueres ligeledes ved opdeling af resultaterne i de ovennævnte grupper.

Materialeudgangspunkt

Resultaterne sorteres efter om udgangspunktet for materialet er primær tumor eller metastase. Der er inkluderet fem primære tumorer og resterende 25 er fra metastaser.

Overensstemmelsen mellem applikationens nøjagtighed og *gold standard* for primære tumor ses i figur 21 og overensstemmelsen for metastaserne ses i figur 22.



Bland Altman plot - applikation, primær

Figur 21 – Tumorcellekernefraktion i primære tumorer. Bland Altman plot af overensstemmelsen mellem applikationens tumorcellekernefraktion og gold standard. Gennemsnittet af forskellen er markeret med en orange linje, hvor 95% konfidensintervallet markeres af grønne stiplede linjer. LOA er markeret med blå linjer.

Det ses i figur 21 at den gennemsnitlige forskel mellem applikationen og *gold standard* på primær tumor er på -26,8%, hvilket indikerer en dårlig overensstemmelse. 95% konfidensintervallet for den gennemsnitlige forskel ligger mellem -9,3% og -44,3%, og indeholder dermed ikke 0, hvilket indikerer, at der foreligger systematiske fejl, som tyder på at applikationen, også ved primære tumorer, konsekvent underestimerer tumorcellekernefraktionen. LOA ligger fra 0,8% til -54,5%, og svarer til en variation på ca. ± 28%, hvilket må vurderes at være en bred afstand, og dermed ligeledes indikerer en dårlig overensstemmelse.



Figur 22 - Tumorcellekernefraktion i metastaser. Bland Altman plot af overensstemmelsen mellem applikationens tumorcellekernefraktion og *gold standard*. Gennemsnittet af forskellen er markeret med en orange linje, hvor 95% konfidensintervallet markeres af grønne stiplede linjer. LOA er markeret med blå linjer.

Det ses i figur 22 at den gennemsnitlige forskel mellem applikationen og *gold standard* på metastaser er på -43,2%, hvilket ligeledes indikerer en dårlig overensstemmelse for metastaserne. 95% konfidensintervallet for den gennemsnitlige forskel ligger mellem -30,4% og -56,1%, og indeholder dermed ikke 0, hvilket indikerer, at der findes systematiske fejl, som tyder på, at applikationen konsekvent underestimerer tumorcellekernefraktionen i vævssnit af metastaser. LOA ligger fra 17,9% til -104,4%, og svarer til en variation på ± 61%, hvilket må vurderes at være en bred afstand, og dermed ligeledes indikerer en dårlig overensstemmelse.

Materialetype

Resultaterne sorteres efter materialetypen. Der er, i projektet, inkluderet 15 excisionsbiopsier af henholdsvis 6 primære tumor og 9 metastaser, og 15 grovnålsbiopsier kun fra metastaser.

Overensstemmelsen mellem applikationens nøjagtighed og *gold standard* for excisionsbiopsier ses i figur 23 og overensstemmelsen for grovnålsbiopsierne ses i figur 24.



Bland Altman plot - applikation, resektat

Figur 23 - Tumorcellekernefraktion i excisionsbiopsier. Bland Altman plot af overensstemmelsen mellem applikationens tumorcellekernefraktion og *gold standard*. Gennemsnittet af forskellen er markeret med en orange linje, hvor 95% konfidensintervallet markeres af grønne stiplede linjer. LOA er markeret med blå linjer.

Det ses i figur 23 at den gennemsnitlige forskel mellem applikationen og gold standard i excisionsbiopsier er på -22,7%, hvilket indikerer en dårlig overensstemmelse. 95% konfidensintervallet for den gennemsnitlige forskel ligger mellem -12% og -33,3%, og indeholder dermed ikke 0, hvilket indikerer en systematisk fejl, og at applikationen underestimerer tumorcellekernefraktionen i excisionsbiopsierne. LOA ligger fra 15,1% til - 60,4%, svarende til en variation på ± 38%, hvilket må vurderes at være en bred afstand og dermed ligeledes indikerer en dårlig overensstemmelse.

Bland Altman plot - applikation, biopsi



Figur 24 - Tumorcellekernefraktion i grovnålsbiopsi. Bland Altman plot af overensstemmelsen mellem applikationens tumorcellekernefraktion og *gold standard*. Gennemsnittet af forskellen er markeret med en orange linje, hvor 95% konfidensintervallet markeres af grønne stiplede linjer. LOA er markeret med blå linjer.

Figur 24 viser en gennemsnitlig forskel mellem applikationen og gold standard i grovnålsbiopsier på -58,3%, hvilket indikerer en dårlig overensstemmelse. 95% konfidensintervallet for den gennemsnitlige forskel ligger mellem -45,2% og -74%, og indeholder dermed ikke 0, hvilket indikerer en systematisk fejl, og at applikationen underestimerer tumorcellekernefraktionen i grovnålsbiopsierne. LOA ligger fra -4,6% til - 112,1%, svarende til en variation på ± 54%, hvilket må vurderes at være en bred afstand og dermed ligeledes indikerer en dårlig overensstemmelse.

5.2.2 Eye ballings nøjagtighed

Forskellen mellem *eye balling* og *gold standard* findes normalfordelt, så der beregnes LOA og 95% konfidensinterval for forskellene.

Eye ballings nøjagtighed i vurderingen af tumorcellekernefraktionen evalueres ligeledes med et Bland Altman plot af overensstemmelsen med *gold standard* i figur 25.



Figur 25 - Bland Altman plot af overensstemmelsen mellem *eye balling* tumorcellekernefraktion og *gold standard*. Gennemsnittet af forskellen er markeret med en orange linje, hvor 95% konfidensintervallet markeres af grønne stiplede linjer. LOA er markeret med blå linjer.

På figur 25 ses en gennemsnitlig forskel mellem *eye balling* og *gold standard* på 0,2%, hvilket indikerer en god overensstemmelse. Yderligere ses det at 95%-konfidensintervallet ligger mellem 9% og -8,5%, og dermed indeholder konfidensintervallet 0, hvilket indikerer, at der ikke findes systematiske fejl og at der ikke ses en konsekvent under- eller overestimering. LOA ligger fra 46,2% til -45,7%, svarende til en variation på ± 46%, hvilket må vurderes at være en bred afstand og dermed en mindre god overensstemmelse.

Dette kapitel opdeles i først en diskussion af den anvendte metode til annotering af cellekerner til træning af U-NET. Herefter diskuteres applikationens evne til at kunne segmentere og klassificere de forskellige celler, og afslutningsvis diskuteres vurderingen af tumorcellekernefraktionen for både applikation og *eye balling*.

6.1 Metode til annotering af træningsdata

I dette projekt er der udviklet en metode til annotering af tumorcellekerner i et HE-farvet vævssnit. Dette er gjort for at imødekomme udfordringerne med et begrænset datasæt ved at automatisere annotationerne til træning af netværket samt for at højne kvaliteten af annoteringerne, fordi SOX-10 er specifik for melanocytære celler, og derved kan anvendes som ground truth.

Første datasæt til træningen blev annoteret ud fra immunfarvede vævssnit, hvor der var anvendt et brunt kromogen. Efter verificering af den *handcraftede* algoritmes annotationer ses der dog en del forkerte annoteringer, fordi algoritmen annoterer områder, som indeholder hudens egen pigment. Dette krævede en del justeringer i annotationerne, hvorfor der blev lavet et nyt datasæt til træning, hvor der anvendes et rødt kromogen for at kunne skelne tumorcellekerner fra pigmentet. Det blev dog besluttet at beholde begge træningssæt for at øge omfanget af træningsdata. Derfor kan det diskuteres, om det kunne have styrket netværkets generaliserbarhed, hvis det første træningssæt var blevet ekskluderet, da der er risiko for, at nogle forkerte annoteringer er overset, og derved indgår i træningsmaterialet.

