



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2022 30 stp
Fakultet for realfag og teknologi

Prediksjon av behandlingsutfall for hode- og halskreftpasienter ved bruk av radiomics og repetert elastisk nett teknikk

Prediction of treatment outcome for head and neck
cancer patients using radiomics and Repeated
Elastic Net Technique

Sofie Fjellvang
Miljøfysikk og fornybar energi

Forord

Denne oppgaven er skrevet ved Fakultet for realfag og teknologi ved Norges miljø- og biovitenskapelige universitet våren 2022, og markerer avslutningen på en mastergrad i Miljøfysikk og fornybar energi. Oppgaven har et omfang på 30 studiepoeng.

Jeg vil takke min hovedveileder, professor Cecilia Marie Futsæther, for god oppfølging og grundige tilbakemeldinger gjennom hele arbeidet med masteroppgaven. En stor takk også til førsteamanuensis Oliver Tomic for gode innspill og spennende diskusjoner på våre ukentlige møter. Takk til ph.d-stipendiater Aurora Rosvoll Grøndahl og Bao Ngoc Huynh som alltid har vært tilgjengelige for spørsmål. Det har vært utrolig spennende å få jobbe med dere. Takk også til professor Eirik Malinen (Universitetet i Oslo) og doktor Einar Dale (Oslo universitetssykehus) for tilgjengeliggjøring av datasettet brukt i oppgaven.

Videre vil jeg rette en takk til foreldrene mine, som har støttet meg gjennom 18 år med skolegang, og litt ekstra til pappa som har lest korrektur på masteroppgaven. Takk til Sigbjørn, som har holdt humøret mitt oppe når motivasjonen dalte. Sist, men ikke minst, takk til alle de fantastiske miljøfysikerene jeg har fått dele studietiden med.

Sofie Fjellvang
Ås, 15. mai 2022

Sammendrag

Radiomics er et fagområde som baserer seg på å hente ut egenskaper fra medisinske bilder ved hjelp av ulike dataalgoritmer. Disse egenskapene defineres ut i fra forholdet mellom nabovokslar, og skal avdekke mønstre og karakteristikker i en kreftsvulst, som videre kan brukes i prediktive modeller.

Denne oppgaven tar for seg radiomics av PET- og CT-bilder av pasienter med hode- og halskreft. Målet var å finne ut om radiomics-egenskapene kunne avdekke informasjon som ikke kom fram fra klinisk data eller egenskaper knyttet til svulstens form og størrelse. Det ble også undersøkt hvilke egenskaper som var viktige indikatorer for de beste prediksjonene og om vekselvirkninger mellom ulike egenskaper kunne påvirke prediksjonen i positiv retning. I denne oppgaven predikeres tre ulike behandlingsutfall; sykdomsfri overlevelse, generell overlevelse og lokaltregionalt tilbakefall.

Datasettet benyttet til trening av de prediktive modellene besto av klinisk informasjon, PET-parametere og PET- og CT-avbildninger av 197 pasienter behandlet for hode- og halskreft ved Oslo universitetssykehus i perioden 2007-2013. Fra bildene ble det hentet ut formegenskaper, førsteordens statistiske egenskaper og teksturegenskaper. For å velge ut hvilke egenskaper som skulle benyttes i modellen ble repertert elastisk nett teknikk, RENT, brukt til å velge ut ulike undersett av egenskaper. De ulike undersettene ble trent på åtte klassifiseringsalgoritmer, og resultatene ble validert ved F1-score for begge klasser, nøyaktighet, ROC AUC og Matthews korrelasjonskoeffisient. Modellene ble til slutt testet på et eksternt datasett bestående av 99 pasienter behandlet ved MAASTRO-klinikken i Nederland.

På det eksterne datasettet oppnådde de beste modellene en Matthews korrelasjonskoeffisient på 0.190 for sykdomsfri overlevelse, 0.324 for generell overlevelse, og 0.125 for lokaltregionalt tilbakefall. Undersettene av egenskaper som ledet til de beste prediksjonene var baserte på hovedsakelig kliniske egenskaper og formegenskaper for generell overlevelse, radiomics-egenskaper for lokaltregionalt tilbakefall

og en kombinasjon av kliniske, form- og radiomics-egenskaper for sykdomsfri overlevelse. Ytelsen til modellene sank kraftig fra kryssvalidering til ekstern validering, og det ble avdekket store ulikheter i datasettene brukt til trening og testing av modellene. For å vite om fremgangsmåten brukt i denne oppgaven er et nyttig verktøy for prediksjon av behandlingsutfall for pasienter med hode- og halskreft vil det være nødvendig å teste modellene på flere datasett.

Abstract

Radiomics is a field based on the extraction of features from medical images using different data algorithms. These features are defined from the relationship between neighbouring voxels, and are intended to uncover patterns and characteristics in a tumor that can further be used in predictive analyses.

This paper looks at radiomics of PET and CT images of patients with head and neck cancer. The goal was to find out whether the radiomics features could provide information to predictive models that were not uncovered in clinical data or features related to the tumor's shape and size. It was also investigated which features were important indicators for the best predictions, and whether including interactions between different features could influence the model in a positive direction. The models were trained and tested for three different treatment outcomes; disease free survival, overall survival and loco-regional control.

The dataset used for training the predictive models contained clinical information, PET parameters and PET and CT images of 197 patients treated at Oslo university hospital between 2007 and 2013. Shape features, first order statistical features and texture features were extracted from the images. Repeated elastic net technique, RENT, was used to select different subsets of features for prediction. The different subsets were trained on eight classification algorithms, and the results were validated using F1-score for both classes, accuracy, ROC AUC and Matthews correlation coefficient. Lastly, the models were tested on an unseen dataset containing 99 patients treated at the MAASTRO-clinic in The Netherlands.

On the external dataset, the best models achieved an Matthews correlation coefficient of 0.190 for disease free survival, 0.324 for overall survival and 0.125 for loco-regional control. The subsets of features that lead to the best predictions were based on mainly clinical features and shape features for overall survival, radiomics features for loco-regional control and a combination of shape features, clinical features and radiomics features for disease free survival. The performance of the

models dropped significantly from cross validation to external validation, and big differences between the datasets used for training and testing were uncovered. To know whether the workflow in this study is a useful tool for prediction of treatment outcome for head and neck cancer patients, the models needs to be tested further on more datasets.

Innhold

1	Introduksjon	1
1.1	Motivasjon	1
1.2	Metode	2
1.3	Mål	2
1.4	Oppbygging	3
2	Teori	4
2.1	Kreft	4
2.1.1	Hode- og halskreft	4
2.1.2	Metastase	5
2.1.3	Behandling	5
2.2	Kjernefysikk	6
2.2.1	Elektromagnetisk stråling	6
2.2.2	Nukleær stråling	7
2.2.3	Vekselvirkning mellom stråling og materie	8
2.3	Diagnosering	12
2.3.1	Positronemisjonstomografi	12
2.3.2	Computertomografi	17
2.4	Maskinlæring	20
2.4.1	Ulike typer maskinlæring	21
2.4.2	Klassifiseringsalgoritmer	21
2.4.3	Tilpasning av modeller	28
2.4.4	Validering av modeller	30
3	Materiale og metode	35
3.1	Datasettet	35
3.2	Programvare	39
3.3	Preprosessering	40
3.3.1	Preprosessering av kliniske data	40
3.3.2	CT-windowing	41
3.4	Radiomics	42

3.4.1	Bildeprosessering	43
3.4.2	Uthenting av egenskaper	44
3.4.3	Imskaper	51
3.5	Arbeidsflyt	52
3.6	Egenskapsutvelging	52
3.6.1	Korrelasjon mellom egenskaper fra ulike grupper	53
3.6.2	Konstruksjon av datasett	53
3.6.3	Repetert elastisk nett teknikk	53
3.6.4	Valg av egenskaper	55
3.6.5	Korrelasjon mellom valgte egenskaper	55
3.7	Etablering av rammeverk for prediksjon	56
3.7.1	Standardisering av data	56
3.8	Validering av modeller	57
3.8.1	Kryssvalidering	57
3.8.2	Grid search-kryssvalidering	58
3.9	Modeller med vekselvirkninger	59
3.10	Etablere en baseline	59
3.11	Testing på eksternt datasett	60
4	Resultater	61
4.1	Preliminær egenskapsutvelging	61
4.1.1	Korrelasjon mellom egenskaper fra ulike grupper	61
4.1.2	Konstruksjon av datasett	62
4.2	Baseline-modeller	63
4.3	DFS	65
4.3.1	Egenskapsutvelging	65
4.3.2	Egenskapsutvelging med vekselvirkningsegenskaper	66
4.3.3	Kryssvalidering	69
4.3.4	Testing på eksternt datasett	72
4.4	OS	75
4.4.1	Egenskapsutvelging	75
4.4.2	Egenskapsutvelging med vekselvirkningsegenskaper	77
4.4.3	Kryssvalidering	79
4.4.4	Testing på eksternt datasett	81
4.5	LRC	85
4.5.1	Egenskapsutvelging	85
4.5.2	Egenskapsutvelging med vekselvirkningsegenskaper	87
4.5.3	Kryssvalidering	88
4.5.4	Testing på eksternt datasett	90
5	Diskusjon	95

5.1	Datasettet	95
5.2	RENT som egenskapsutvelger	96
5.3	Evaluering av utvalgte egenskaper	97
5.4	Evaluering av resultater	99
5.4.1	DFS	99
5.4.2	OS	100
5.4.3	LRC	101
5.5	Ulikheter i datasett brukt til trening og testing	102
5.6	Diskretisering av bildene	108
5.7	Overoptimistiske resultater i kryssvalidering	108
5.8	Forslag til videre arbeid	109
6	Konklusjon	111
	Bibliografi	112
	Tillegg	
A	Parametere	122
A.1	Logistisk regresjon	122
A.2	Random forest	123
A.3	K-nærmeste naboer	123
A.4	Support Vector Machines	124
A.5	AdaBoost	125
A.6	Kvadratisk diskriminantanalyse	125
A.7	Passive aggressive	126
A.8	Light gradient boosting machine	126
B	Bildeegenskaper	128
B.1	Formegenskaper	128
B.2	Førsteordens statistiske egenskaper	131
B.3	Teksturegenskaper	133
C	Resultater fra kryssvalidering og prediksjon	148
D	Undersett 0.2-0.4 - LRC	158
E	Undersett 0.5v for LRC	162

Figurer

2.2.1 Fundamental prosess for positronemisjon	8
2.2.2 Fotoelektrisk effekt	9
2.2.3 Comptonspredning	9
2.2.4 Koherent spredning	10
2.2.5 Pardannelse	11
2.2.6 Annihilering	11
2.3.1 PET-avbildning	12
2.3.2 Fluordeoxyglukose	13
2.3.3 Prinsippet bak PET-avbildning	14
2.3.4 PET-avbildning: Spredt og tilfeldig deteksjon	15
2.3.5 Sinogram	16
2.3.6 CT-avbildning	18
2.3.7 Filtered back propagation	20
2.4.1 Datamatriser til trening og testing	22
2.4.2 Bestemmelsestre	24
2.4.3 K-nærmeste naboer	25
2.4.4 Lineære support vector machines	26
2.4.5 Ikke-lineære support vector machines	26
2.4.6 Tilpasning av en modell	29
2.4.7 L2- og L1-regularisering	30
2.4.8 Feilmatrix for prediksjon	31
2.4.9 ROC-kurve	32
3.3.1 CT-avbildning med og uten preprosessering	42
3.4.1 Arbeidsflyt i radiomics-prosedyren	42
3.4.2 Diskretisering av PET-avbildning	43
3.4.3 Orientering til nabovokslar relativt til en sentervoksel	45
3.4.4 Gray Level Size Zome Matrix	46
3.4.5 Gray Level Run Length Matrix	47
3.4.6 Neighbouring Gray Tone Difference Matrix	48
3.4.7 Gray Level Dependence Matrix	49

3.4.8 Gray Level Co-occurrence Matrix	50
3.4.9 Local Binary Pattern	51
3.5.1 Arbeidsflyt for å finne prediksjonsmodell med høyest ytelse	52
3.8.1 5-foldet kryssvalidering	58
4.1.1 Korrelasjon mellom egenskaper fra ulike egenskapsgrupper	62
4.3.1 Frekvensplott for DFS	65
4.3.2 Oversikt over undersett basert på ulike τ -verdier for DFS	67
4.3.3 Resultater fra kryssvalidering - DFS	70
4.3.4 Ytelse på eksternt datasett - DFS	72
4.3.5 PCA av undersettet med høyest ytelse under kryssvalidering for DFS	75
4.4.1 Frekvensplott for OS	76
4.4.2 Oversikt over undersett basert på ulike τ -verdier for OS	77
4.4.3 Resultater fra kryssvalidering - OS	80
4.4.4 Ytelse på eksternt datasett - OS	82
4.4.5 PCA av undersettet med høyest ytelse under kryssvalidering for OS	85
4.5.1 Frekvensplott for LRC	86
4.5.2 Oversikt over undersett basert på ulike τ -verdier for LRC	87
4.5.3 Resultater fra kryssvalidering - LRC	89
4.5.4 Ytelse på eksternt datasett - LRC	91
4.5.5 PCA av undersettet med høyest ytelse under kryssvalidering for LRC	94
5.3.1 Fellestrekk mellom egenskaper valgt for ulike responser	97
5.5.1 PCA på begge datasett	102
5.5.2 Distribusjonen av variablene <i>age</i> og <i>pack_years</i>	103
5.5.3 Distribusjon av binære egenskaper	104
5.5.4 Klassedistribusjon for OUS-data og MAASTRO-data	105
5.5.5 Oversikt over antall egenskaper valgt for de ulike datasettene	105
5.5.6 PCA på utvalgte egenskaper	107

Tabeller

3.1	Klassedistribusjon OUS-datasett	36
3.2	Oversikt over kliniske egenskaper i datasett	37
3.3	Oversikt over PET-parametere i datasett	38
3.4	Klassedistribusjon i MAASTRO-datasett	39
3.5	Ny kategoriinndeling for <i>hpv_related</i> og <i>uicc8_III – IV</i>	41
3.6	Metoder for egenskapsuthenting	44
3.7	Parametere i RENT	54
4.1	Oversikt over egenskaper	63
4.2	Baseline-ytelse for de ulike responsene	64
4.3	Undersett 0.9v for responsen DFS	68
4.4	Modellene uten vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for DFS	69
4.5	Modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for DFS	71
4.6	Modellene uten vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for DFS testet på eksternt datasett , a), og modellene med høyest ytelse på det eksterne datasettet, b).	73
4.7	Modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for DFS testet på eksternt datasett, a), og modellene med vekselvirkningsegenskaper med høyest ytelse under testing på eksternt datasett for DFS, b)	74
4.8	Undersett 0.5v for responsen OS	78
4.9	Modellene uten vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for OS	79
4.10	Modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for OS	81
4.11	Modellene uten vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for OS testet på eksternt datasett, a), og modellene uten vekselvirkningsegenskaper med høyest ytelse under testing på eksternt datasett for OS, b)	83