Den *handcraftede* algoritme er opbygget efter *thresholds* for henholdsvis en blå farve for almindelige cellekerner og en brun eller rød farve for tumorcellekerner. Ved annotering af normale cellekerner er *threshold* for den blå farve sat i et niveau, hvor det sikres, at det kun er cellekerner som detekteres, og derved begrænses annotering af baggrundsstøj i farvningen. Dog ses der en variation i farveintensiteten af cellekernerne både imellem forskellige celletyper samt snittene imellem pga. varierende snittykkelse, som normalt ses i histologiske vævssnit, hvorfor nogle normale cellekerner ikke bliver annoteret med algoritmen. Derfor kan det diskuteres om den *handcraftede* algoritme med fordel kunne have et bredere *threshold*, således at der blev annoteret flere normale cellekerner. Der kan dog argumenteres for at bredere *thresholds* vil annotere baggrund og derved en forringelse af kvaliteten. Alternativ kunne der udvælges flere områder til annotering i vævssnittene til træning, således at der annoteres flere både tumorcellekerner og normale cellekerner i træningssættet.

Annoteringen af tumorcellekernerne er lavet ud fra en IHC farvning med SOX-10 som bl.a. er specifik for tumorceller, både i primære maligne melanomer såvel som i metastaser med

udgangspunkt i malignt melanom (63). Ud over tumorceller er SOX-10 ligeledes positiv i normale melanocytter i epidermis samt i forskellige ekkrine kirtler og Schwann celler i de perifere nerver (63), hvorfor områder til annoteringen skal udvælges omhyggeligt for at undgå at normale celler annoteres, og ligeledes at annoteringerne gennemgås og korrigeres for fejl. Dette kan dog være en begrænsning for netværkets generaliserbarhed, fordi de udvalgte områder ikke repræsenterer den hele læsion. Studiet af Bulten et. al. (73) har anvendt en kombination af immunhistokemiske farvninger til identificering af epitelceller og basalceller i glandler i prostata, og derved til adskillelse og annotering af henholdsvis normale glandler og maligne glandler. Studiet har, ligesom i dette projekt, annoteret de forskellige glandler ud fra en handcrafted algoritme med deconvolution og thresholds, og har efterfølgende brugt imellem 45 og 60 minutter per vævssnit på at gennemgå og korrigere annoteringerne (73). Der er ikke taget tid på gennemgang og korrektion af annotationerne i dette projekt, men det er oplevet som et omstændeligt arbejde. På baggrund af dette kan det diskuteres om SOX-10 farvningen med fordel kunne være kombineret med en anden specifik markør for melanom celler, således det kun er ved kombination af de to markører som skal annoteres som tumorceller. Denne kombination vil dog kun undgå Schwann cellerne og de ekkrine kirtler, men ikke normale celler i epidermis, hvorfor det stadig vil være nødvendigt at korrigere annotationerne.

Ud over udfordringen i at normale cellekerner kan annoteres som tumorcellekerner, ses det at lavt differentierede tumorceller kan miste udtrykket af SOX-10. Dette kan give et begrænset repræsentativt datasæt til træning, fordi netværket derved ikke trænes på lavt differentierede celler. Yderligere kan det resultere i, at lavt differentierede tumorceller kan annoteres som normale celler, fordi de ikke er farvet med tumormarkøren. Ud fra ovenstående kan det konkluderes, at denne metode til annotering af forskellige cellekerner ikke kan anvendes uden omhyggelig verificering af, at der er annoteret korrekt. Dog viser erfaringen at dette er en mindre krævende opgave end den manuelle annotering, hvorfor den handcraftede algoritme stadig vil foretrækkes til opgaven. Studiet af Bulten et. al. (73) har dog anvendt annotationerne fra deres handcraftede algoritme til træning af et neuralt netværk på yderligere IHC farvede vævssnit med formålet at annotere endnu flere celler, som efterfølgende kan overføres til det korresponderende HE-farvede vævssnit. Det kan derfor være relevant for lignende fremtidige projekter at optimere annoteringerne ved at kombinerede annoteringsmetoden med en DL algoritme. Alternativt har studiet af Jackson, Sriharan og Vaickus (74) anvendt en anden automatiseret metode til annotering af melanocytære celler ved først at træne et neuralt netværk til segmentering af alle kerner i HE-farvede vævssnit, for efterfølgende at lable de segmenterede kerner ud fra en SOX-10 farvning på et korresponderende vævssnit. Derved har de annoteret samtlige kerner i det digitaliserede vævssnit med enten en tumorlabel eller en normal label. Der er med denne metode annoteret henholdsvis i omegnen af 306.000 tumorcellekerner og 3.396.000 normale cellekerner. I dette projekt er der udvalgt mindre områder i læsionerne til annoteringen, for at begrænse omfanget af normale SOX-10 positive celler, hvor studiet af Jackson, Sriharan og Vaickus har annoteret hele læsionen i de forskellige vævssnit. Dette indikerer, at det er muligt med denne eller andre metoder at annotere mange cellekerner til træning, dog skal der stadig tages højde for at anvendelsen af SOX-10 som ground truth kræver omfattende tilretning.

6.2 Applikationens evne til segmentering og klassificering

Evalueringen af applikationens evne til segmentering og klassificering er foretaget ud fra en *gold standard*. Denne *gold standard* er det korresponderede SOX-10 farvede vævssnit, som den *handcraftede* algoritme har lablet med henholdsvis tumorcellekerne label og normal cellekerne label. Udgangspunktet for *gold standard* er manulet tilrettet, fordi ikke alle cellerne i vævssnittene er lablet, jf. figur 15. Dog er det i forbindelse med dette projekt identificeret mange SOX-10 negative tumorceller, hvilket kan betyde at nogle tumorceller er karakteriseret som normale celler, til trods for den manuelle tilretning.

Applikationens overordnede evne til segmentering og klassificering belyses ved hjælp af den beregnede sensitivitet, specificitet og nøjagtighed. Sensitiviteten beregnes til 45%, hvilket viser testens evne til korrekt at segmentere og klassificere de sandt positive, som i dette tilfælde er tumorcellekerner. Omvendt er testes evnen til korrekt at segmentere og klassificere de sandt negative, normale cellekerner vist ved specificiteten, som er udregnet til 66,4%. Dette betyder, at applikationen identificerer 45% af tumorcellekerner korrekt og 66,4% af de normale cellekerner korrekt. Hermed ses det, at applikationen præsterer bedst i identificeringen af normale cellekerner, omend identificeringen er begrænset. Samlet ses applikationens nøjagtighed på 52,8% for korrekt identificering uanset klassen. Derved må applikationens overordnede evne til segmentering og klassificering af cellekernerne, i dette projekt, betragtes som utilstrækkelige til trods for at der ikke, forud for projektet, er besluttet et antal af fejl, som accepteres.