4.12	Modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for OS testet på eksternt datasett, a), og modellene med vekselvirkningsegenskaper med høyest ytelse under testing på eksternt datasett for OS, b)	84
4.13	Modellene uten vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for LRC	88
4.14	Modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for LRC	90
4.15	Modellene uten vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for LRC testet på eksternt datasett, a), og modellene uten vekselvirkningsegenskaper med høyest ytelse under testing på eksternt datasett for LRC, b)	92
4.16	Modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering for LRC testet på eksternt datasett, a), og modellene med vekselvirkningsegenskaper med høyest ytelse under testing på eksternt datasett for LRC, b)	93
A.1	Tilpassede hyperparametere for logistisk regresjon.	122
A.2	Tilpassede hyperparametere for random forest.	123
A.3	Tilpassede hyperparametere for KNN.	123
A.4	Tilpassede hyperparametere for SVM.	124
A.5	Tilpassede hyperparametere for AdaBoost.	125
A.6	Tilpassede hyperparametere for QDA.	125
A.7	Tilpassede hyperparametere for passive aggressive.	126
A.8	Tilpassede hyperparametere for LightGBM.	126
B.1	Oversikt over formegenskapene brukt i oppgaven.	128
B.2	Oversikt over de førsteordens statistiske egenskapene brukt i oppgaven.	131
B.3	Oversikt over GLSZM-egenskapene brukt i oppgaven.	133
B.4	Oversikt over GLRLM-egenskapene brukt i oppgaven.	136
B.5	Oversikt over NGTDM-egenskapene brukt i oppgaven.	139
B.6	Oversikt over GLDM-egenskapene brukt i oppgaven.	141
B.7	Oversikt over GLCM-egenskapene brukt i oppgaven.	143
B.8	Oversikt over LBP-egenskapene brukt i oppgaven.	147
D.1	Oversikt over egenskapene inkludert i undersett 0.2-0-4 for LRC . .	158

Liste over forkortelser

Forkortelse	Betydning
HPV	Humant papillomavirus
CT	Computertomografi
SPECT	Single-photon emission computerized tomography
PET	Positronemisjonstomografi
(¹⁸ F-)FDG	Flourdeoxyglukose
LOR	Line of response
TOF	Time of flight
SUV	Standard uptake value
FBP	Filtered back projection
HU	Hounsfield unit
IG	Information gain
KNN	K-nærmeste naboer
SVM	Support vector machines
QDA	Quadratic discriminant analysis
LDA	Lineær diskriminantanalyse
TP	True positive
FP	Falsk positiv
FN	Falsk negativ
TN	True negative
TPR	True positive rate
FPR	False positive rate
ROC	Receiver operator characteristics
AUC	Area under curve
PRE	Precision

REC	Recall
MCC	Matthews correlation coefficient
OUS	Oslo universitetssykehus
ROI	Region of interest
DFS	Desease free suvival
LRC	Loco-regional control
OS	Overall survival
MTV	Metabolsk tumorvolum
TLG	Total lesion glycosis
WL	Window level
WW	Window width
IBSI	Image Biomarker Standardization Initiative
GLSZM	Gray level size zone matrix
GLRLM	Gray level run length matrix
NGTDM	Neighbouring gray tone difference matrix
GLDM	Gray level dependence matrix
GLCM	Gray level co-occurence matrix
LBP	Local binary pattern
RENT	Repetert elastisk nett teknikk
RV	Relativ varians
PCA	Principal component analysis

Liste over symboler

Symbol	Betydning	Verdi
B_e	Bindingsenergi	
BW	Kroppsvekt	
c	Lysets hastighet	299 792 458 m/s
C	Konsentrasjon (i computertomografi)	
C	Reguleringsparameter (i maskinl�ring)	
C_{st}	Konstant	
E	Energi	
h	Planckkonstanten	$6,6260710^{-34}$ Js
I	Intensitet	
ID	Injisert dose	
L	Lengde	
λ	B�lgelengde (i fysikk)	
λ	Reguleringsparameter (i maskinl�ring)	
m	Masse	
μ	Dempningskoeffisient (i fysikk)	
μ	Gjennomsnitt (i statistikk)	
ν	Frekvens	
p	Sannsynlighet	
σ	Standardavvik	
w	Vekt	
x	Avstand	
z	Tilfelle	

Kapittel 1

Introduksjon

1.1 Motivasjon

Kreft er en fellesbetegnelse på sykdommer som skyldes ukontrollert celledeling, og som rammet 35 515 nordmenn i 2020 [1], [2]. Prognosene varierer for ulike krefttyper, og med nye behandlingsmetoder blir andelen pasienter som overlever kreft stadig høyere [3]. I dag overlever så mange som tre av fire som mottar en kreftdiagnose i minst fem år [2]. Samtidig diagnostiseres stadig flere med kreft, og det anslås at rundt en tredjedel av alle nordmenn kommer til å motta en kreftdiagnose før fylte 75 år [2]. I 2020 var kreft den aller vanligste dødsårsaken i Norge, foran hjerte- og karsykdommer som lå øverst frem til 2017 [4]. En kreftdiagnose stilles blant annet ved bruk av medisinske bilder, som PET og CT [5].

Hode- og halskreft er fellesbetegnelsen på kreft i hode-og halsregionen, som inkluderer svelg, nese og bihuler, leppe og munnhule, strupehode og spyttkjertler [6]. Infeksjon av humant papilomavirus (HPV), tobakk og fedme knyttes ofte til kreft i denne regionen. Blant alle som fikk kreft i Norge i 2020 ble 1.9% diagnostert med hode- og halskreft [6].

De vanligste formene for kreftbehandling er stråleterapi, cellegift og operasjon [3], [7]. Disse behandlingene har som mål å stanse den ukontrollerte celledelingen uten å påføre skade på områdene rundt svulsten. Cellegift og overflødig stråling til friskt vev kan føre til senskader som kan frembringe plager og fare for ny sykdom hos pasienten resten av livet [8]. Det er derfor ikke ønskelig å gi pasienter overflødig behandling, og heller ikke å gi behandling som ikke fungerer. Dersom det er mulig

å predikere utfallet av en behandling før den påbegynnes, vil det være mulig å skreddersy behandlingen for den enkelte pasienten som vil kunne bedre prognosen og minske risikoen for, og alvorligheten av, senskader.

1.2 Metode

Maskinlæring er en form for kunstig intelligens som kan benyttes til å lage prediktive modeller [9]. Modellene trenes på et datasett med kjent respons, før de brukes til å predikere respons for usett data. Innen medisin kan maskinlæring blant annet brukes til å predikere utfallet av kreftbehandling, for eksempel ved å se på de kliniske egenskapene, som alder, HPV-status og kreftens lokasjon, til pasienten og formegenskapene til svulsten.

Radiomics er en metode som bruker ulike dataalgoritmer til å trekke ut egenskaper fra medisinske bilder, som PET og CT [10]. Disse egenskapene skal avdekke mønstre og karakteristikk til kreftsvulsten, og målet er at de skal kunne styrke ytelsen til de prediktive modellene ved å finne viktige sammenhenger som ikke kommer til syne i de kliniske dataene.

Hode- og halskreft har lav overlevningsrate for høyere kreftstadier, og det er funnet sammenhenger mellom karakteristikkene avdekket av radiomics-egenskaper og overlevelse [11]. Andre studier indikerer at ulike radiomics-baserte modeller kan gi bedre prediksjoner av lokaltregionalt tilbakefall og spredning enn modeller kun basert på klinisk data, mens overlevelse best predikeres av kun kliniske egenskaper [12]. Radiomics er et fagfelt i sterk vekst, og ved å inkludere prediksjonsmodeller basert på radiomics i behandlingsfasen av hode- og halskreftpasienter er målet at prediksjonene skal kunne støtte onkologene i beslutningstakingen om den enkelte pasients behandling [13].

1.3 Mål

Målet med denne oppgaven er å bruke radiomics-egenskaper beregnet fra PET- og CT-avbildninger sammen med kliniske egenskaper til å predikere behandlingsutfallet for pasienter med hode- og halskreft. Dersom det er mulig å predikere hvordan en behandlingsform vil fungere på en svulst med gitt karakteristikk vil det være mulig å tilpasse behandlingen til hver enkelt pasient. Det er også et mål i denne

oppgaven å avdekke hvilke egenskaper av svulsten som er de viktigste indikatorene for korrekt prediksjon av behandlingsutfall.

1.4 Oppbygging

Denne oppgaven starter med et teorikapittel som presenterer prinsippene bak PET- og CT-avbildning og maskinlæring. Deretter følger kapittel 3 om materiale og metode, som tar for seg datasettet brukt i oppgaven, radiomics-prosedyren, utvelging av svulstegenskaper og modellering ved hjelp av flere maskinlæringsalgoritmer. Resultatene presenteres i kapittel 4, etterfulgt av diskusjon i kapittel 5. Til slutt følger en konklusjon og forslag til videre arbeid.

Kapittel 2

Teori

2.1 Kreft

Kreft er en samlebetegnelse på om lag 200 sykdommer som skyldes unormal cellevekst, og kan oppstå i alle kroppens organer og vev [1], [3]. Når celler dør, vil kroppen kontinuerlig erstatte disse med nye celler ved å doble sitt arvestoff, DNA, og deretter dele seg i to [3]. Dersom denne celledelingen blir ukontrollert grunnet en skade eller mutasjon i arvestoffet vil cellene etter hvert hoppe seg opp i organet eller vevet og danne en kreftsvulst. Sykdomsforløp, symptomer, prognose og behandling vil variere for de ulike krefttypene [3].

Kreft var den vanligste dødsårsaken i Norge i 2020, med 222 dødsfall per 100 000 innbyggere [4]. Rundt en tredjedel av alle kreftforekomster kan knyttes til levemåte, blant annet bruk av alkohol og tobakk, lite mosjon og dårlig kosthold [3]. Andre faktorer som øker risikoen for kreft er virus, stråling og kjemikalier, i tillegg til arv og miljø .

2.1.1 Hode- og halskreft

Hode- og halskreft er fellesbetegnelsen på kreft i hode- og halsregionen, derunder strupehode (larynx), leppe og munnhule (cavum oris), svelg (pharynx), nese og bihuler og spyttkjertler [6]. Faktorer som gir økt risiko for kreft i dette området er infeksjon av HPV, røyk og snus, alkohol, kjemiske stoffer og fedme [6]. Tobakk knyttes oftest til kreft i munnhule, strupesvelg (hypopharynx) og strupehode, mens

HPV knyttet til kreft i munnsvelg (oropharynx) som inkluderer mandler og tungebase [14].

686 mennesker ble diagnostisert med hode- og halskreft i 2020, og av disse var cirka to tredjedeler menn [6]. Hode- og halskreft utgjorde dermed 1,93 prosent av alle krefttilfeller i Norge [2], [6]. Denne kreftformen viser en årlig svak stigning, som har blitt knyttet til infeksjon med HPV [6], [15]. Fem år etter diagnose er 70,7 prosent av menn og 75,6 prosent av kvinnene fortsatt i live [6].

2.1.2 Metastase

Metastase er spredning av kreftceller til andre kroppsdeler enn der den opprinnelige svulsten oppsto [16]. Kreftcellene løsrives fra svulsten og spres gjennom blod eller lymfbaner til andre steder i kroppen [3]. Spredningen kan føre til nye svulster som vil gjøre behandlingen mer utfordrende.

Hode- og halskreft kan spre seg til lymfeknutene på halsen, men det er sjeldent at den sprer seg videre i kroppen [6].

2.1.3 Behandling

Målet med kreftbehandlingen er å drepe kreftcellene og/eller stanse den ukontrollerte celledelingen [17]. De vanligste formene for kreftbehandling er cellegift, kirurgi og strålebehandling, og det er også disse som dominerer i behandlingen av hode- og halskreft [3], [7]. Ofte brukes en kombinasjon av behandlingsformene for å øke sannsynligheten for overlevelse og minske sannsynligheten for, og alvorlighetsgraden av, senskader. Vanlige senskader etter hode- og halskreft er stråleskader i hud, munntørrhet, skader i tyggemuskulatur og taleutfordringer (knyttet til kreft i strupehodet) [6]. I tillegg medfører strålebehandling og cellegift en risiko for sekundær kreft, det vil si kreft som oppstår i et annet organ grunnet kreftfremkallende virkning fra den primære behandlingen, som oftest etter 10-30 år [8].

For å optimalisere pasientenes prognose er det viktig å stille rett diagnose hurtig, slik at korrekt behandlingsmetode kan igangsettes [18]. For hode- og halskreft øker sannsynligheten for overlevelse med cirka 15 prosent dersom riktig diagnose stilles før metastase og behandlingen hindrer videre spredning [3].

2.2 Kjernefysikk

I denne seksjonen forklares kjernefysiske begreper relevante for seksjon 2.3.

2.2.1 Elektromagnetisk stråling

Elektromagnetisk stråling er energi som beveger seg med lysets hastighet gjennom fritt rom eller materie i form av elektromagnetiske bølger [19]. All elektromagnetisk stråling består av fotoner, som er en type elementærpartikkel uten hvilemasse, men med energi og bevegelsesmengde [20]. Energien til et foton er gitt ved

$$E = h\nu = \frac{hc}{\lambda} \quad (2.1)$$

der h er Planckkonstanten, ν er fotonets frekvens, c er lysets hastighet og λ er fotonets bølgelengde [21]. Eksempler på elektromagnetisk stråling er synlig lys, radiobølger, røntgenstråler og gammastråler [19], [22].

Røntgenstråling

Røntgenstråler er fotoner som oppstår som bremsestråling når ulike ladede partikler bremses eller aksellereres, og frembringes oftest i et røntgenrør [22]. Disse fotonene har energi opp til 150 keV [23]. Røntgenstråler kan passere gjennom biologisk materie, og gjør det mulig å skille mellom bein og andre typer vev grunnet deres ulike dempningsegenskaper [21]. Et røntgenbilde lages ved at fotonene som har passert gjennom kroppen treffer en detektor eller film på den andre siden [24]. Beinvev demper røntgenstrålene i stor grad, og vil fremstå som hvitt på røntgenbildet, mens bløtvev, som slipper strålene lettere igjennom, vil vises i ulike gråtoner. I medisinsk sammenheng brukes røntgenbilder til deteksjon av brudd og ved kartlegging av sykdom [24]. Røntgenstråling danner grunnlaget for computertomografi (CT) [21].