Ved fokusering på applikationens evne til korrekt segmentering af cellekernerne beregnes sensitiviteten, specificiteten og nøjagtigheden for denne, hvor tumorcellekernerne defineres som de positive, og de normale cellekerner defineres som de negative. De falske negative og positive defineres som de cellekerner applikationen ikke har segmenteret. Her er applikationens sensitivitet for segmenteringen beregnet til 59,9%. Dvs. at testen evner korrekt at segmentere 59,9% af tumorcellerne. Applikationens specificitet er beregnet til 74,6%, hvilket svarer til at der korrekt segmenteres 74,6% af de normale cellekerner. Samlet ses applikationen at have en nøjagtighed på 65,6% for en overordnet korrekt segmentering, hvilket må betragtes som utilstrækkelig. Ved manuel inspektion af applikationens resultater ses det, at applikationen i flere tilfælde, ikke segmenterer overlappende kerner, hvormed der kun tælles én kerne i stedet for flere kerner. Studiet at Jørgensen et. al. (75) belyser udfordringen i den individuelle cellekernesegmentering, og beskriver forskellige tiltag som eventuelt kan anvendes i postprocessering. Dette kan f.eks. være at tilføje en detektering af konkave hjørner, som kan indikere en adskillelse af tætliggende celler. (75)

Applikationens evne til korrekt klassificering af cellekernerne undersøges ligeledes ved beregning af sensitiviteten, specificiteten og nøjagtigheden. Her defineres de falske som værende de klassificeringer, applikationen har gjort forkert, ligeledes indeholder denne gruppe applikationens klassificereringer, hvor der ikke er en cellekerne. Applikationens sensitivitet for klassificering er 64,5% og specificiteten er 85,8%, yderligere ses nøjagtigheden

at være 72,7%. Hvilket, ligesom med segmenteringen, må betragtes at være utilstrækkelig. Dog kan det diskuteres, at applikationen præsterer en smule bedre i klassificeringen end i segmenteringen. Studiet af Wang et. al. (76) har trænet et CNN til identificering af forskellige celler i hepatocellulære carcinomer, og finder en samlet nøjagtighed i identificeringen af alle celler uanset klasse på 92%. Hvilket er nærmere et acceptabelt niveau, sammenlignet med dette projekts resultater. Studie af Wang et.al. har anvendt i alt 65.000 annotationer, hvilket er mindre end det antal som er anvendt i dette projekt. Dog er der i studiet kun anvendt vævssnit fra den primære tumor og yderligere kun begrænset materiale fra metastaserende tumorer (76), hvorfor cellekernerne må formodes at være højt differentieret, således tumorcellerne ligner mere deres egentlige udgangspunkt, i modsætning til materialet inkluderet i dette projekt. Studiet af Jackson, Sriharan og Vaickus (74) har trænet et CNN til identificering af melanocytære celler, i både primære tumorer og metastaser, og studiet beskriver en udfordring i, at have samlet træningsmateriale på tværs af forskellige vævstyper. Her italesættes at diversiteten i de forskellige vævstyper, kan have øget den overordnede generaliserbarhed, men på bekostning af nøjagtigheden. Derfor kan det være relevant at opdele træningen vævsspecifikt, hvilket dette projekts resultater ligeledes indikerer.

For at teste hvorvidt applikationen præsterer bedre på et mere målrettet datasæt, opdeles applikationens resultater i grupperne; materialeudgangspunkt og materialetype. Dette gøres for at kunne vurdere hvorvidt træningsdata med fordel, kan målrettes enten et bestemt materialeudgangspunkt eller materialetype. Dog er der i undergruppen materialeudgangspunkt kun inkluderet 5 primære tumorer, i modsætning til 25 metastaser, hvilket ikke giver det bedste grundlag for sammenligning, men er klinisk repræsentativt for materialet til BRAF-analyse.

Ved opdeling af resultaterne i materialeudgangspunkt er den samlede evaluering, at applikationen overordnet præsterer bedst på primær tumor, dog gældende for både primære og metastaser ses specificiteten at være bedre end sensitiviteten. Dette kan skyldes at tumorcellekernerne i metastaserne, ofte er meget lavt differentierede, fordi de kan efterligne de oprindelige celler i det metastatiske organ (28). Et tidligere studie har beskrevet en omfattende variation i melanomcellernes morfologi også indenfor den enkelte tumor, hvilket kan vanskeliggøre den histopatologiske diagnose, selv for en trænet speciallæge i patologi (28). Den store variation i cellekernernes morfologi, vil kræve et stort, og nøje udvalgt, træningssæt for at sikre, at netværket kan lære tilstrækkeligt. Indledningsvis var metastaserne i træningssættet udvalgt tilfældigt uden hensyntagen til en repræsentativ udvælgelse. I løbet af træningen identificeres udfordringen i ikke tilstrækkeligt repræsentativt materiale, fordi der primært var udvalgt store subkutane metastaser og lymfeknudemetastaser, hvorfor der efterfølgende prioriteres at inkludere specifikt metastatisk materiale fra fem grovnålsbiopsier. Opdeles resultaterne i materialetype ses det, at applikationen præsterer bedst på excisionsbiopsierne. Dette kan skyldes at grovnålsbiopsiser indeholder begrænset materiale, fordi der udtages en vævsprøve på ca. 1 mm i diameter og 1 – 2 cm i længden. Derfor er der risiko for et mindre repræsentativt udsnit ved en tumor med øget morfologisk variation. (77) Dog er det vist, at anvendelse af grovnålsbiopsier til diagnostik er fuldt ud tilstrækkeligt som diagnostisk materiale (78), hvorfor forskellen imellem disse grupper sikkert skyldes at den primære tumor altid er en excisionsbiopsi, og derfor er inkluderet i denne gruppe.

Ovenstående giver anledning til, at fremtidige studier kan fokusere på at målrette træningen af netværket til enten primære tumorer eller metastaser, og såfremt det skal målrettes metastaserne, skal det sikres, at træningssættet er udvalgt repræsentativt for de forskellige vævstyper, såvel som materialetyper.

6.3 Vurdering af tumorcellekernefraktionen

Evalueringen af applikationens evne til estimering af tumorcellefraktionen er gjort ud fra overensstemmelsen med en manuel celletælling som *gold standard*. Den manuelle celletælling er udført på det korresponderende SOX-10 farvede vævssnit, som tidligere er beskrevet som mangelfuld. Dette kan have medført, at nogle tumorceller er talt som normale celler, selvom tællingerne er foretaget af en erfaren patolog.

Overensstemmelsen mellem tumorcellekernefraktionen for applikationen og *gold standard* vurderes utilstrækkelig. Den gennemsnitlige forskel mellem applikationen og *gold standard* er -40,5% og ved LOA ses en variation i forskellen mellem applikationen og *gold standard* på ca. ± 58%. Dette kan skyldes at applikationen, i dette projekt, er trænet på forskellige materialetyper samt forskellige vævstyper for at sikre et klinisk repræsentativt datasæt til træningen. Dog tyder resultaterne på, at netværket skal trænes yderligere for at opnå en tilfredsstillende nøjagtighed.

Ved opdeling af resultaterne vurderes det for gruppen materialeudgangspunkt, at applikationen præsterer bedre på primære tumorer, fordi den gennemsnitlige forskel er mindre ved primær tumor i modsætning til ved metastaser; henholdsvis -26,8% og -43,24%. Dog er overensstemmelsen for ingen af grupperne tilstrækkelig til, at applikationen kan anvendes. Ligeledes ses det, for begge grupper, at applikationen konsekvent underestimerer tumorcellekernefraktionen. LOA vurderes at være bred for begge undergrupper, dog spænder LOA for de primære kun over ca. \pm 28%, hvor LOA spænder over ca. \pm 61% for metastaserne. Opdeles resultaterne i materialetype, er den samlede evaluering, at applikationen præsterer bedre på excisionsbiopsier, fordi den gennemsnitlige forskel er tættere på 0 i modsætning til ved grovnålsbiopsierne; henholdsvis -22,7% og -58,3%. Dog er overensstemmelsen for ingen af grupperne tilstrækkelig til applikationen kan anvendes. Ligeledes ses det ved begge grupper, at applikationen konsekvent underestimerer tumorcellekernefraktionen. LOA ses ved begge at være bred, dog spænder LOA for excisionsbiopsierne kun over ca. ± 38%, hvor LOA spænder over ca. ± 54% for grovnålsbiopsierne. Dette indikerer ligeledes, at overensstemmelsen er bedre ved excisionsbiopsierne end ved grovnålsbiopsierne. Dette kan igen skyldes, at gruppen af excisionsbiopsierne indeholder de primære tumorer, hvor gruppen af grovnålsbiopsier kun omfatter metastaser. Ovenstående understøtter, at applikationen med formål kan målrettes materialeudgangspunktet, hvormed træning kan gøres mere repræsentativ.

Sammenlignes resultaterne fra applikations overordnede overensstemmelse og *eye ballings* overensstemmelse ses det at *eye balling* præsterer bedre i vurderingen af tumorcellekernefraktionen. Den gennemsnitlige forskel for *eye balling* er tæt på 0 (0,23%),

hvor applikationens gennemsnitlige forskel er på -40,5%. Dog ses LOA for begge at spænde bredt, henholdsvis ca. \pm 46% for eye balling og ca. \pm 58% for applikationen. Det er dog intet generelt mål for hvor snævert LOA skal være, før dette er acceptabelt, men at det skal afgøres efter de biologiske eller kliniske omstændigheder, hvor metoden skal anvendes (72). Eftersom sensitiviteten af BRAF-analysen er afhængig af en vis mængde tumorceller i forhold til normale celler, vil en usikkerhed på ± 46%, for eye balling, kompromittere BRAF-analysen uhensigtsmæssigt. Ligeledes vil usikkerheden i tumorcellekernefraktionen kompromittere tolkningen af BRAF-analysen. (36,77) Studiet af Lhermitte et.al. (36) italesætter, at problemet med en upræcis tumorcellekernefraktion er størst i tilfælde, hvor denne er overestimeret, netop i forhold til sensitiviteten af analysen, hvorfor tumorcellekernefraktionen bør revurderes ved en negativ BRAF. Omvendt vil en underestimeret tumorcellekernefraktion, som ligger i nærheden af analysens grænseværdi resultere i, at en negativ BRAF eventuelt gentages, og besvares som inkonklusiv. Ligeledes siger studiet, at det er vigtigt med en præcis tumorcellekernefraktion til tolkning af den positive analyses resultater, således at det er muligt at vurdere heterogeniteten af BRAF-mutationen i tumoren. (36) En korrekt vurdering af tumorcellekernefraktionen er, som nævnt, essentielt for BRAF-analysens sensitivitet, og kan dermed være afgørende for, om den enkelte patient tilbydes den rette behandling. Endvidere er det oftest eneste mulighed for behandling til denne gruppe af patienter, fordi der ikke tilbydes andre effektive behandlinger (43), og det er derfor essentielt for den enkelte patients forlængede overlevelse.

I ovenstående er det belyst at applikationen, for nuværende, ikke præsterer tilstrækkeligt, dog er det ligeledes vist, at *eye balling* heller ikke er præcis nok til at sikre hverken BRAF-analysens sensitivitet eller tolkning af analysen. Derfor kan en applikation udviklet ved træning af U-NET stadig være et bedre redskab til nøjagtig vurdering af tumorcellekernefraktionen. Dog vil det kræve yderligere træning af netværket samt en eventuel mere fokuseret træning på henholdsvis primær tumor eller metastaser.

Konklusion 7

Formålet med dette projekt var at udvikle en applikation til bestemmelse af tumorcellekernefraktionen, i vævssnit af malignt melanom og med udgangspunkt i følgende opstilledes problemformulering:

Kunstig intelligens kan anvendes til udvikling af en applikation, som kan estimere tumorcellekernefraktionen i et digitaliseret HE-farvet vævssnit fra både primær tumor og metastaser, såvel som fra excisionsbiopsier og grovnålsbiopsier, med større nøjagtighed end den subjektive vurdering, som foretages i forbindelse med diagnostikken.

I projektet udvikles applikationen ved træning af U-NET til segmentering og klassificering af de forskellige cellekerner i vævssnittene. Annotering af cellekernerne til træning af U-NET er gjort ved hjælp af en *handcrafted* algoritme, således det både var nemmere at annotere flere cellekerner samt at sikre kvaliteten af annoteringerne. Resultaterne viser dog, at algoritmens annotationer, skal korrigeres i et stort omfang, både for at sikre kvaliteten af annotationerne såvel som at sikre, at de er tilstrækkeligt repræsentative for den variation, som ses i vævssnit. Derfor er der i dette projekt kun annoteret mindre områder af læsionerne, for at begrænse det omfattende arbejde. Her viser litteraturen at alternative metoder, i anvendelsen af en *handcrafted* algoritme, med fordel kan anvendes til annoteringer af flere cellekerner.

Den udviklede applikation har vist sig, at have vanskeligt ved både komplet segmentering af de enkelte cellekerner samt korrekt klassificering tumorcellekernerne. Kombinationen af ikke tilstrækkeligt segmenterede cellekerner, såvel som forkert klassificering kommer til udtryk i, at applikationen konsekvent underestimerer tumorcellekernefraktionen. Dog ses det at applikationen præsterer en anelse bedre på de primære tumorer end på metastaserne, såvel som på excisionsbiopserne frem for grovnålsbiopsierne. Derfor er konklusionen, i dette projekt, at applikationen udviklet med kunstig intelligens, ikke for nuværende, kan estimere en nøjagtig tumorcellekernefraktion i et digitaliseret HE-farvet vævssnit fra både primær tumor og metastaser såvel som fra excisionsbiopsier og grovnålsbiopsier. Dog viser projektets resultater ligeledes at selvom den subjektive vurdering, som anvendes i diagnostikken heller ikke er tilstrækkelig nøjagtig i denne vurdering, er den dog stadig mere nøjagtig end applikationen. Litteraturen viser dog, at anvendelsen af kunstig intelligens kan præstere bedre end den subjektive vurdering, såfremt det anvendte netværk er trænet tilstrækkeligt. Her viser projektets resultater, at træningen af netværket med fordel kan målrettes udgangspunktet for materialet, således den biologiske og morfologiske variation imellem vævstyperne begrænses i træningen. Det skal dog sikres at der ved træningen af metastatisk væv, skal inkluderes et repræsentativt omfang af de forskellige metastatiske lokalisationer, lige som de forskellige materialetyper skal repræsenteres. Applikationen skal efterfølgende optimeres i postprocesseringen til bedre segmentering af de individuelle cellekerner.

Perspektivering 8

Dette projekt bidrager med et videnskabelige grundlag, som kan anvendes i fremtidige studier med samme formål. Såfremt applikationen havde præsteret optimalt, kunne det ligeledes bidrage med den videnskabelige dokumentation, der er nødvendig for en fremtidig implementering som et diagnostisk redskab.

Projektet identificerer en række udfordringer i træningen af et neuralt netværk, herunder både udvælgelse af træningsmateriale, samt annoteringen af cellekernerne. I forhold til udvælgelse af træningsmateriale, er det som tidligere nævnt relevant at målrette træningen til bestemte vævstyper. Eftersom malignt melanom metastaserer til mange forskellige organer, f.eks. subkutan fedt- og bindevæv, lever, hjerne og lymfekirtler, er det ikke optimalt at målrette netværket til vævstyper. Her kan det være relevant at målrette et netværk til den primære tumor og et andet til metastaser generelt, hvor det sikres at de forskellige vævstyper er tilstrækkeligt repræsenteret i træningssættet. Alternativt kunne det belyses, hvorvidt BRAF mutationen er tilstrækkelig homogen til at der som udgangspunkt kunne analyseres BRAF på den primære tumor, til trods for at det ofte er metastasen, som behandles med BRAF hæmmere.