Gammastråling

Gammastråling er fotoner som oppstår når en atomkjerne desintegrerer, eller ved annihiling, som forklares i seksjon 2.2.3. [23]. Gammastråler har kortere bølgelengde

enn røntgenstråler, og fra likning 2.1 derav også høyere energi. Energiområdet til gammastråler strekker seg opp til 10 MeV, og overlapper med energinivået til røntgenstråler for de laveste energinivåene [23]. Derfor skilles røntgen- og gammastråler som oftest på metoden for fremstilling.

2.2.2 Nukleær stråling

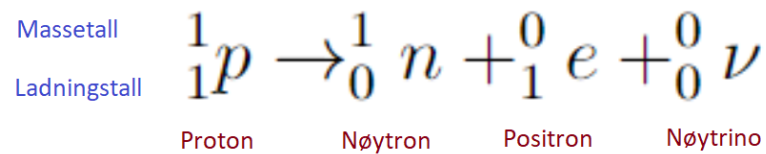
Nukleær stråling er strålingen som sendes ut fra kjernen til radioaktive stoffer, også kalt radionuklider [25]. Sammen med gammastråling er de vanligste formene for nukleær stråling alfa- og betastråling. I alfastråling sendes en alfapartikkel ut fra atomkjernen [26]. Denne består av to protoner og to nøytroner, og er derfor positivt ladet. Alfapartikkelen er tung og har dårlig penetreringsevne, men kan gjøre stor skade på områdene den treffer. Betastråling består av elektroner eller positroner, β^- - og β^+ -partikler, som sendes ut fra kjernen når et nøytron omgjøres til et proton eller motsatt [27]. Betapartiklene er vanskeligere å stoppe enn alfapartiklene, da de er mindre og har høyere hastighet [25]. Gammastråling etterfølger oftest alfa- eller betastråling, der datterkjernen, som er i en eksitert tilstand etter utsendelsen, kvitter seg med eksitasjonsenergien [25].

Gammastråling kan detekteres av et gammakamera, som brukes til å gi et todimensjonalt bilde av stråleaktiviteten i et organ, basert på den oppsamlede gammastrålingen fra et radioaktivt legemiddel injisert i pasienten [28]. Gammakameraet benyttes også i avbildningsteknikken *single-photon emission computed tomography* (SPECT), som gir et tredimensjonalt bilde av pasienten [28]. Beta-pluss-stråling og gammastråling danner grunnlaget for positronemisjonstomografi (PET), der emisjon av positroner, som videre omgjøres til gammastråler, gjør det mulig å oppdage områder med økt metabolsk aktivitet [29].

Positronemisjon

Positronemisjon finner sted når det er en ubalanse i forholdet mellom antallet protoner og nøytroner som skal til for at en atomkjerne holder seg stabil [30]. Dersom kjernen inneholder for mange protoner vil den emitte et positron i forsøk på å oppnå stabilitet. Et proton omgjøres til et nøytron i kjernen, og overskuddsladningen sendes ut i form av et positron og et nøytrino [30]. Nøytrinoet har ingen ladning og svært liten masse, som gjør at det vekselvirker lite med materie og er svært vanskelig å oppdage [27]. For at positronemisjon skal være mulig må energiforskjellen mellom morkjernen og datterkjernen være over 1.022 MeV [30]. Den

fundamentale prosessen bak positronemisjon er illustrert i figur 2.2.1.



Figur 2.2.1: Fundamental prosess for positronemisjon. Massetallet forteller hvor mange nukleoner (protoner og nøytroner) det er i partikkelen. Ladningstallet beskriver ladningen til partikkelen i forhold til en elementærladning. Figur adaptert fra [30].

2.2.3 Vekselvirkning mellom stråling og materie

Elektromagnetisk stråling vekselvirker med materie hovedsakelig gjennom fotoelektrisk effekt, Comptonspredning og pardannelse, avhengig av fotonenergien og materien fotonet vekselvirker med [31].

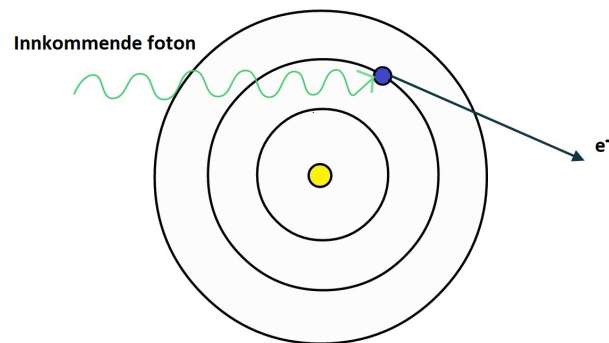
Alfa- og betapartikler, vekselvirker hovedsakelig ved at de ioniseres og eksiteres etter kollisjon med elektroner og positroner [31]. Fordi alfapartikkelen er tung, blir den minimalt avvøyd i kollisjonene, i motsetning til betapartiklene som også kan endre retning [26].

Fotoelektrisk effekt

Fotoelektrisk effekt er den dominerende vekselvirkningen mellom fotoner og materie ved energinivåer lavere enn 0.1 MeV [31]. Når et foton kolliderer med et bundet elektron i atomet absorberes all fotonenergien, og elektronet løsriveres, som illustrert i figur 2.2.2 [26]. Noe energi absorberes av atomet, samt at bindingsenergien til elektronet må overvinnnes for at det skal løsriveres. Den kinetiske energien til elektronet gis derfor ved

$$E_k = h\nu - E_A - B_e \quad (2.2)$$

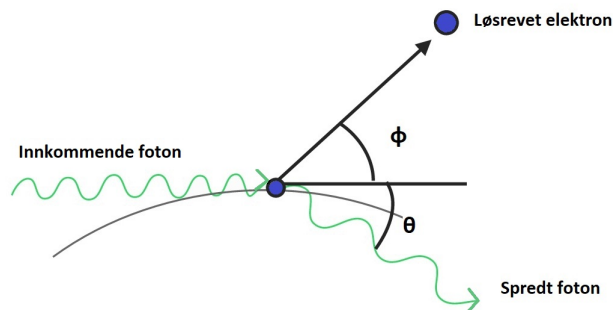
der E_A er energien som absorberes av atomet og B_e er bindingsenergien til elektronet [31]. Fotoelektrisk effekt skjer ikke i frie elektroner, da bevegelsesmengden ikke vil bevares [26].



Figur 2.2.2: Et innkommende foton kolliderer med et bundet elektron, og elektronet løsriveres. Adaptert fra [31].

Comptonspredning

Comptonspredning blir den dominerende vekselvirkningen ved fotonenergier over 25 keV, og er hovedårsaken til spredt stråling i vev [21], [31], [32]. Spredningen oppstår når et foton treffer løst bundne elektroner i valensskallet til et atom, og fører til at fotonet blir spredt og elektronet løsriveres, som illustrert i figur 2.2.3.



Figur 2.2.3: Et innkommende foton kolliderer med et valenselektron, og endrer retning etter sammenstøtet. Adaptert fra [31].

Energibevaringen i støtet gis ved

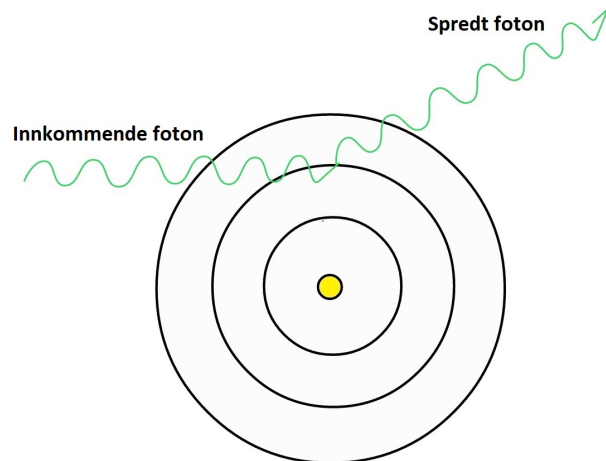
$$E_\gamma = h\nu = E'_\gamma + E_e \quad (2.3)$$

der E_γ er energien til innkommende foton, E'_γ er energien til det spredte fotonet og E_e er energien til det løsrevne elektronet [26]. Likning 2.3 viser at energien

til det innkommende fotonet deles mellom det spredte fotonet og det løsrevne elektronet. Det spredte fotonet kan videre kollidere med andre elektroner, og vil dermed fortsette å miste energi.

Koherent spredning

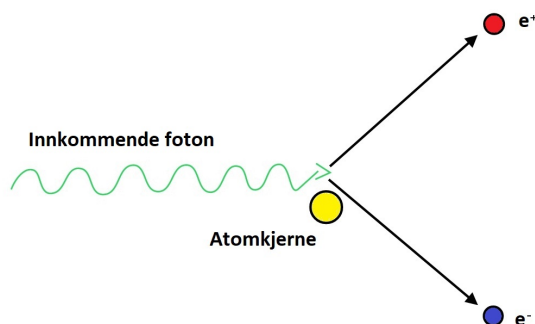
Koherent spredning forekommer når fotoner med lav energi spres av partikler med radius mindre enn en tiendedel av fotonets bølgelengde [21], [33]. Fotonet absorberes av alle elektronene i atomet, for så å bli sendt ut igjen i en annen retning, men med samme frekvens, som illustrert i figur 2.2.4 [26]. Det vil si at sammenstøtet mellom fotonet og atomet er elastisk. Koherent spredning fører ikke til at ladede partikler eksiteres, ioniseres eller får overført energi [26].



Figur 2.2.4: Et innkommende foton absorberes av atomet, og sendes ut i en annen retning med samme bølgelengde.

Pardannelse

Pardannelse er den dominerende vekselvirkningen ved energinivåer høyere enn det dobbelte av elektronets hvilemasse, $2m_e c^2$, eller 1.02 MeV [31]. Et energirikt foton i Coloumbfeltet til en atomkjerne omdannes til et elektron-positron par, som illustrert i figur 2.2.5 [21]. Videre vil elektronene og positronene kollidere med andre betapartikler i en prosess som kalles annihilering [26].



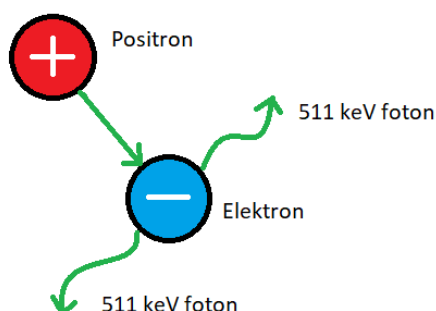
Figur 2.2.5: Et innkommende foton omdannes til et elektron-positron-par i nærheten av en atomkjerne. Adaptert fra [31].

Annihilering

Annihilering oppstår når et positron kolliderer med et elektron [29]. Positroner har lik masse og spinn som elektroner, men motsatt ladningstall [34]. Elektroner og positroner er dermed det som kalles et partikkel-antipartikkelpar, og vil tiltrekkes hverandre grunnet den motsatte ladningen. Når et positron og et elektron kolliderer frigjøres energi i form av to fotoner som sendes ut med en vinkel på 180 grader mellom seg, som illustrert i figur 2.2.6 [35]. Konverteringen fra masse til energi beskrives av masseenergiloven,

$$E = mc^2 = m_p c^2 + m_e c^2 \quad (2.4)$$

, der m er den samlede massen til positronet og elektronet, m_p er massen til positronet og m_e er massen til elektronet [35]. Fra likning 2.4 vil energien til hvert av fotonene tilsvare 511 keV.



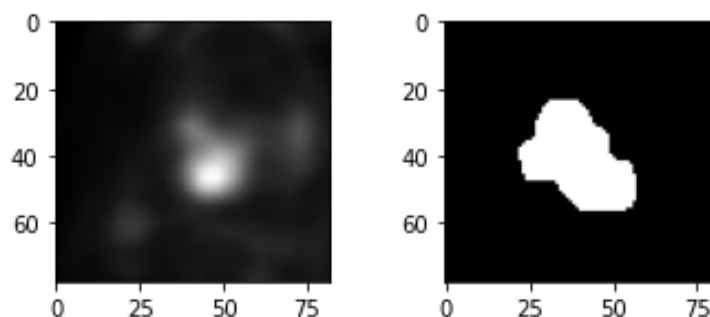
Figur 2.2.6: Når et proton kolliderer med et elektron sendes to fotoner med energi 511 keV ut med 180 grader i mellom seg. Adaptert fra [29].

2.3 Diagnoser

En kreftdiagnose kan stilles ved hjelp av ulike undersøkelser, blant annet celleprøver, vevsprøver, ultralyd og ulike medisinske bilder, herunder PET og CT, og ofte en kombinasjon av disse [5]. Undersøkelsene utføres ved symptomer eller mistanke om kreft, eller i forbindelse med rutinekontroller. PET og CT er svært avanserte bildeundersøkelser, som blir forklart i detalj i seksjon 2.3.1 og 2.3.2. Kombinasjonen av PET- og CT- bilder gjør det mulig å finne ut hvor i kroppen det har oppstått en opphopning av kreftceller og omfanget av opphopningen [36].

2.3.1 Positronemisjonstomografi

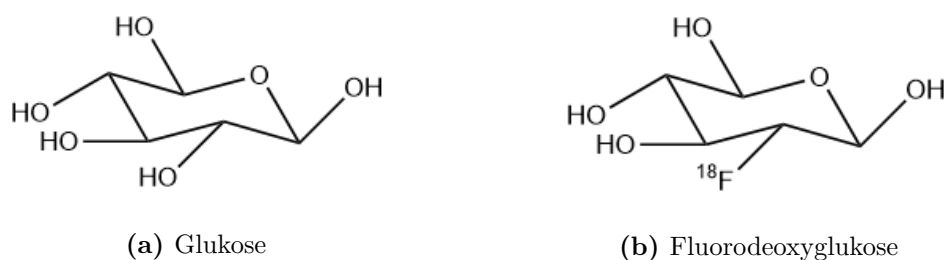
Positronemisjonstomografi (PET) benyttes til å stille diagnose, følge opp behandling, kartlegge utbredelsen til svulsten, skille mellom godartede og ondartede svulster og oppdage metastase [35], [37]. PET baserer seg på deteksjon av radioaktivitet, etter at en liten mengde radioaktivt sporingstoff injiseres i kroppen gjennom en perifer vene. Dette stoffet emitterer positroner gjennom β^+ -desintegrasjon. De emitterte positronene fra sporingstoffet kolliderer med elektroner i vevet og annihileres til gammastråler som detekteres av en detektorring rundt pasienten. Her registreres signalene og sendes til en datamaskin som omdanner dem til tredimensjonale bilder [35], [38]. Moderne PET-skannere produserer bilder med en romlig oppløsning på 1 – 6 mm [29]. En PET-avbildning med tilhørende maske som markerer svulsten vises i figur 2.3.1.