Annoteringen af cellekernerne kan med fordel annoteres med den anvendte metode, dog kunne denne modificeres til annotering af endnu flere cellekerner, hvor litteraturen opstiller forskellige relevante eksempler. F.eks. kan annoteringer lavet med en *handcrafted* algoritme trænes yderligere i et neuralt netværk, med formålet at nemmere kunne annotere et repræsentativt omfang af cellekerner til træningen af den endelige algoritme.

Et fremtidigt studie kan sammenligne den estimerede tumorcellekernefraktion, med den fundne allelfrekvens ved BRAF analysen. Her ville det forventes ved en homogen mutation af identificere et tilsvarende antal allelfrekvenser, og vil derved kunne bruges som et mål for estimeringen. Studiet af Dudley et. al. (77) viser at allelfrekvensen og tumorcellekernefraktionen er sammenhængende og anvender omvendt tumorcellekernefraktionen som kvalitetsmål for en mindre sensitiv BRAF-analyse. Dog vil et sådant studie skulle tage højde for en nøjagtig makrodissektion, hvor ROI skæres ud, til BRAFanalysen, for at sikre at der kun makrodissekeres det ROI, hvor tumorcellekernefraktionen ligeledes vurderes fra. Makrodissektionen har indflydelse på antallet af de forskellige celler, som analyseres for BRAF og derved afspejle sig i allelfrekvensen.

- 1. Kim B, Kang SY, Kim K-M. DNA-protein biomarkers for immunotherapy in the era of precision oncology. J Pathol Transl Med. 9. november 2020;
- Jakobsen AS, Kjems E. Behandling ved spredning af modermærkekræft (metastaser) [Internet]. Kræftens Bekæmpelse. 2019 [henvist 11. januar 2021]. Tilgængelig hos: https://www.cancer.dk/modermaerkekraeft-malignt-melanom/behandlingmodermaerkekraeft/behandling-ved-spredning/
- Dean L. Vemurafenib Therapy and BRAF and NRAS Genotype. I: Pratt VM, McLeod HL, Rubinstein WS, Scott SA, Dean LC, Kattman BL, m.fl., redaktører. Medical Genetics Summaries [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2012 [henvist 11. januar 2021]. Tilgængelig hos: http://www.ncbi.nlm.nih.gov/books/NBK447416/
- 4. Cheng L, Lopez-Beltran A, Massari F, MacLennan GT, Montironi R. Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine. Mod Pathol. januar 2018;31(1):24–38.
- 5. Smits AJJ, Kummer JA, de Bruin PC, Bol M, van den Tweel JG, Seldenrijk KA, m.fl. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. Mod Pathol. februar 2014;27(2):168–74.
- 6. Peikari M, Salama S, Nofech-Mozes S, Martel AL. Automatic cellularity assessment from post-treated breast surgical specimens: Automatic Cellularity Assessment. Cytometry A. november 2017;91(11):1078–87.
- Globocan 2018. All cancers fact sheet [Internet]. WHO; 2019 [henvist 11. januar 2021]. Tilgængelig hos: https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-factsheet.pdf
- 8. American Cancer Society, Inc. The Burden of Cancer [Internet]. The Cancer Atlas. 2019 [henvist 11. januar 2021]. Tilgængelig hos: http://canceratlas.cancer.org/Xoh
- 9. Globocan 2018. Denmark fact sheet, 2018 [Internet]. 2020 [henvist 11. januar 2021]. Tilgængelig hos: https://gco.iarc.fr/today/data/factsheets/populations/208-denmarkfact-sheets.pdf
- Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, m.fl. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 15. april 2019;144(8):1941–53.
- 11. Stewart BW, Wild C, International Agency for Research on Cancer, World Health Organization. World cancer report 2014. 11/01/2021; 2014.

- 12. Jönsson B, Hofmarcher T, Lindgren P, Wilking N. The cost and burden of cancer in the European Union 1995–2014. Eur J Cancer. oktober 2016;66:162–70.
- 13. Globocan 2018. Melanoma of the skin fact sheet [Internet]. 2019 [henvist 11. januar 2021]. Tilgængelig hos: https://gco.iarc.fr/today/data/factsheets/cancers/16-Melanoma-of-skin-fact-sheet.pdf
- 14. Medicinerrådet. Baggrund for Medicinrådets behandlingsvejledning vedrørende lægemidler til adjuverende behandling af modermærkekræft [Internet]. 2020 [henvist 19. marts 2021]. Tilgængelig hos: https://medicinraadet.dk/media/hjymd4lv/baggrund-formedicinraadets-behvejl-vedr-modermaerkekraeft-vers-10_adlegacy.pdf
- Mogensen M, Jensen AN, Karlsmark T, Sachs C. Malignt melanom Lægehåndbogen på sundhed.dk [Internet]. 2019 [henvist 11. januar 2021]. Tilgængelig hos: https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/hud/tilstande-ogsygdomme/pigmenterede-laesioner/malignt-melanom/
- Cancer.Net Editorial Board. Melanoma Statistics [Internet]. Cancer.Net. 2020 [henvist 31. januar 2021]. Tilgængelig hos: https://www.cancer.net/cancertypes/melanoma/statistics
- Ellebæk E, Øllegaard TH, Højberg L. Melanoma Guidelines inoperabel metastatisk melanom [Internet]. DMG; 2019 [henvist 17. januar 2021]. Tilgængelig hos: https://www.melanoma.dk/assets/files/Melanoma_Guidelines_inoperabel-metastatiskmelanom.pdf
- 18. Marcussen N. Patologi. Kbh.: Roskilde; 2019.
- 19. Garibyan L, Fisher DE. How Sunlight Causes Melanoma. Curr Oncol Rep. september 2010;12(5):319–26.
- 20. Markovic SN, Erickson LA, Rao RD, McWilliams RR, Kottschade LA, Creagan ET, m.fl. Malignant Melanoma in the 21st Century, Part 1: Epidemiology, Risk Factors, Screening, Prevention, and Diagnosis. Mayo Clin Proc. marts 2007;82(3):364–80.
- 21. Maul M, Dinesen MØ, Rasmussen LM, Brixen K. Pakkeforløb for modermærkekræft i huden [Internet]. Sundhedsstyrelsen; 2020 [henvist 11. januar 2021]. Tilgængelig hos: https://www.sst.dk/da/udgivelser/2020/pakkeforloeb-for-modermaerkekraeft
- 22. Pakkeforloeb-og-opfoelgningsprogrammer-Begreber-forloebstider-og-monitorering.pdf [Internet]. [henvist 17. december 2020]. Tilgængelig hos: https://www.sst.dk/-/media/Udgivelser/2019/Pakkeforloeb-kraeft-2015-2019/Pakkeforloeb-ogopfoelgningsprogrammer-Begreber-forloebstider-og-monitorering.ashx
- 23. Sundhedsstyrelsen. Pakkeforløb for modermærke­kræft i huden [Internet]. 2020
[henvist 11. januar 2021]. Tilgængelig hos:
https://www.sst.dk/da/udgivelser/2020/pakkeforloeb-for-modermaerkekraeft

- 24. DMCG.dk. DMCG [Internet]. [henvist 11. januar 2021]. Tilgængelig hos: https://www.dmcg.dk/
- 25. DMG. Malign melanom for klinikere [Internet]. [henvist 11. januar 2021]. Tilgængelig hos: https://www.melanoma.dk/page3.html#header1-10
- 26. Wagenblast AL. Biopsi ved mistanke om primært melanom [Internet]. DMCG; 2020 [henvist 17. januar 2021]. Tilgængelig hos: https://www.melanoma.dk/assets/files/Melanoma_Guidelines_3.pdf
- 27. Klausen S, Spaun E. Melanoma Guidelines patologi [Internet]. DMG; 2018 [henvist 17. januar 2021]. Tilgængelig hos: https://www.melanoma.dk/assets/files/Melanoma_Guidelines_Patologi.pdf
- 28. Banerjee, Harris. Morphological and immunophenotypic variations in malignant melanoma: Variations in malignant melanoma. Histopathology. maj 2000;36(5):387–402.
- Hölmich LR. Melanoma Guidelines Tumor klassifikation i henhold til UICCs/AJCCs TNMklassifikation, 8. Udgave [Internet]. DMG; 2017 [henvist 17. januar 2021]. Tilgængelig hos: https://www.melanoma.dk/assets/files/Melanoma_Guidelines_TNM%20klassifikation.p df
- 30. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. marts 2011;144(5):646–74.
- 31. Hupé P. Hallmarks of cancer [Internet]. 2011 [henvist 21. februar 2021]. Tilgængelig hos: https://commons.wikimedia.org/wiki/File:Hallmarks_of_cancer.svg
- 32. Nielsen LB, Dabrosin N, Sloth K, Bønnelykke-Behrndtz ML, Steiniche T, Lade-Keller J. Concordance in BRAF V600E status over time in malignant melanoma and corresponding metastases. Histopathology. april 2018;72(5):814–25.
- Fagudvalget for metastaserende malignt melanom. RADS baggrundsnotat malignt melanom [Internet]. 2016 [henvist 2. marts 2021]. Tilgængelig hos: https://rads.dk/media/4244/baggrundsnotat-malignt-melanom.pdf
- 34. Yancovitz M, Litterman A, Yoon J, Ng E, Shapiro RL, Berman RS, m.fl. Intra- and Inter-Tumor Heterogeneity of BRAFV600EMutations in Primary and Metastatic Melanoma. Brandner JM, redaktør. PLoS ONE. 3. januar 2012;7(1):e29336.
- 35. Kristensen TK, Høgdall E, Christensen M, Hansen Bonde J, Pallisgaard N, Spaun E, m.fl. Detektion af V600E-muteret BRAF ved fremskreden (ikke-resekterbar eller metastaserende) malignt melanom – anbefalinger til metodik. [Internet]. DPAS; 2015 [henvist 31. januar 2021]. Tilgængelig hos: https://danskpatologi.org/wpcontent/uploads/2016/02/anbefalinger_molekyaerbiologisk_analyse_for-BRAF_-febr-2015.pdf
- 36. Lhermitte B, Egele C, Weingertner N, Ambrosetti D, Dadone B, Kubiniek V, m.fl. Adequately defining tumor cell proportion in tissue samples for molecular testing

improves interobserver reproducibility of its assessment. Virchows Arch. 16. november 2016;470(1):21–7.

- Jakobsen AS, Kjems E. Behandling af modermærkekræft [Internet]. Kræftens Bekæmpelse.
 2019 [henvist 31. januar 2021]. Tilgængelig hos: https://www.cancer.dk/modermaerkekraeft-malignt-melanom/behandlingmodermaerkekraeft/
- 38. Drzewiecki KT, Gjørup C. Melanoma Guidelines behandling af primær tumor [Internet].
 2013 [henvist 17. januar 2021]. Tilgængelig hos: https://www.melanoma.dk/assets/files/Melanoma_Guidelines_7.pdf
- 39. Larsen MS, Keller JL, Klausen S. Melanoma Guidelines lymfeknude patologi [Internet].
 2020 [henvist 17. januar 2021]. Tilgængelig hos: https://www.melanoma.dk/assets/files/Melanoma_Guidelines_4.pdf
- 40. Haldrup M, Stolle LB, Hölmich LR, Chakera A. Sentinel node-positive melanoma. Ugeskr Laeger. 8. oktober 2018;180(41).
- 41. Ammundsen IN, Kjems E. Immunterapi [Internet]. Kræftens Bekæmpelse. 2019 [henvist 1. februar 2021]. Tilgængelig hos: https://www.cancer.dk/hjaelpviden/kraeftbehandling/behandlingsformer/immunterapi/
- 42. Ny L, Hernberg M, Nyakas M, Koivunen J, Oddershede L, Yoon M, m.fl. BRAF mutational status as a prognostic marker for survival in malignant melanoma: a systematic review and meta-analysis. Acta Oncol. 2. juli 2020;59(7):833–44.
- Hauschild A, Dummer R, Schadendorf D, Santinami M, Atkinson V, Mandalà M, m.fl. Longer Follow-Up Confirms Relapse-Free Survival Benefit With Adjuvant Dabrafenib Plus Trametinib in Patients With Resected *BRAF* V600–Mutant Stage III Melanoma. J Clin Oncol. 10. december 2018;36(35):3441–9.
- 44. Schadendorf D, Hauschild A, Santinami M, Atkinson V, Mandalà M, Chiarion-Sileni V, m.fl. Patient-reported outcomes in patients with resected, high-risk melanoma with BRAFV600E or BRAFV600K mutations treated with adjuvant dabrafenib plus trametinib (COMBI-AD): a randomised, placebo-controlled, phase 3 trial. Lancet Oncol. maj 2019;20(5):701–10.
- 45. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology new tools for diagnosis and precision oncology. Nat Rev Clin Oncol. november 2019;16(11):703–15.
- 46. Hanna MG, Pantanowitz L. Digital Pathology. I: Narayan R, redaktør. Encyclopedia of Biomedical Engineering [Internet]. Oxford: Elsevier; 2019 [henvist 11. marts 2021]. s. 524–32. Tilgængelig hos: https://www.sciencedirect.com/science/article/pii/B9780128012383999586
- 47. Moxley-Wyles B, Colling R, Verrill C. Artificial intelligence in pathology: an overview. Diagn Histopathol. november 2020;26(11):513–20.

- 48. Grunkin M, Raundahl J, Foged NT. Practical Considerations of Image Analysis and Quantification of Signal Transduction IHC Staining. I: Signal Transduction Immunohistochemistry: Methods and Protocols. 2011. s. 143–54. (Springer protocols).
- 49. Nielsen PS, Spaun E, Riber-Hansen R, Steiniche T. Automated quantification of MART1verified Ki-67 indices: useful diagnostic aid in melanocytic lesions. Hum Pathol. juni 2014;45(6):1153–61.
- 50. Nielsen M. Machine learning automatisk læring med computere. I: Den digitale revolution : fortællinger fra datalogiens verden. 2010.
- 51. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, CM-Path AI in Histopathology Working Group, m.fl. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. J Pathol. oktober 2019;249(2):143–50.
- 52. Lægemiddelstyrelsen. CE-mærkning [Internet]. Lægemiddelstyrelsen. 2015 [henvist 28. februar 2021]. Tilgængelig hos: https://laegemiddelstyrelsen.dk/da/udstyr/ce-maerkning/
- 53. Visiopharm A/S. Visiopharm APP Center APPsolute Image Analysis [Internet]. Visiopharm
 Pathology image analysis software. [henvist 28. februar 2021]. Tilgængelig hos: https://visiopharm.com/app-center/
- 54. Visiopharm A/S. Visiopharm Software Features Powerful image analysis [Internet]. Visiopharm · Pathology image analysis software. [henvist 22. marts 2021]. Tilgængelig hos: https://visiopharm.com/visiopharm-digital-image-analysis-software-features/
- 55. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. maj 2015;521(7553):436–44.
- 56. Wani MA, Bhat FA, Afzal S, Khan AI. Advances in Deep Learning [Internet]. Singapore: Springer Singapore; 2020 [henvist 7. marts 2021]. (Studies in Big Data; bd. 57). Tilgængelig hos: http://link.springer.com/10.1007/978-981-13-6794-6
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. I: Navab N, Hornegger J, Wells WM, Frangi AF, redaktører. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing; 2015. s. 234–41.
- Bengio Y. Practical recommendations for gradient-based training of deep architectures. ArXiv12065533 Cs [Internet]. 16. september 2012 [henvist 3. maj 2021]; Tilgængelig hos: http://arxiv.org/abs/1206.5533
- 59. Visiopharm A/S. AI/DeepLearning Deep Learning Section from VIS Help Manual. 2020.
- 60. Yazdani M. Example architecture of U-Net for producing k 256-by-256 image masks for a 256-by-256 RGB image [Internet]. 2019 [henvist 1. april 2021]. Tilgængelig hos: https://commons.wikimedia.org/wiki/File:Example_architecture_of_U-Net_for_producing_k_256-by-256_image_masks_for_a_256-by-256_RGB_image.png