Figur 2.3.1: Bildet til venstre viser et tverrsnitt av PET-avbildningen av pasient nummer 40 fra datasettet brukt i denne oppgaven. Til høyre vises den tilhørende masken som markerer hvor svulsten befinner seg.

Fluorodeoxyglucose

Fluorodeoxyglucose (^{18}F -FDG) er et mye brukt sporingstoff ved PET-avbildning av hode og hals [39]. ^{18}F -FDG består av et glukosemolekyl der hydroksylgruppen i C-2-posisjon har blitt byttet ut med radionukleidet ^{18}F , illustrert i figur 2.3.2 [40]. Kreftceller tar opp sukker, noe som fører til at sporingstoffet, som har utgangspunkt i et glukosemolekyl, hoper seg opp der det befinner seg kreftceller [36].



Figur 2.3.2: Forskjellen mellom glukose og fluorodeoxyglukose, der den ene OH-gruppen er erstattet med radionukleidet fluor-18.

Når ^{18}F -FDG benyttes som sporingstoff kan et PET-bilde reflektere glukosetransporten på tvers av cellemembranene, og vise hvordan glukosen fordeler seg i kroppen [35], [38]. Den radioaktive ^{18}F -kjernen har en fysisk halveringstid på 110 minutter, og hele 97% av den avgitte strålingen stammer fra positronemisjon [41]. De emmiterte positronene vil vekselvirke med elektronene i vevet gjennom annihilering [35]. Strålebelastningen fra sporingstoffet er ikke større enn ved vanlig røntgen, da det er en veldig liten mengde som injiseres, og sporingstoffet vil være helt ute av kroppen etter 4-5 timer [36], [38]. Under en PET-avbildning eksponeres pasienten for opp mot 25 mSv, som tilsvarer den gjennomsnittlige bakgrunnsstrålingen et menneske utsettes for i løpet av åtte år [42].

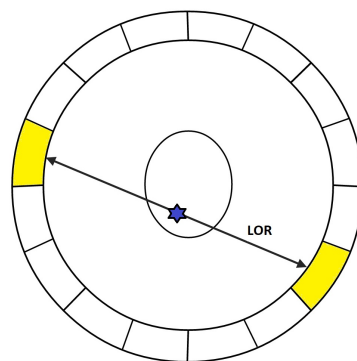
Radionuklider som ^{18}F produseres i en syklotron, som er en elementærpartikkelakselerator [29]. Tyngre partikler som protoner og deuteriumkjerner akselereres til et energinivå på 5 – 100 MeV, før en målkjerne bombarderes med partiklene for å produsere de nødvendige radionukleidene. For ^{18}F brukes H_2^{18}O som målkjerne, det vil si et vannmolekyl der oksygenatomet inneholder to nøytroner flere enn protoner [29].

Undersøkelsen

En PET-skanner består av en trommel, en bevegbare seng og en datastasjon for innsamling av rådata og rekonstruksjon av bilder [29]. Etter at det radioaktive sporingsstoffet har blitt injisert emitterer det positroner fra atomkjernene, som kolliderer med elektroner i vevet, og annihilerer [35]. Fordi sporingsstoffet har hopet seg opp ved kreftcellene vil annihileringene skje nær svulsten, og det er fotonene som dannes her som danner grunnlaget for PET-avbildningen.

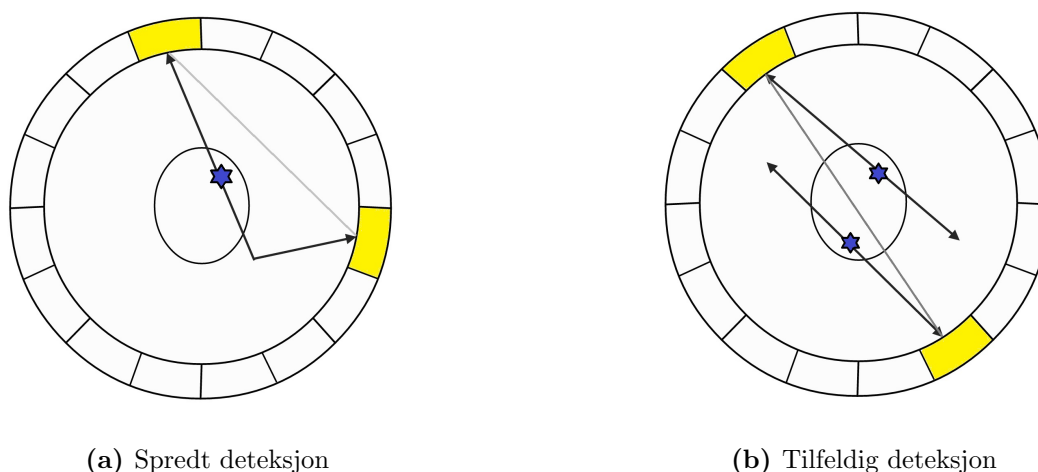
En times tid etter at sporingsstoffet har blitt injisert plasseres pasienten på sengen og føres inn i trommelen [36]. I trommelen er det flere detektorringer, som er bygget opp av kjeder av detektorer [29]. Alle kommersielle PET-maskiner bruker uorganiske scintillatormaterialer som strålingsdetektorer, oftest i form av en krystall [43]. En scintillatorkrystall konverterer høyenergetisk stråling, som gammastråling, til synlig lys [44]. En fotomultiplikator koblet til scintillatorkrystallen omgjør lysglimt til en elektrisk puls, og denne detekteres og prosesseres av den tilkoblede datamaskinen [29], [44].

Hvis to av detektorene rundt pasienten detekterer hvert sitt foton i det annihilerte paret innen et gitt tidsintervall, vil annihileringspunktet ligge på linjen mellom detektorene [35]. Denne linjen kalles *line of response* (LOR) og er illustrert i figur 2.3.3 [43]. Annihileringspunktet trenger ikke befinne seg midt på LOR-en, men forskjellen i ankomsttid for de to fotonene, kalt *time-of-flight* (TOF)-forskjellen gjør det mulig å bestemme riktig kollisjonspunkt mellom positron og elektron [29], [43]. Dersom det ene fotonet treffer detektorringen tidligere enn det andre, vil annihileringen ha skjedd nærmere detektoren som først detekterte fotonet.



Figur 2.3.3: Prinsippet bak en PET-avbildning. Fotonparet (representert av hver sin pil) treffer hver sin detektor innenfor et tidsintervall for å registrere et signal. Stjernen markerer annihileringspunktet. Adaptert fra [29].

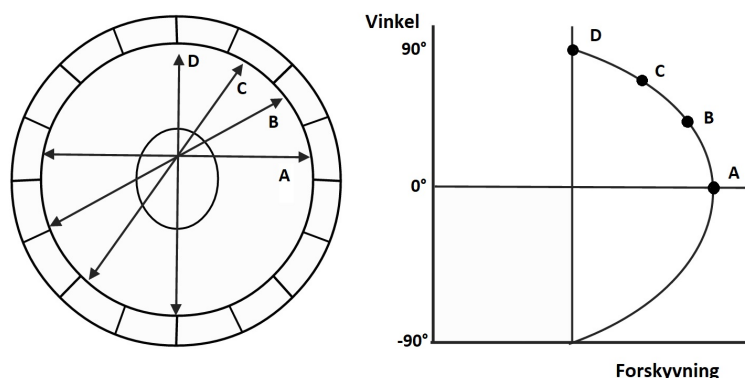
I tillegg til å registrere de ønskede innkommende fotonene vil detektorene også registrere spredte fotoner og tilfeldige sammentreff fra ulike annihileringer [29]. Spredte fotoner er fotoner som har blitt avbøyd etter annihilering, og dermed vil ikke den registrerte LOR-en gå igjennom annihileringspunktet. Spredte sammentreff relaterer i de fleste tilfeller til comptoneffekten [29]. Tilfeldige sammentreff dukker opp når to fotoner som stammer fra ulike annihileringer tilfeldigvis treffer detektoren samtidig og danner en falsk LOR [29], [45]. Tilfeldige og spredte deteksjoner vises i figur 2.3.4.



Figur 2.3.4: Uriktige registreringer som resultat av spredte gammastråler (piler) eller tilfeldige deteksjoner. Stjernene markerer annihileringspunktene, og de grå linjene mellom detektorene viser de falske LOR-ene. Adaptert fra [45].

Fra signal til bilde

Deteksjonene lagres i polarkoordinater i et sinogram, som viser forholdet mellom vinkel og forskyvning [29]. Vinkelen angir deteksjonens orientering i forhold til annihileringspunktet, mens forskyvningen er avstanden fra den horisontale linjen som skjærer gjennom samme punkt. Sinogrammet er dermed en kurve som representerer antall registrerte hendelser for hver LOR, vist i figur 2.3.5. For at hendelsen skal registreres er det viktig at begge fotonene når hver sin detektor innenfor et forhåndsbestemt tidsintervall [29], [46]. Fra dette sinogrammet er det mulig å lage en representasjonsmatrise som representerer de ulike scintillatorkrystallene og gir koordinater til annihileringspunktet [29]. Jo flere linjer som går gjennom et annihileringspunkt, jo lettere er det å bestemme koordinatene nøyaktig.



Figur 2.3.5: Deteksjonene i bildet til venstre danner grunnlaget for sinogrammet til høyre. Adaptert fra [47].

For å skille de ønskede deteksjonene fra spredte og tilfeldige deteksjoner med lavere energi, benyttes et energivindu med senter på 511 keV [29]. De spredte fotonene med energi under denne terskelverdien blir da luket ut. En annen metode for å luke ut uønskede deteksjoner er ved dempningskorreksjon.

Dempningskorreksjon

Dempningskorreksjon er nødvendig for å oppnå høyere presisjon i PET-avbildningen. PET-bilder kan korrigeres ved hjelp av CT-bilder, som gir informasjon om hvordan gammastråling dempes i ulike typer vev [46], [48]. En andel av fotonene absorberes i vevet og når aldri fram til detektorene, og informasjonen i CT-bildet brukes til å beregne sannsynligheten for at et foton absorberes [48]. Dersom det ikke gjøres en dempningskorreksjon, vil objekter nærmere overflaten vises sterkere enn objekter lenger inn i vevet, da det er mindre sannsynlig at fotonene herifra absorberes [48]. Ulike typer vev har også ulik absorpsjonsevne [46].

Et dempningskart lages fra CT-bildet, og legges til PET-bildet [29], [48]. Det er viktig at PET- og CT-bildet blir tatt i nøyaktig samme posisjon, som er optimalt med de kombinerte PET- og CT-skannerene som hovedsakelig er i bruk i dag [46]. Dempningskorreksjon er også viktig i beregningen av standard uptake value [48].

Standard Uptake Value

Standard uptake value (SUV) er en dimensjonsløs verdi som brukes for å bestemme aktiviteten i PET-bilder [29]. SUV gis ved

$$SUV = \frac{C(t)}{ID \cdot BW} \quad (2.5)$$

hvor $C(t)$ er konsentrasjonen av radioaktivitet målt i kBq/mL ved tiden t , ID målt i MBq er den injiserte dosen og BW er pasientens kroppsvekt [29]. Intensitetsverdiene i det ferdige PET-bildet representerer de ulike SUV-verdiene i hver voksel (volumerisk piksel).

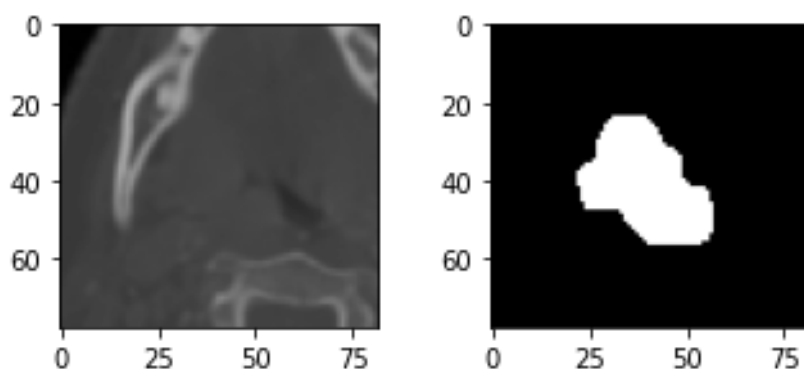
2.3.2 Computertomografi

Computertomografi (CT) bruker røntgenstråling til avbildning og viser et tverrsnitt av et område i kroppen, eksempelvis et organ [49]. Bildene kan stables oppå hverandre for å danne et tredimensjonalt bilde av interesseområdet. Fordelen med CT-bilder er at de inneholder mer informasjon enn konvensjonelle røntgenbilder, som gjør det enklere å identifisere og lokalisere svulster [50].

CT-bilder gjør det mulig å skille skjelett, organer og eventuelle abnormaliteter som skal identifiseres [21]. En fordel med CT-bilder er at det er mulig å rotere 3D-bildet slik at man kan betrakte et interesseområde fra alle vinkler [21]. I tillegg til ved diagnostisering, benyttes CT-avbildning også under og etter behandling, for å vurdere behandlingens effekt [49]. CT-bilder har en romlig oppløsning helt nede i 0.5 – 0.675 mm [51]. En CT-avbildning med tilhørende maske som markerer svulsten vises i figur 2.3.6.

Undersøkelsen

En CT-maskin består av samme komponenter som en PET-skanner, og pasienten plasseres også her på en bevegbare seng som sakte føres gjennom en trommel [50]. Et røntgenrør roterer rundt pasienten, og bestråler pasienten med en smal røntgenstråle. CT baserer seg på å måle hvor stor del av røntgenstrålene som dempes i ulike vev. Årsakene til dempingen er hovedsakelig fotoelektrisk effekt, comptonspredning, koherent spredning og pardannelse [21]. Summen av disse fire



Figur 2.3.6: Bildet til venstre viser et tverrsnitt av CT-avbildningen av pasient nummer 40 fra datasettet brukt i denne oppgaven. Til høyre vises den tilhørende masken som markerer hvor svulsten befinner seg

bidragsyterene bestemmer den totale dempingen av røntgenstråler i det penetrerte objektet. Den totale dempningskoeffisienten gis ved

$$\mu = \mu_{pe} + \mu_{compt} + \mu_{ko} + \mu_{pp} \quad (2.6)$$

der μ_{pe} , μ_{compt} , μ_{ko} og μ_{pp} er dempningskoeffisienten forårsaket av henholdsvis fotoelektrisk effekt, comptonspredning, koherent spredning og pardannelse [21]. Bidraget fra hver enkelt bidragsyter avhenger av energien til røntgenstrålene og egenskapene til de ulike vevstypene.

Røntgenstråler mister energi eksponensielt, og for et homogent legeme kan intensiteten til strålen uttrykkes ved Beer-Lamberts lov,

$$I(x) = I_0 e^{-\mu x} \quad (2.7)$$

hvor I_0 er den innkommende strålens intensitet, μ er dempningskoeffisienten og x er avstanden strålen har reist gjennom materie [21]. For å ta høyde for den varierende dempningskoeffisienten til ulikt vev skrives likningen til følgende:

$$I(L) = I_0 e^{-\int_0^L \mu(x) dx} \quad (2.8)$$

som tar høyde for linjeintegralet langs veien L fotonet beveger seg langs, og at denne veien består av ulike typer vev [21].

Dempingen måles ved at de uttgående røntgenstrålene registreres av en røntgenstråledetektor plassert på motsatt side av pasienten [21], [50]. Detektoren, som er laget av et scintillatormateriale, konverterer røntgenstrålen til en lyspuls, som

videre konverteres til et elektrisk signal [21], [52]. Hver gang strålekilden har rotert en runde rundt pasienten, vil en datamaskin rekonstruere et todimensjonalt bilde ut i fra de registrerte utgående røntgenstrålene [50]. Deretter vil pasienten føres videre inn i trommelen, der neste bilde vil bli tatt på samme måte. Tykkelsen på vevet som representeres i hver bildeskive varierer fra en til ti millimeter [50].

Kontrastvæske

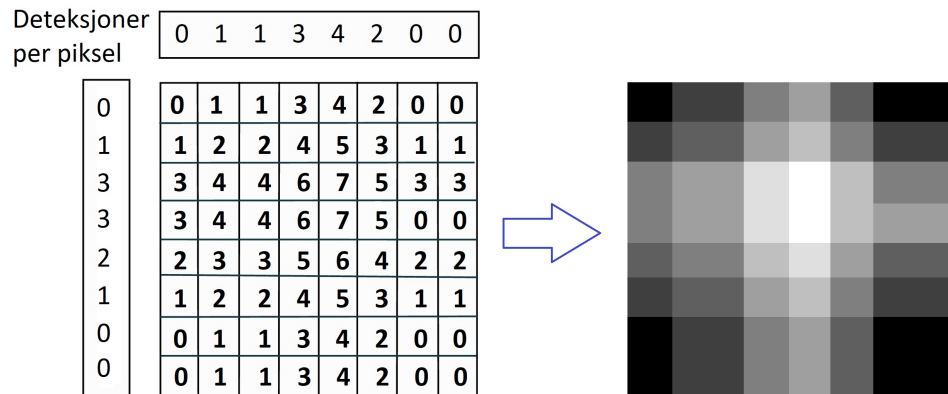
I mange tilfeller gis pasienten kontrastvæske intravnt før undersøkelsen [49]. Tette strukturer, som bein, er lett å avbilde, mens mykere vev har varierende dempningssegenskaper og kan dermed oppleves som vanskelige å se [50]. Kontrastvæsken inneholder stoffer som demper røntgenstråler i større grad, og vil gjøre det myke vevet mer synlig i et CT-bilde.

Fra signal til bilde

Meningen med rekonstruksjonen er å tilegne rett verdi av μ til hver voksel kun ved hjelp av de dempede intensitetsverdiene oppfattet av detektoren [21]. Resultatet blir en serie todimensjonale bilder der hver piksel representerer dempningskoeffisienten i det gitte punktet.

I CT-avbildning er *filtered back projection* (FBP) den vanligste rekonstruksjonsteknikken [21]. I likhet med PET-avbildning lagres signalene fra CT-avbildningen i et sinogram [52]. FBP utføres ved at dempningsprofilen fra hver bestrålingsvinkel projiseres tilbake over bildet. Dempningsverdien fra sinogrammet deles på antall piksler langs retningen for projeksjonen, og den gjennomsnittlige dempningsverdien gis til disse pikslene [52]. Dette repeteres for hver vinkel, og det endelige bildet er summen av alle de tilbakeprojiserte dempningsprofilene. FBP er illustrert i figur 2.3.7.

FBP resulterer i et veldig utydelig bilde, som motvirkes ved at et frekvensfilter legges til dempningsprofilene før tilbakepropageringen [52]. Filteret blokkerer ut lave frekvenskomponenter hos dempningsprofilene, slik at områdene med raske overganger, for eksempel mellom ben og bløtvev, blir tydeligere [52]. Dette reduserer uskarphet, men fører også til mer støy i bildet.



Figur 2.3.7: Dempningsprofilen fra hver bestrålingsvinkel projiseres tilbake over bildet og danner matrisen som gjøres om til et CT-bilde.

CT-bildet er et todimensjonalt dempningskart av et tverrsnitt av pasienten. Hvert punkt på kartet gis av en lineær dempningskoeffisient relativ til vann, som defineres som en *Hounsfield unit* (HU), gitt ved

$$HU = \frac{\mu_{materiale} - \mu_{vann}}{\mu_{vann}} \quad (2.9)$$

der $\mu_{materiale}$ og μ_{vann} er dempningskoeffisientene til henholdsvis det aktuelle punktet i vevet og vann [52].

2.4 Maskinlæring

Maskinlæring er et felt innenfor kunstig intelligens som gjennom selvlærende algoritmer henter ut informasjon fra data [9]. Denne informasjonen kan videre brukes til å lage modeller som utfører prediksjoner for usett data eller avdekker mønster i dataen. Videre er det ønskelig å gradvis forbedre ytelsen til modellene for å ta avgjørelser basert på prediksjonene. Et av målene med maskinlæring er å effektivt utføre serier av matematiske operasjoner som vil være ekstremt tidkrevende med menneskelige ressurser.

I denne oppgaven brukes maskinlæring til å predikere behandlingsutfall for pasienter med hode- og halskreft.

2.4.1 Ulike typer maskinlæring

Maskinlæring kan deles inn i tre hovedtyper; overvåket læring, forsterkende læring og ikke-overvåket læring [9]. I overvåket læring brukes data som allerede har blitt tildelt en merkelapp til å trene en modell [9]. Modellen mates deretter med umerket usett data, og skal predikere merkelappen til denne. Merkelappene kan representere både binære klasser, multiple klasser eller regresjonsverdier.

I forsterkende læring er målet å utvikle et system som forbedrer ytelsen sin basert på hvordan det vekselvirker med omgivelsene [9]. Ikke-overvåket læring går ut på å avdekke gjemte strukturer i umerket data [9]. I denne oppgaven brukes kun overvåkede læringsalgoritmer.

2.4.2 Klassifiseringsalgoritmer

En klassifiseringsalgoritme brukes for å tildele en klasse til umerket testdata bestående av en serie prøver med tilhørende egenskaper [9]. Algoritmene trenes med data som allerede har fått tildelt klasser, og målet er at klassifiseringsalgoritmen etter trening skal kunne motta ny data og korrekt tildele klasse basert på det den har lært fra treningsdataen. Responsvariabelen, y , er kategorisk, og kan uttrykkes som en funksjon av en matrise X . Datamatriksen brukt til trening kalles ofte X_{train} , med tilhørende responsvariabler y_{train} , som regel en vektor, og datamatriksen brukt til testing, og som mates inn i den trente modellen, kalles X_{test} . De ulike datamatriksene som brukes i trening og testing er illustrert i figur 2.4.1. Algoritmen ønsker å tildele y -verdier tilhørende X_{test} basert på forholdet mellom X_{train} og y_{train} [9].

Lineære klassifiseringsalgoritmer utfører klassifiseringen basert på en lineær kombinasjon av prøvenes egenskaper kombinert med vektorer tildelt hver egenskap [53]. For data som ikke er lineært separabel kan en ikke-lineær klassifiseringsalgoritme kartlegge hvert tilfelle over til en vektor av høyere dimensjonalitet der dataen kan separeres lineært [53]. De lineære klassifiseringsalgoritmene brukt i denne oppgaven er logistisk regresjon og lineære *support vector machines*. I tillegg brukes de ikke-lineære klassifiseringsalgoritmene *random forest*, *k-nærmeste naboer*, ikke-lineære *support vector machines*, *adaptive boosting*, *kvadratisk diskriminantanalyse*, *passive aggressive* og *light gradient boosting machine*.

Treningsdata				
Tilfelle	Egenskap 1	Egenskap 2	Egenskap 3	Klasse
1	0
2	1
3	1
4	0
5	1

Testdata			
Tilfelle	Egenskap 1	Egenskap 2	Egenskap 3
6
7
8

Figur 2.4.1: En datamatrix for trening oppdelt i X_{train} som inneholder egenskapene til prøvene og y_{train} som viser den tilhørende klassen. En modell trent på denne dataen skal predikere y -verdier tilhørende datamatrixen for testing av modellen, X_{test} .

Klassifiseringsalgoritmene kan brukes til både binære- og flerklasseproblemer, og implementeres gjennom ulike tilleggspakker. I denne oppgaven utføres kun binær klassifisering, som vil si at prøvene kun kan tildeles to ulike klasser.

Logistisk regresjon

Til tross for navnet er logistisk regresjon en lineær klassifiseringsalgoritme der sannsynligheten for de ulike utfallene til en binær problemstilling modelleres ved hjelp av en logistisk funksjon [54]. Oddsratioen, det vil si oddsen til fordel for en hendelse, gis av

$$OR = \frac{p}{1-p}$$

der p er sannsynligheten for positivt utfall [9]. Fra likning 2.4.2 kommer logit-funksjonen, som er logaritmen til oddsratioen,

$$\text{logit} = \ln\left(\frac{p}{1-p}\right) \quad (2.10)$$

som tar verdier mellom 0 og 1 og transformerer de til verdier som dekker hele nummerspekteret [9]. For å predikere sannsynligheten for at en ny prøve hører til

en bestemt klasse, brukes den inverse formen av likning 2.10,

$$\Phi(z) = \frac{1}{1 + e^{-z}} \quad (2.11)$$

der z er de ulike prøvene matet inn i modellen [9]. Utdataen fra likning 2.11 er en verdi mellom 0 og 1, og gir sannsynligheten for at prøven tilhører klasse 1. En terskelfunksjon vil deretter tildele prøven klasse 1 dersom $\phi(z)$ er større enn terskelverdien, og 0 ellers [9].

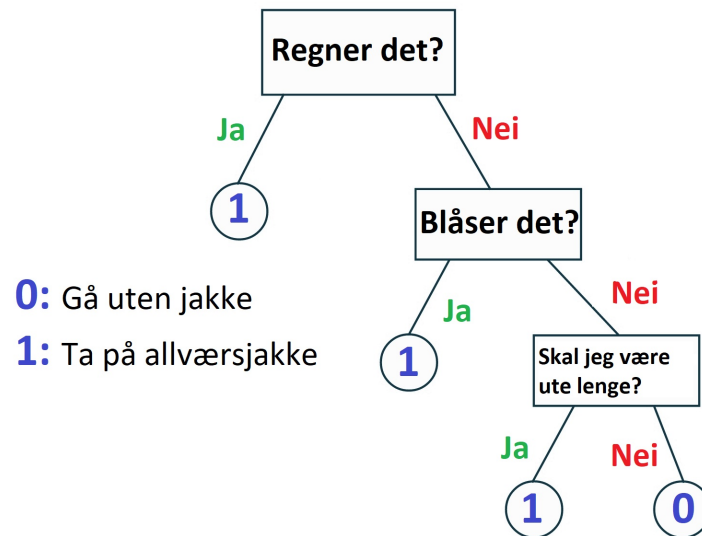
I denne oppgaven brukes pakken LogisticRegression fra scikit-learn [55].

Random Forest

Random Forest-klassifiseringsalgoritmen er en samling av bestemmelsestrær. Et bestemmelsestre bryter opp dataen ved å stille en serie spørsmål lært fra treningssettet [9]. Dette kan være spørsmål som: «Har pasienten røkt mer enn 20 sigaretter daglig i mer enn 10 år?» eller «Er diameteren til svulsten større enn 5 cm?». I den første delingen splittes dataene ved egenskapen som resulterer i mest innhentet informasjon (IG fra information gain), og treet får to nye grener [9]. Prosessen kan repeteres helt til alle prøvene i det ytterste leddet på hver gren, bladene, tilhører samme klasse [9]. Ved å splitte dataen helt til alle bladene er rene kan modellen lett bli overtilpasset, som forklares i seksjon 2.4.3 [9]. Derfor er det vanlig å stanse algoritmen ved en maksimal dybde, det vil si etter et gitt antall splittelser fra utgangspunktet. Et eksempel på et bestemmelsestre er vist i figur 2.4.2.

Random Forest er et ensemble av beslutningstrær, som skal gjøre modellen mer generaliserende enn et individuelt tre [9]. Random forest-algoritmen kan oppsummeres med følgende fire steg [9]:

1. Tilfeldig trekke et utvalg av n prøver med tilbakelegging fra treningssettet.
2. Gro et bestemmelsestre fra utvalget. Velg d antall egenskaper fra utvalget uten tilbakelegging, og splitte grenen ved egenskapen som gir høyest IG.
3. Repetere steg 1 og 2 for hvert av k trær.
4. Bruke prediksjonen fra hvert tre til å angi klasse til nye prøver, ved at prøven tildeles klassen majoriteten av trærne har predikert.



Figur 2.4.2: Et bestemmelsestre som avgjør om en person skal ta på allværsjakke eller gå uten, basert på vær og tid utendørs. Bladene er markert med 0 og 1, og avgjør om prøven, som her er en person, skal ta på jakke eller ikke.

I denne oppgaven brukes pakken `RandomForestClassifier` fra `scikit-learn` [55].

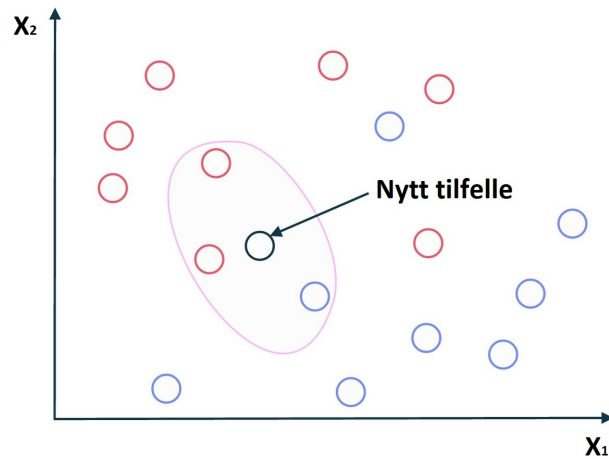
K-nærmeste naboer

K-nærmeste naboer (KNN), er det som kalles en lat klassifiseringsalgoritme [9]. Det vil si at i stedet for å lære en funksjon fra treningsdaten, memorerer KNN alle datapunktene, og tildeler klasser til testdataen basert på hukommelsen. KNN klassifiserer nye prøver basert på klassen til de nærmeste naboene sine, det vil si de prøvene fra datamatriksen brukt til trening som ligger nærmest den nye prøven.