- 61. Ibrahim A, Gamble P, Jaroensri R, Abdelsamea MM, Mermel CH, Chen P-HC, m.fl. Artificial intelligence in digital breast pathology: Techniques and applications. Breast Edinb Scotl. februar 2020;49:267–73.
- 62. Tizhoosh HR, Pantanowitz L. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. J Pathol Inform. 2018;9:38.
- 63. Nonaka D, Chiriboga L, Rubin BP. Sox10: A Pan-Schwannian and Melanocytic Marker. Am J Surg Pathol. september 2008;32(9):1291–8.
- 64. NVK. Hvad skal jeg anmelde? [Internet]. 2020 [henvist 5. marts 2021]. Tilgængelig hos: https://www.nvk.dk/forsker/naar-du-anmelder/hvilke-projekter-skal-jeg-anmelde
- 65. Region Midt. Hvilke projekter skal anmeldes? [Internet]. 2020 [henvist 5. marts 2021]. Tilgængelig hos: https://www.rm.dk/sundhed/faginfo/forskning/de-videnskabsetiskekomiteer/anmeldelse/hvilke-projekter-skal-anmeldes/
- 66. Datatilsynet. Hvad er databeskyttelse? [Internet]. 2020 [henvist 5. marts 2021]. Tilgængelig hos: http://www.datatilsynet.dk/databeskyttelse/hvad-er-databeskyttelse
- 67. Region Midt. Intern fortegnelse over forskningsprojekter med Region Midtjylland som dataansvarlig [Internet]. 2020 [henvist 5. marts 2021]. Tilgængelig hos: https://www.rm.dk/sundhed/faginfo/forskning/Forskningsprojekter/
- 68. Sundhedsdatastyrelsen. Vævsanvendelsesregisteret [Internet]. 2021 [henvist 5. marts 2021]. Tilgængelig hos: https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-de-nationale-sundhedsregistre/testamenter-og-organdonation/vaevsanvendelsesregisteret
- 69. Dabrosin N, Sloth Juul K, Bæhr Georgsen J, Andrup S, Schmidt H, Steiniche T, m.fl. Innate immune cell infiltration in melanoma metastases affects survival and is associated with BRAFV600E mutation status. Melanoma Res. februar 2019;29(1):30–7.
- 70. Bendsen T. Noter i statistik [Internet]. 2009 [henvist 7. maj 2021]. Tilgængelig hos: https://statnoter.dk/index.php?pageID=56
- 71. Nyengaard JR. Stereologic methods and their application in kidney research. J Am Soc Nephrol JASN. maj 1999;10(5):1100–23.
- 72. Giavarina D. Understanding Bland Altman analysis. Biochem Medica. 2015;25(2):141–51.
- 73. Bulten W, Bándi P, Hoven J, Loo RV, Lotz J, Weiss N, m.fl. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. Sci Rep. 2019;9(1):864.
- 74. Jackson CR, Sriharan A, Vaickus LJ. A machine learning algorithm for simulating immunohistochemistry: development of SOX10 virtual IHC and evaluation on primarily melanocytic neoplasms. Mod Pathol. 2020;33(9):1638–48.

- 75. Jørgensen AS, Rasmussen AM, Andersen NKM, Andersen SK, Emborg J, Røge R, m.fl. Using cell nuclei features to detect colon cancer tissue in hematoxylin and eosin stained slides: Detection of Colon Cancer Tissue. Cytometry A. august 2017;91(8):785–93.
- 76. Wang H, Jiang Y, Li B, Cui Y, Li D, Li R. Single-cell spatial analysis of tumor and immune microenvironment on whole-slide image reveals hepatocellular carcinoma subtypes. Cancers. 2020;12(12):1–15.
- 77. Dudley JC, Gurda GT, Tseng L-H, Anderson DA, Chen G, Taube JM, m.fl. Tumor Cellularity as a Quality Assurance Measure for Accurate Clinical Detection of BRAF Mutations in Melanoma. Mol Diagn Ther. august 2014;18(4):409–18.
- 78. Birgin E, Yang C, Hetjens S, Reissfelder C, Hohenberger P, Rahbari NN. Core needle biopsy versus incisional biopsy for differentiation of soft-tissue sarcomas: A systematic review and meta-analysis. Cancer. maj 2020;126(9):1917–28.

Indledningsvist i projektet, er der udført er ustruktureret litteratursøgning på Google, Google Scholar og PubMed, med formålet at skabe overblik over eksisterende litteratur, såvel som at identificere relevante søgetermer.

Der er efterfølgende udført en struktureret litteratursøgning som en bloksøgning, med 4 blokke; cellesegmentering, kunstig intelligens, vævssnit, farvning. Alle søgeordene er søgt i *title/abstract* og der identificeres ligeledes relevante kontrollerede emneord i databaserne, se tabel A-1. Blokke er kombineret med boolske operatorer, hvor ord inden for hver blok er kombineret med OR og AND er anvendt til kombination af de 4 blokke. Søgningen er udført i PubMed og Embase.

Cellesegmentering	Kunstig intelligens	Vævssnit	Farve
Percentage	Artificial intelligence	Whole slide	HE
Fraction	Digital pathology	Histology	Hematoxylin
Segmentation	Deep learning	Pathology	
Cellularity	Neural network		
Immunophenotype		MESH	
	MESH	Histology	
EMTREE	Artificial intelligence		
Cell segmentation	Ai Artificial intelligence	EMTREE	
Tumor cellularity		Histology	
	EMTREE	Whole slide imaging	
	Artificial intelligence		
	Digital pathology		
	Deep learning algorithm		
	Neural network		
	Artificial neural network		

Tabel A-1 - Oversigt over søgeord anvendt i den strukturerede litteratursøgning.

I nedenstående ses søgninger udført i det forskellige databaser. Søgning er udført i marts 2021. Figur A-2 viser søgning udført i Embase og figur A-3 viser den udførte søgning i PubMed.

Figur A-2 - Søgning udført i Embase.