KNN-algoritmen består av 3 steg [9]:

1. Velge et nummer for k antall naboer, og bestemme hvordan avstanden skal måles, for eksempel i luftlinje mellom punktene - euklidisk, eller summen av avstanden i x og y retning - Manhattan-avstanden.
2. Finne de k -nærmeste naboene til den nye prøven som skal klassifiseres.
3. Tildel klasse til den nye prøven basert på den dominerende klassen blant de k nærmeste naboene.

Valget av k er viktig for ytelsen til modellen, og hvilken verdi som er optimal vil variere for hvert enkelt datasett [9]. Lave verdier av k vil tillate mer støy i modellen, mens høye verdier av k krever mye kapasitet. En klassifisering etter k -nærmeste naboer-prinsippet vises i figur 2.4.3.

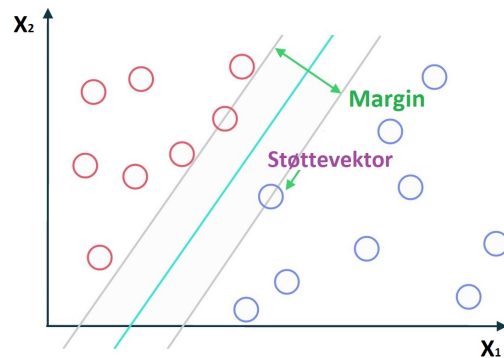


Figur 2.4.3: KNN utført for en ny prøve basert på de tre nærmeste naboene. I dette eksempelet vil den nye prøven tildeles den røde klassen. Adaptert fra [9].

I denne oppgaven brukes pakken `KNeighbors` fra `scikit-learn` [55].

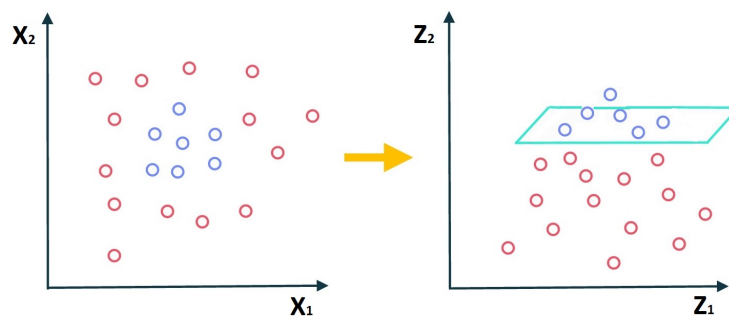
Support Vector Machines

Support vector machines (SVM), er en serie algoritmer som blant annet kan brukes til klassifisering, regresjon og deteksjon av ekstremverdier [56]. Målet med SVM er å maksimere avstanden mellom grensen som skiller klassene og treningsprøvene som ligger nærmest denne grensen [9]. Denne avstanden kalles for marginen, og de nærmeste prøvene kalles støttevektorer. Med andre ord ønsker man å sette opp klassifiseringsgrensen slik at den ligger lengst mulig unna de nærmeste datapunktene. Dette illustreres i figur 2.4.4.



Figur 2.4.4: Klassifisering ved hjelp av SVM algoritmen. Marginen er avstanden mellom de nærmeste datapunktene på hver side av klassifiseringsgrensen, markert som lyseblå linje.

SVM kan også brukes til å løse ikke-lineære problemer, ved å bruke ulike kjernefunksjoner som sørger for en ikke-lineær grense mellom klassene. Disse kjernefunksjonene projiserer datapunktene over i et rom der de blir lineært separable, som illustrert i figur 2.4.5 [9].



Figur 2.4.5: Et ikke-lineært separabelt datasett projiseres over i et plan av høyere dimensjon, hvor det kan separeres av en flate, her markert i lyseblått. Adaptert fra [9].

I denne oppgaven brukes pakken `svm.SVC` fra `scikit-learn` [55].

Boosting

Boosting består av et ensemble av klassifiseringsalgoritmer som så vidt yter bedre enn tilfeldig gjetning [9]. Disse skal lære fra misklassifisert trening. Målet er å

fokusere på prøvene som ofte klassifiseres feil, og gjennom dette gradvis bedre ytelsen til ensemblet. Følgende fire steg gir en oppsummering av boosting-prinsippet [9]:

1. Trekk et tilfeldig sett av treningsprøver d_1 uten tilbakelegging for å trene en klassifiseringsalgoritme C_1 .
2. Trekk et nytt tilfeldig sett av treningsprøver d_2 uten tilbakelegging, legg til halvparten av prøvene som ble feilklassifisert i steg en, og tren en klassifiseringsalgoritme C_2 .
3. Tren en tredje klassifiseringsalgoritme C_3 med prøvene der C_1 og C_2 er uenige.
4. Tildel prøvene en klasse via flertallsstemmegivning fra C_1 , C_2 og C_3 .

En variant av boosting-algoritmen er *adaptive boosting*, AdaBoost, der hele treningssettet brukes på å trene klassifiseringsalgoritmene, men prøvene tildeles nye vekter i hver runde for å forsterke klassifiseringsalgoritmen [9]. I den første boosting-runden forsøker AdaBoost å trekke en bestemmelsesgrense mellom klassene. Prøvene som havner på feil side av grensen tildeles en større vekt i neste boosting-runde, mens de korrekt klassifiserte prøvene tildeles et lavere vekttall. Runde 2 vil dermed fokusere mer på prøvene som ble feilklassifisert i første runde. Slik fortsetter prosessen i j antall runder, før hver prøve tildeles en klasse basert på flertallsstemmegivning fra klassifiseringsalgoritmene i hver av de j boosting-rundene [9].

I denne oppgaven brukes pakken AdaBoostClassifier fra scikit-learn [55].

En annen boosting-variant er *light gradient boosting machine*, LightGBM, som benytter seg av bestemmelsestrær for klassifisering [57]. LightGBM-algoritmen er histogrambasert, og setter kontinuerlige egenskaper i diskrete bokser for å få fortgang i treningen og redusere minnebruk. I stedet for å gro bestemmelsestrærne nivåvis, gror LightGBM ett og ett blad av gangen, og nye trær legges til ensemblet en etter en for å forsøke å klassifisere prøvene som tidligere har blitt feilklassifiserte [57].

I denne oppgaven brukes pakken LightGBM utviklet av Microsoft [58].

Kvadratisk diskriminantanalyse

Kvadratisk diskriminantanalyse (QDA) er en klassifiseringsalgoritme som antar at hver klasse følger en normalfordelt distribusjon [59]. Fordelen med QDA sam-

menliknet med lineær diskriminantanalyse (LDA) er at QDA kan sette opp en kvadratisk bestemmelsesgrense for å skille klassene, mens LDA kun lager lineære grenser [60]. QDA utledes fra en enkel sannsynlighetsbasert modell, og prediksjoner gjøres basert på Bayes regel,

$$P(y = k | x) = \frac{P(x | y = k)P(y = k)}{P(x)} \quad (2.12)$$

for hver treningsprøve x , der k representerer klassen [60]. $P(y = k | x)$ er sannsynligheten for at $y = k$ for en bestemt prøve x . Den klassespesifikke sannsynligheten for prøven, $P(x | y = k)$ er andelen av prøver fra treningssettet som tilhører klasse k . $P(y = k)$ angir sannsynligheten for at prøven predikeres som klasse k , mens $P(x)$ gir sannsynligheten for at prøven x faktisk tilhører klassen. Logaritmen til likning 2.12 gis ved

$$\ln P(y = k | x) = \ln P(x | y = k) + \ln P(y = k) + C_{st} \quad (2.13)$$

hvor C_{st} er en konstant som korresponderer til nevneren i likning 2.12 [60]. Den nye prøven tildeles klassen k som maksimerer uttrykket i likning 2.13.

I denne oppgaven brukes pakken `qda.QDA` fra `scikit-learn` [55].

Passive aggressive

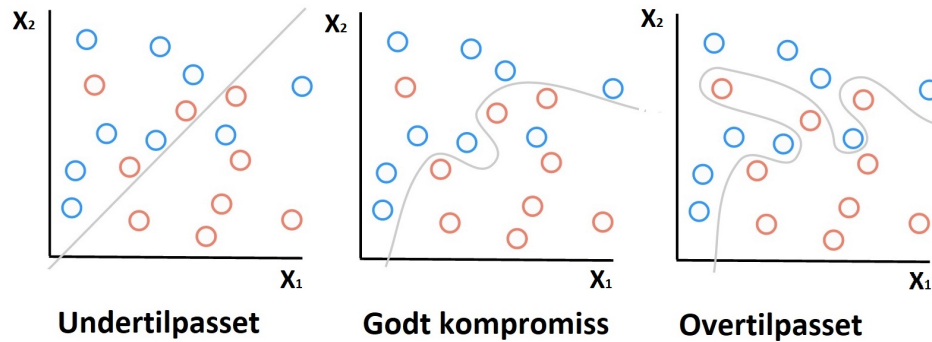
Passive aggressive er en *online* klassifiseringsalgoritme. *Online* læring er en form for maskinlæring som går utenfor hovedtypene, der algoritmen lærer fra hver prøve i datasettet brukt til trening underveis heller enn fra hele datasettet på en gang [9]. Prinsippet bak passive aggressive-algoritmen er å respondere passivt til korrekte klassifiseringer og aggressivt til gale klassifiseringer [61]. Jo lengre unna bestemmelsesgrensen en feilklassifisert prøve befinner seg, jo sterkere vil algoritmen reagere.

I denne oppgaven brukes pakken `PassiveAggressiveClassifier` fra `scikit-learn` [55].

2.4.3 Tilpasning av modeller

Ofte kan en modell yte bra på treningsdataen, mens når den blir testet på usett data faller andelen korrekte prediksjoner kraftig. Dette kalles for overtilpasning, og betyr at modellen har høy varians som følge av for mange parametere som gjør

modellen for kompleks [9]. Modellen kan også være utsatt for undertilpasning, som innebærer at modellen ikke er kompleks nok til å fange mønstrene i treningsdataen, og dermed også yter dårlig på usett data. Over og undertilpasning er illustrert i fig 2.4.6.



Figur 2.4.6: Modellen til venstre viser en lineær klassifiseringsalgoritme der en ganske stor andel prøver har blitt feilklassifisert. Modellen til høyre har trukket bestemmelsesgrensen slik at alle prøvene i datamatriksen brukt til trening havner på riktig side. Denne modellen er lite generaliserende. Modellen i midten viser et kompromiss mellom de to andre modellene, der prediksjonen på både treningsdataen og testdataen er ivaretatt. Adaptert fra [9].

De ulike klassifiseringsalgoritmene har ulike hyperparametere som kan justeres for å tilpasse modellene [9]. En mye brukt hyperparameter for å takle overtilpasning er regulariseringsparameteren C , som brukes i logistisk regresjon, SVM og *Passive Aggressive*.

Andre eksempler på hyperparametere som kan tilpasses er antall naboer i KNN, tredybde og antall trær i *random forest* og kjernefunksjon i SVM. En oversikt over parameterene som tilpasses for de ulike klassifiseringsalgoritmene finnes i tillegg A.

Regularisering

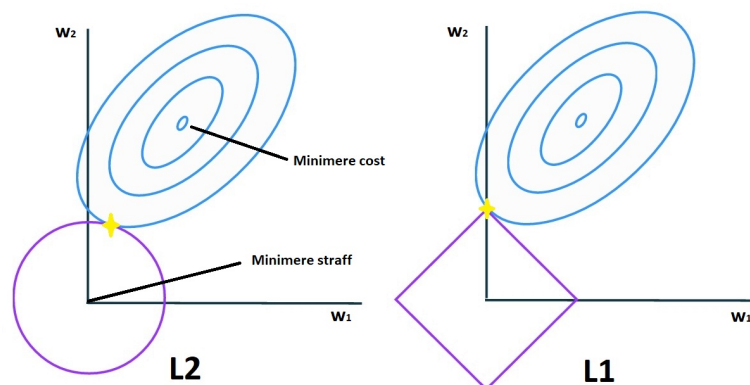
Regularisering skal ta hensyn til korrelasjon mellom egenskaper, filtrere ut støy og takle overtilpasning [9]. Konseptet går ut på å introdusere tilleggsinformasjon som straffer ekstreme vektverdier under klassiferingen. Den vanligste formen for regularisering er L2-regularisering, som kan uttrykkes gjennom

$$L2 = \frac{\lambda}{2} \|w\|^2 \quad (2.14)$$

hvor λ er regulariseringsparameteren og den inverse av C , og w er vektene som brukes i oppdateringen av klassifiseringsalgoritmen [9]. En annen metode for regularisering er L1-regularisering. Her uttrykkes tilleggsinformasjonen som

$$L1 = \frac{\lambda}{2} \|w\| \quad (2.15)$$

L1-regularisering fører til at mange av vektene som hører til egenskapene settes til null, og fungerer på den måten også som egenskapsutvelger for å redusere modellkompleksitet [9]. L1- og L2-leddene legges til som straffeparametere for å redusere *cost*-funksjonen. En *cost*-funksjon benyttes for å evaluere ytelsen til algoritmen, og bruker de sanne og predikerte verdiene av responsvariabelen til å regne ut hvor mye feil algoritmen tok i prediksjonen [9]. Det er dermed ønskelig å minimere *cost*-funksjonen, slik at antall feil reduseres. Løsningen som minimerer *cost*-funksjonen vil være en verdi mellom 0 og 1 for L2, mens for L1 vil den være enten 0 eller 1 [9]. Dette er illustrert i figur 2.4.7.



Figur 2.4.7: L2-regularisering til venstre, og L1-regularisering til høyre. Punktene markert i gult viser hvilke verdier av vekten w som minimerer *cost*-funksjonen.

2.4.4 Validering av modeller

For å måle modellene opp mot hverandre brukes en felles metrikk som vurderer prediksjonen opp mot den ekte sannheten. For å kunne gjøre dette er det nødvendig

med et valideringssett som benyttes til testing, der klassene allerede er kjente. En enkel metode for å teste ytelsen til modeller er ved bruk av nøyaktighet, som ganske enkelt tildeler poeng dersom den predikerte klassifiseringen stemmer overens med den sanne klassifiseringen, før den til slutt deler på antall oppnåelige poeng [9]. Med andre ord er nøyaktighet et mål på andelen korrekte klassifiseringer. Ulempen med nøyaktighet er at dersom klassedistribusjonen ikke er balansert, vil det ikke nødvendigvis gis et korrekt bilde av modellytelsen. Eksempelvis, hvis 90% av prøvene tilhører klasse 0, og modellen tildeler samtlige prøver klasse 0 vil modellen ha en nøyaktighet på 90% selv om den har feilpredikert en hel klasse.

De predikerte resultatene kan deles inn i fire grupper; sann positiv prediksjon (TP fra *true positive*), sann negativ prediksjon (TN fra *true negative*), falsk positiv prediksjon (FP) og falsk negativ prediksjon (FN) [9]. Disse fire gruppene er illustrert i figur 2.4.8.

		Predikert verdi	
		Positiv	Negativ
Faktisk verdi	Positiv	TP	FN
	Negativ	FP	TN

Figur 2.4.8: Matrise som viser de predikerte sanne og falske resultatene. Adaptert fra [9].

Fra disse gruppene kan nøyaktighet defineres ved

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.16)$$

Receiver operating characteristics area under curve

Receiver operator characteristics- (ROC)-kurve er en graf som visualiserer ytelsen til en modell ved å se på forholdet mellom sann positiv rate (TPR fra *true positive rate*) og falsk positiv rate (FPR) [9]. TPR og FPR gis av henholdsvis

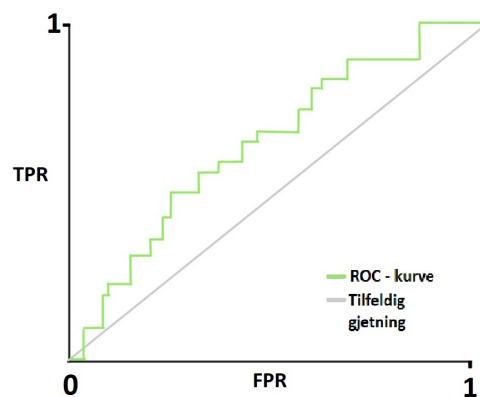
$$TPR = \frac{TP}{FN + TP} \quad (2.17)$$

og

$$FPR = \frac{FP}{FP + TN} \quad (2.18)$$

[9]. TPR forteller hvor stor del av de faktisk positive prøvene som klassifiseres riktig, mens FPR fortelle hvor stor andel av de faktisk negative prøvene som klassifiseres som positive. TPR og FPR har alltid verdier mellom 0 og 1, og summen av de to er alltid 1.

TPR og FPR beregnes ved å flytte bestemmelsesterskelen til klassifiseringsalgoritmen, og ulike sammenhenger mellom TPR og FPR plottes i et koordinatsystem med TPR på x-aksen og FPR på y-aksen [9]. ROC-kurven strekker seg fra punkt (0,0) til punkt (1,1) og kobler sammen de plottede datapunktene. Den rette linjen mellom (0,0) og (1,1) viser en modellytelse tilsvarende tilfeldig gjetning, det vil si at TPR er lik FPR. Et hvert punkt på denne linjen tilsvarer at for alle korrekt klassifiserte prøver i klasse 1, finnes det like mange feilklassifiserte prøver i klasse 0. Alle kurver som ligger over denne linjen vil ha en bedre ytelse enn tilfeldig gjetning, mens alle under vil yte dårligere. En perfekt klassifisering algoritme vil ligge helt øverst i venstre hjørne. En vilkårlig ROC-kurve er illustrert i figur 2.4.9.



Figur 2.4.9: Den grønne kurven viser en vilkårlig ROC-kurve. ROC AUC er arealet mellom denne og x-aksen. Den grå linjen tilsvarer en modell som yter på samme nivå som tilfeldig gjetning. Adaptert fra [9].

Basert på ROC-kurven regnes ROC *area under the curve* (ROC AUC) ut. Denne verdien mellom 0 og 1 er arealet som befinner seg under ROC-kurven, og kan brukes til å bedømme ytelsen til en klassifiseringsmodell [9].

I denne oppgaven brukes pakken `roc_auc_score` fra `scikit-learn` [55].

F1-score

Metrikkene *precision* (PRE) og *recall* (REC) er sterkt relaterte til TPR og FPR, og defineres av henholdsvis

$$PRE = \frac{TP}{TP + FP} \quad (2.19)$$

og

$$REC = TPR = \frac{TP}{FN + TP} \quad (2.20)$$

[9]. En kombinasjon av disse gir metrikken *F1-score*, definert som

$$F1 = 2 \frac{PRE \cdot REC}{PRE + REC} \quad (2.21)$$

[9]. En høy F1-score er derfor avhengig av at både PRE og REC er høye, det vil si at både andelen predikerte positive som faktisk er positive og andelen faktisk positive som bli korrekt klassifisert er høye.

I denne oppgaven brukes pakken `flscore` fra `scikit-learn` [55].

F1-score for motsatt klasse

I beregningen av F1-score står den positive klassen i fokus, og det er vanskelig å si noe om hvor nøyaktig modellen predikerer den negative klassen. For å få et bilde av dette er det mulig å snu om på klassene, det vil si at prøvene som har klasse 0 får tildelt klasse 1, og motsatt. Det samme gjøres for den predikerte responsen. Ved å sammenligne disse nye verdiene med F1-score for den positive klassen er det mulig å se om modellen er signifikant bedre til å predikere den ene klassen enn den andre.

Matthews korrelasjonskoeffisient

Matthews korrelasjonskoeffisient (MCC) ble utviklet spesielt som et alternativ til nøyaktighet for å ta høyde for ubalanserte klassedistribusjoner [62]. MCC gis av følgende likning:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad (2.22)$$

[62]. MCC kan inneha verdier mellom -1 og 1, og er den eneste binære klassifiseringsmetrikken som gir et høy resultat kun om den binære prediktoren klarer å predikere både majoriteten positive prøver og majoriteten negative prøver. En MCC med verdi 1 tilsier at alle prøvene ble korrekt klassifisert. En MCC med verdi 0 tilsier at modellen yter like godt som tilfeldig gjetning.

I denne oppgaven brukes pakken `matthewscorecoef` fra `scikit-learn` [55].

Kapittel 3

Materiale og metode

Dette kapittelet tar for seg materialet og fremgangsmåten benyttet i oppgaven. Koden til dette prosjektet er tilgjengelig fra <https://github.com/sofiefj/Bellossom>, der annet ikke er spesifisert.

3.1 Datasettet

Datasettet brukt til trening besto av klinisk informasjon og tredimensjonale PET- og CT-bilder med tilhørende PET-parametere fra pasienter som ble behandlet for hode- og halskreft ved Oslo universitetssykehus (OUS) i perioden 2007 til 2013 [63]. Datasettet inkluderte 197 pasienter, som tidligere har blitt beskrevet i Moan et al. [63] og Groendahl et al. [64]. Blant annet ble pasienter med kreft i nesesvelget (nasopharynx) ekskludert fra det originale datasettet, i tillegg til pasientene som av ulike årsaker ikke fikk kontrastforsterket CT [64].

Samtlige av pasientene mottok stråleterapi, og alle under 70 år mottok også cellegiftbehandling [63]. Dette er en retrospektiv studie, det vil si at pasientdataen ble hentet ut etter endt behandling. Pasientene ble delt inn i to ulike klasser, 0 og 1, for hver av tre ulike responser. Pasientene fulgte et femårig oppfølgingsprogram, og de som fortsatt var i live og/eller ikke fikk tilbakefall innen fem år etter endt behandling ble betraktet som friske og fikk tildelt klasse 0, selv om det er mulig at de har fått tilbakefall senere. Pasienter som fikk tilbakefall eller døde før oppfølgingsperioden var ferdig fikk tildelt klasse 1. Datasettet kunne også inneholde pasienter som har dødd av andre årsaker enn tilbakefall av kreft [63].

De tre ulike responsene modellene ble trent for var sykdomsfri overlevelse (DFS fra *disease free survival*), lokalregionalt tilbakefall (LRC fra *local-regional control*) og generell overlevelse (OS fra *overall survival*). Det vil si at for hver pasient var det tre ulike merkelapper, hvor klasse 1 representerte tilbakefall for DFS og LRC, og død for OS. Klassedistribusjonen til de ulike responsene vises i tabell 3.1.

Tabell 3.1: Oversikt over de ulike responsene og tilhørende klassedistribusjon for OUS-datasettet.

Respons	Beskrivelse	Distribusjon
DFS	1 hvis pasienten ikke overlevde, fikk lokalt eller regionalt tilbakefall eller metastatisk tilbakefall, 0 ellers	Klasse 1: 45,7% Klasse 0: 54,3%
LRC	1 hvis pasienten fikk lokalt eller regionalt tilbakefall, 0 ellers	Klasse 1: 24,4% Klasse 0: 75,6%
OS	1 hvis pasienten ikke overlevde, 0 ellers	Klasse 1: 38,6% Klasse 0: 61,4%

En oversikt over de kliniske egenskapene i datasettet finnes i tabell 3.2. Beskrivelsen av de ulike egenskapene er hentet fra Grøndahl [65] og er inndelt som i Moan et al. [63].

Bildene som ble brukt i denne studien var PET- og CT-bilder fra de 197 pasientene i datasettet. Intensitetsverdiene i PET-bildene representerte SUV-verdier, mens intensitetsverdiene i CT-bildene representerte HU. For hver pasient fantes det i tillegg en maske tilhørende hvert par av PET og CT-bilder som viste avgrensningen til primærsvulsten og eventuelle lymfeknuter med spredning, og definerte regionen av interesse (ROI fra *region of interest*). Masken var et binært bilde, der alle voksler med verdi 1 lå innenfor ROI, og vokslene med verdi 0 lå utenfor. Disse maskene var tegnet inn av en onkolog [63].

Tabell 3.2: Oversikt over de kliniske egenskapene i datasettet. μ viser gjennomsnittsverdien, og σ viser standardavviket til de kontinuerlige egenskapene. **uicc8_III – IV* mangler for de samme pasientene som *hvp_related*.

Egenskap	Type	Beskrivelse	Distribusjon
age	Kontinuerlig	Pasientens alder (år)	$\mu = 60.32$ $\sigma = 7.53$
female	Binær	1 hvis kvinne, 0 ellers	Kvinne: 24,9% Mann: 75,1%
cavum_oris	Binær	1 hvis primærsvulst er i munnhulen, 0 ellers	Munnhule: 6,6%
oropharynx	Binær	1 hvis primærsvulst er i munnsvelg, 0 ellers	Munnsvelg: 74,6%
hypopharynx	Binær	1 hvis primærsvulst er i strupesvelg, 0 ellers	Strupesvelg: 8,1%
larynx	Binær	1 hvis primærsvulst er i strupehodet, 0 ellers	Strupehode: 10,7%
histgrade_high	Binær	1 hvis histologisk grad av svulstvev er klasse G3, 0 ellers	G3: 69,5%
hvp_related	Binær	1 dersom pasienten har positiv HPV-status, 0 ved negativ	Positiv: 41,6% Negativ: 32,0% Ukjent: 26,4%
ecog	Binær	1 hvis ECOG-status er 1- 3, 0 hvis 0	1-3: 35,5%
charlson	Binær	1 hvis Charlson Comor- dibity Index (CII) er 1-6, 0 hvis CII er 0	CII 1-6: 34,5%
pack_years	Kontinuerlig	Antall år der pasienten har røyket minst 20 sigaretter daglig	$\mu = 25,7$ $\sigma = 24,2$
uicc8_III_IV	Binær	1 for kreftstadie 3-4, 0 for stadie 1-2	I-II: 37,1% III-IV: 36,5% Ukjent: 26,4%*

Bildene var reskalerte til en isotropisk vokselstørrelse på $1 \times 1 \times 1 \text{ mm}^3$ og registrert til et felles referansesystem [64]. Fra PET-, CT-bildene og de tilhørende maskene ble det trukket ut egenskaper som sammen med de kliniske egenskapene og utvalgte PET-parametere dannet datasettet som ble brukt til trening.

Intensitetsverdiene i PET-bildene var representerte som SUV-verdier, det vil si at de var normaliserte med hensyn på pasientens vekt [64]. Fra disse verdiene var det utledet tre parametere som tilhørte datasettet, *SUVpeak*, *MTV* (metabolsk tumorvolum) og *TLG* (total lesion glycosis). Tidligere studier har vist en sammenheng mellom disse PET-parametere og overlevelse, og disse kunne derfor være viktige egenskaper i prediksjonen, som vist i Moan et al. [63] hvor disse PET-parametere også er definert. En oversikt over PET-parametere brukt i denne oppgaven er gitt i tabell 3.3.

Tabell 3.3: Oversikt over PET-parametere i datasettet.

Parameter	Type	Beskrivelse	Distribusjon
SUVpeak	Kontinuerlig	Maksimal snittverdi av SUV i en sfære med volum 1 cm^3 , der senteret til sfæren må ligge innenfor ROI	$\mu = 10,81$ $\sigma = 5,25$
MTV	Kontinuerlig	Volum med $SUV \geq 0.5 \cdot SUV_{peak}$	$\mu = 11,31 \text{ cm}^3$ $\sigma = 12,16 \text{ cm}^3$
TLG	Kontinuerlig	Total Lesion Glycosis, $MTV \times SUV_{mean}$ (snittverdi av SUV)	$\mu = 111,32 \text{ cm}^3$ $\sigma = 170,47 \text{ cm}^3$

Datsett til testing

Datsettet brukt til testing var et eksternt datsett som opprinnelig besto av 183 pasienter behandlet for hode- og halskreft ved MAASTRO-klinikken i Nederland mellom 2008 og 2014. Datsettet ble redusert til 99 pasienter, da enkelte pasienter hadde ukomplett bildedata, ikke fikk kontrastforsterket CT eller manglet klinisk data. På samme måte som dataen fra OUS hadde bildene fra MAASTRO blitt reskalert til vokselstørrelser på $1 \times 1 \times 1 \text{ mm}^3$. Klassedistribusjonen til datsettet benyttet til testing vises i tabell 3.4.

Tabell 3.4: Oversikt over de ulike responsene og tilhørende klassedistribusjon for MAASTRO-datsettet.

Respons	Beskrivelse	Distribusjon
DFS	1 hvis pasienten ikke overlevde, fikk lokalt eller regionalt tilbakefall eller metastatisk tilbakefall, 0 ellers	Klasse 1: 57,9% Klasse 0: 42,1%
LRC	1 hvis pasienten fikk lokalt eller regionalt tilbakefall, 0 ellers	Klasse 1: 23,4% Klasse 0: 76,6%
OS	1 hvis pasienten ikke overlevde, 0 ellers	Klasse 1: 52,3% Klasse 0: 47,7%

Datsettet brukt til testing inkluderte de samme egenskapene som treningssettet fra OUS, men manglet *ecog*-egenskapen. Det er uvisst hvorfor dette var tilfelle, men trolig er det ikke standard prosedyre å registrere denne variabelen ved MAASTRO-klinikken.