Em	base Session Results	
No.	Query	Results
#5	#1 AND #2 AND #3 AND #4	98
#4	he :ti,ab,kw OR hematoxylin :ti,ab,kw OR 'hematoxylin' /exp	363,626
#3	'whole slide ':ti,ab,kw OR histology :ti,ab,kw OR pathology :ti,ab,kw OR ' histology '/exp OR ' whole slide imaging '/exp OR ' whole slide image'/exp OR ' pathology '/exp	3,116,126
#2	' artificial intelligence ':ti,ab,kw OR ' digital pathology ':ti,ab,kw OR ' deep learning ':ti,ab,kw OR ' neural network ':ti,ab,kw OR ' artificial intelligence'/exp OR ' digital pathology '/exp OR ' deep learning algorithm '/exp OR ' neural network '/exp OR ' artificial neural network'/exp	127,292
#1	percentage:ti,ab,kw OR fraction:ti,ab,kw OR segmentation:ti,ab,kw OR cellularity:ti,ab,kw OR immunophenotype:ti,ab,kw OR 'segmentation'/exp OR 'cellularity'/exp	1,130,967

Search	Actions	Details	Query	Results	Time
#5		>	Search: ((((((((percentage[Title/Abstract]) OR (fraction[Title/Abstract])) OR (segmentation[Title/Abstract])) OR (cellularity[Title/Abstract])) OR (immunophenotype[Title/Abstract])) OR (tumor burden[MeSH Terms])) AND (((("artificial intelligence" [Title/Abstract]) OR ("digital pathology"[Title/Abstract])) OR ("deep learning"[Title/Abstract])) OR ("neural network"[Title/Abstract]))) AND (((("whole slide"[Title/Abstract]) OR (histology[Title/Abstract]))) OR (pathology[Title/Abstract])) OR (histology[MeSH Terms]))) AND (((HE[Title/Abstract]) OR (hematoxylin[Title/Abstract])) Sort by: Most Recent	45	04:44:14
#4	•••	>	Search: (HE[Title/Abstract]) OR (hematoxylin[Title/Abstract]) Sort by: Most Recent	213,929	04:43:50
#3		>	Search: ((("whole slide"[Title/Abstract]) OR (histology[Title/Abstract])) OR (pathology[Title/Abstract])) OR (histology[MeSH Terms]) Sort by: Most Recent	836,948	04:43:11
#2		>	Search: ((("artificial intelligence"[Title/Abstract]) OR ("digital pathology"[Title/Abstract])) OR ("deep learning"[Title/Abstract])) OR ("neural network"[Title/Abstract]) Sort by: Most Recent	61,799	04:41:35
#1		>	Search: (((((percentage[Title/Abstract]) OR (fraction[Title/Abstract])) OR (segmentation[Title/Abstract])) OR (cellularity[Title/Abstract])) OR (immunophenotype[Title/Abstract])) OR (tumor burden[MeSH Terms]) Sort by: Most Recent	828,069	04:40:19

Den identificerede litteratur overføres til Zotero, hvor dubletter fjernes, efterfølgende ekskluderes litteratur på baggrund af opstillede inklusions- og eksklusionskriterier, se tabel A-4. Først sorteres litteraturen ud fra relevans i titel og resume, og efterfølgende ekskluderes litteratur på baggrund af relevans ved gennemlæsning af publikationerne, for endeligt kun at inkludere litteratur, som har relevans for dette projekt.

Tabel A-4 - Inklusions- og eksklusionskriterier for annotering af træningsdata.

Inklusionskriterier	Eksklusionskriterier
Sprog: engelsk eller dansk	Studier på andet end humant materiale
Anvendelse af IHC som reference	
Direkte overførte annoteringer	
Segmentering af specifikke elementer	
HE-farvet vævssnit	
Anvendelse af CNN	

Nedenstående flowchart (figur A-5) illustrerer processen i udvælgelse af litteratur.



Figur A-5 - Flowchart over udvælgelse af litteratur.

Dataopsamlingsskema-Bilag ${\sf B}$

ID	Væv	Væv Type	TC% -	% Арр	Gold Standard		Klassificeringsalgortime						
	, and the second s		eye		normal	tumor	%	SN	FN klasse	FN seg	SP	FP klasse	FP seg
1	metastase	Excisionsbiopsi	65	25	66	271	%=(271/(271+66)*100 = 80	25	1	13	35	6	6
2	metastase	Grovnål	65	29	55	94	%=(94/94+55)*100 = 63	30	20	1	29	0	24
3	metastase	Excisionsbiopsi	75	35	318	380	%=(380/(380+318)*100 = 54	30	20	7	41	5	20
4	metastase	Grovnål	100	51	25	317	%=(317/(317+25)*100 = 93	4	36	0	51	2	33
5	primær	Excisionsbiopsi	60	37	63	137	%=(137/(137+63)*100 = 69	9	10	1	32	1	16
6	primær	Excisionsbiopsi	60	52	160	196	%=(196/(196+160)*100 = 55	3	4	1	43	2	20
7	metastase	Excisionsbiopsi	70	20	238	87	%=(87/(87+238)*100 = 27	20	9	10	32	6	10
8	metastase	Excisionsbiopsi	75	43	43	397	%=(397/(397+43)*100 = 90	7	4	3	36	6	11

9	metastase	Excisionsbiopsi	60	44	108	198	%=(198/(198+108)*100 = 65	6	12	25	23	16	13
10	metastase	Excisionsbiopsi	50	42	132	186	%=(186/(186+132)*100 = 58	2	11	3	30	29	30
11	metastase	Excisionsbiopsi	55	18	188	63	%=(63/(63+188)*100 = 25	34	20	8	11	1	14
12	metastase	Excisionsbiopsi	70	28	410	89	%=(89/(89+410)*100 = 18	2	1	0	20	0	0
13	primær	Excisionsbiopsi	20	17	337	268	%=(268/(268+337)*100 = 44	31	10	1	18	2	14
14	metastase	Grovnål	70	11	34	180	%=(180/(180+34)*100 = 84	26	26	4	11	0	35
15	metastase	Grovnål	95	18	15	161	%=(161/(161+15)*100 = 91	31	21	6	22	2	24
16	metastase	Grovnål	90	4	1	146	%=(146/(146+1)*100 = 99	35	8	4	8	0	9
17	metastase	Grovnål	70	20	0	244	%=(244/(244+0)*100 = 100	15	28	1	36	1	38
18	primær	Excisionsbiopsi	40	26	189	375	%=(375/(375+189)*100 = 66	28	4	18	41	3	29
19	metastase	Grovnål	80	16	420	154	%=(154/(154+420)*100 = 27	26	9	12	26	1	6
20	metastase	Grovnål	50	24	272	261	%=(261/(261+272)*100 = 49	7	47	8	29	0	37
21	metastase	Grovnål	60	15			%=(369/(369+7)*100						
----	-----------	-----------------	-----------	----	-----	-----	----------------------	----	----	------	----	----	----
					7	369	= 98	36	10	3	32	2	15
22	metastase	Grovnål	70	14			%=(121/(121+0)*100						
					0	121	= 100	27	42	14	28	2	52
22	metastaco	Groupål	<u>00</u>	10			%=(440/(440+67)*100						
25	metastase	Grovitar	80	19	67	440	= 87	30	14	5	37	0	11
24	metastase	Excisionsbiopsi	80	41			%=(236/(236+40)*100						
					40	236	= 86	2	8	0	34	0	18
25	primær	Grovnål	50	17			%=(172/(172+65)*100						
					65	172	= 73	80	4	23	0	5	0
26	primær	Excisionsbiopsi	50	13			%=(223/(223+278)*100			_		_	
					278	223	= 45	18	1	0	61	0	18
27	metastase	Grovnål	80	13	100	222	%=(223/(223+199)*100	64	17	11	6	6	15
					199	223	= 53	04	17	- 11	0	0	15
28	metastase	Grovnål	85	4			%=(259/(259+24)*100						
					24	259	= 92	32	48	40	17	10	36
29	metastase	Grovnål	50	23			%=(99/(99+125)*100						
					125	99	= 44	7	17	0	33	1	7
30	metastaso	Excisionshionsi	40	24			%=(51/(51+173)*100						
30	metastase		40	24	173	51	= 23	39	1	18	18	8	2