3.2 Programvare

Programmeringsspråket som ble brukt i denne oppgaven var Python [66]. For egenkapsuthenting fra bildene ble versjon 3.7 brukt, mens for øvrig arbeid ble versjon

3.9 brukt. Som programmeringsverktøy ble Spyder og Jupyter Notebook fra Anaconda benyttet [67]. Numpy [68] og Pandas [69] ble brukt til databehandling, mens metrikk, standardisering, klassifiseringsalgoritmer (ekskl. LightGBM [58]) ble hentet fra ScikitLearn [55]. Egenskaper ble hentet ut fra bildene ved hjelp av innskaper [70], og egenskapsutvelging ble gjort ved bruk av RENT [71]. Nibabel [72] ble brukt til å lese PET- og CT-bildene, og til visualisering ble matplotlib [73], hoggorm [74], hoggormplot [75], seaborn [76] og Microsoft Excel [77] tatt i bruk .

3.3 Preprosessering

Preprosessering av de kliniske dataene og PET/CT bildene var en viktig del av forarbeidet til modelleringen, og sikret at dataen var klargjort til egenskapsutvelging og klassifisering. Gjennom preprosessering av klinisk data og bildedata var målet å formatere dataen slik at prediksjonene ble gode og reproduserbare.

3.3.1 Preprosessering av kliniske data

I de kliniske dataene manglet 26,4% av pasientene HPV-status (*hpv_related*) og kreftstadium (*wicc8_III – IV*). En pasients HPV-status er signifikant i prediksjonen av behandlingsutfallet, som vist av Moan et al. [63], ved at det er en sterkere relasjon mellom ikke-HPV-relatert hode- og halskreft og overlevelse, enn for hode- og halskreft relatert til HPV. Pasientene som manglet HPV-status hadde heller ikke registrert kreftstadium. Kreftstadium defineres på en skala fra 1 til 4, der stadium 1 er enklest å behandle, mens stadium 4 regnes som uhelbredelig og behandlingen som gis har som mål å være livsforlengende eller lindrende [78]. *wicc8_III_IV* er derfor også en viktig egenskap for å predikere om en pasient blir frisk eller ikke.

Før modellen ble trent måtte de manglende verdiene skiftes ut med numeriske verdier. I denne oppgaven ble de ukjente HPV-statusene og kreftstadiene tildelt en egen kategori i datasettet bruk til trening. Den nye kategoriinndelingen til egenskapene *hpv_related* og *wicc8_III – IV* er gitt i tabell 3.5. For at datasettene brukt til trening og testing skulle inneholde de samme egenskapene ble egenskapen *ECOG* fjernet fra datasettet brukt til trening.

Tabell 3.5: Oversikt over de nye kategoriinndelingene for *hpv_related* og *uicc8_III–IV* i datasett brukt til trening.

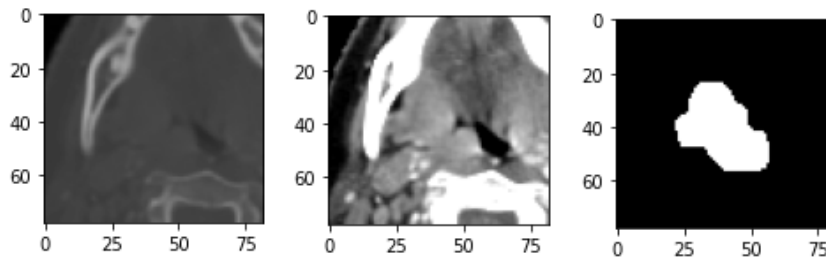
Egenskap	Ny kategoriinndeling	Distribusjon
hpv_related	Kategori 0: Negativ	Kategori 0: 32,0%
	Kategori 1: Positiv	Kategori 1: 41,6%
	Kategori 2: Ukjent	Kategori 2: 26,4%
uicc8_III-IV	Kategori 0: I-II	Kategori 0: 37,1%
	Kategori 1: III-IV	Kategori 1: 36,5%
	Kategori 2: Ukjent	Kategori 2: 26,4%

3.3.2 CT-windowing

Mens PET-bildene, som viste SUV-verdier, kunne benyttes direkte til egenskapsut-hentig, var det nødvendig å preprocessere CT-bildene for å øke kontrasten mellom pikslene. *CT-windowing* er en prosess der CT-bilder manipuleres gjennom histo-grammodifisering for å øke kontrasten mellom ulike elementer i bildet [79]. Lys-styrken til bildet justeres av vindusnivået (WL fra *window level*), og kontrasten justeres av vindusvidden (WW, fra *window width*).

Intensitetsverdiene i CT-bildene representerte HU, som strekker seg over en skala fra -1000 til 1000 [79]. Det vil si at når potensielt 2000 ulike dempningsverdier skal representeres av 256 gråtoner (et vanlig bilde), vil mange av de ulike dempnings-verdiene representeres av samme gråtone, og ulike HU blir vanskelige å skille [79]. Dette gjelder spesielt for ulike typer bløtvev. For å ta hensyn til dette manipuleres bildet ved at WW setter grenser for hvilke HU som skal representeres. Alle verdier som havner utenfor vinduet vises som enten helt svarte eller helt hvite. Pikslene in-nenfor vinduet representeres av en gråskala på 256 verdier der de ulike gråtonene representerer vesentlig færre HU, og dermed øker kontrasten mellom strukturer med liknede HU.

WL er definert som midpunktet i spekteret av HU som vises, og vil være midt på gråtoneskalaen. I denne oppgaven ble en WW på 350 HU brukt, og WL på 60 HU. Forskjellen på et manipulert CT-bilde og et rådatabilde er vist i figur 3.3.1.

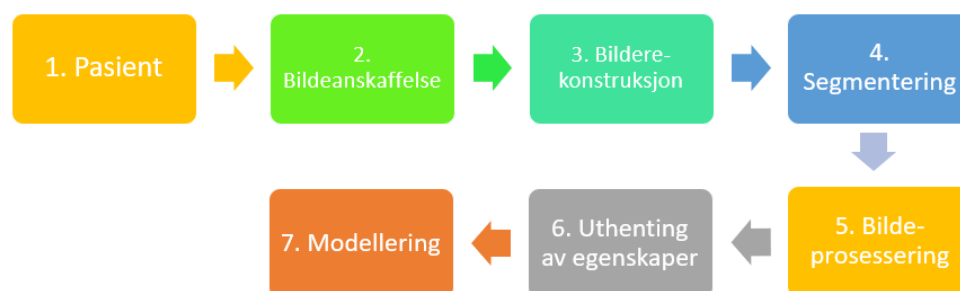


Figur 3.3.1: Bildet til venstre viser et tverrsnitt av et rådata-bilde fra CT-avbildningen. I midten vises den manipulerede versjonen av samme bilde. Masken til høyre viser ROI for bildet. I dette eksempelet er en WW på 350 HU og WL brukt.

3.4 Radiomics

Radiomics er en metode som brukes for å trekke ut egenskaper fra medisinske bilder ved bruk av ulike dataalgoritmer [80]. Målet er at disse egenskapene skal avdekke mønstre og karakteristikker i svulsten ved å se på forskjellene i de ulike vokslens intensitetsverdier [10]. I denne oppgaven ble det hentet ut radiomics-egenskaper fra PET- og CT-bildene.

Tidligere har det ikke vært en standardisering av radiomics-egenskaper, og derfor ble Image Biomarker Standardization Initiative (IBSI) opprettet [80]. Målet med IBSI er å etablere felles definisjoner og navn på egenskapene for å gjøre de reproduserbare. Det har blitt definert 169 radiomics-egenskaper brukt for å kvantifisere morfologiske karakteristikker, førsteordens statistikk og romlig forhold mellom vokslene innen ROI [80]. Arbeidsflyten i radiomics-prosedyren oppsummeres i figur 3.4.1.

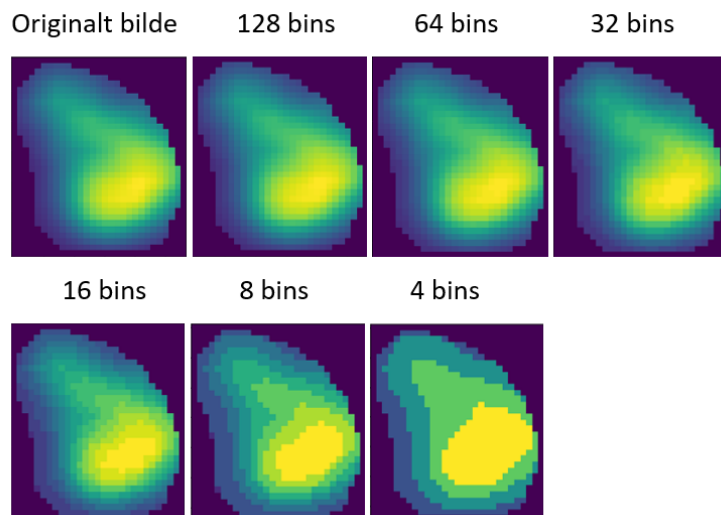


Figur 3.4.1: Arbeidsflyten i radiomics-prosedyren. Adaptert fra [10].

Steg 1-4 var i denne oppgaven dekket av det tildelte materialet: Ferdig rekonstruerte bilder av 197 pasienter (steg 1-3) med tilhørende masker som viste omfanget av svulsten (steg 4). Steg 5 og 6 forklares videre i denne seksjonen, mens steg 7 utføres i seksjon 3.5 til og 3.8.

3.4.1 Bildeprosessering

En viktig del av bildeprosesseringssteget var diskretisering av bildene. Det vil si at intensitetsverdiene ble delt inn i ulike grupper, herved omtalt som bins, for å minske antallet intensitetsverdier som ble vist i hvert bilde. Det finnes to hovedformer for diskretisering; fiksert binantall og fiksert binstørrelse [10]. Fiksert binantall, relativ diskretisering, er basert på bredden av intensitetsverdier funnet i ROI [10]. Her kan bredden på de ulike binsene variere fra bilde til bilde, men antall bins er konstant. Fiksert binstørrelse derimot, lager den samme størrelsen på alle bins på tvers av flere bilder, men antallet bins kan variere [10]. Begge metoder har en diskretiseringsnivå-parameter, henholdsvis antall bins, og binvidde. I denne oppgaven ble fiksert binantall benyttet, og det ble trukket ut egenskaper fra bilder diskretisert til 4, 8 og 16 bins for PET, og 16 og 32 bins for CT. Figur 3.4.2 viser en sammenlikning mellom et originalt PET-bilde innenfor ROI og samme bilde diskretisert til 128, 64, 32, 16, 8 og 4 bins.



Figur 3.4.2: En PET-avbildning med ulike verdier av diskretiseringsparameterem *antall bins*. Det originale bildet er representert av 256 gråtoner.

3.4.2 Uthenting av egenskaper

Egenskapene som ble trukket ut av de medisinske bildene var formegenskaper, førsteordens statistiske egenskaper og teksturegenskaper. De førsteordens statistiske egenskapene og teksturegenskapene (med unntak av Local Binary Pattern, LBP) finnes blant de IBSI-listede radiomics-egenskapene. Tabell 3.6 gir en oversikt over de ulike metodene brukt til egenskapsuthenting, og antall uthentede egenskaper for hver metode. En oversikt over alle de uthentede egenskapene finnes i tillegg B.

Tabell 3.6: Oversikt over de ulike metodene for uthenting av egenskaper og antall egenskaper hentet ut for hver metode.

Uthentingsmetode	Antall egenskaper
Formegenskaper	14
Førsteordens statistiske egenskaper	18
GLSZM	16
GLRLM	16
NGTDM	5
GLDM	14
GLCM	24
LBP	10
Totalt	85

Formegenskaper

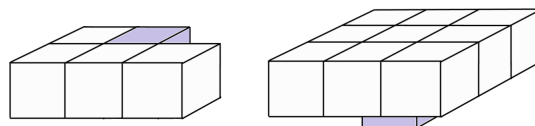
Formegenskapene ble hentet ut ifra masken som definerte ROI. Dette var egenskaper som baserte seg på størrelsen og formen til svulsten, som volum, diameter og rundhet [81].

Førsteordens statistiske egenskaper

De førsteordens statistiske egenskapene var baserte på distribusjonen av intensitetsverdier innen ROI [81]. Eksempler på førsteordens statistiske egenskaper er gjennomsnitt, maksimumsverdi og minimumsverdi av intensitetsverdiene i bildene og den representerte bredden av intensitetsverdier (spekter).

Teksturegenskaper

Teksturegenskapene ble bestemt ut i fra forholdet mellom nabovokslar, og hentet ut ved hjelp av ulike teksturmatriser. Disse egenskapene skulle avdekke egenskaper som ikke ble fanget av formegenskapene og de statistiske egenskapene [10]. Flere av teksturmatrisene brukt til egenskapsuthenting var avhengige av hvilken vinkel bildematrixen ble lest fra. En todimensjonal piksel har 8 nabovokslar orientert i 4 ulike vinkler, mens en tredimensjonal voksel har totalt 26 nabovokslar orientert i 13 ulike vinkler. Det vil si at vokslene orientert 180 grader ovenfor hverandre i forhold til sentervokselen representerer den samme vinkelen. Nabolaget kan ses på som en kube med $3 \times 3 \times 3$ vokslar. Disse totalt 27 vokslene utgjør en sentervoksel, i midten av kuben, med 26 naboer. Dersom denne kuben deles i tre deler i horisontal retning vil den øverste skiven representere ni vokslar orientert i ni ulike vinkler i forhold til senterpikselen. De ni vokslene i den nederste skiven vil representere de samme vinklene. Den midterste skiven består av åtte nabovokslar orientert i fire ulike vinkler. De ni vokslene i den øverste skiven og de fire vokslene i den midterste skiven som representerer de 13 ulike retningene er illustrert i figur 3.4.3.



Figur 3.4.3: En sentervoksel, markert i lilla, med 13 nabovokslar fordelt på et øvre plan (høyre) og et midtre plan (venstre). Disse nabovokslene er orientert i 13 ulike vinkler i forhold til sentervokselen. Orienteringen til de resterende 13 nabovokslene i forhold til sentervokselen kan representeres av en av vokslene representert i denne figuren.

Teksturegenskapene ble hentet ut ved hjelp av de ulike teksturmatrisene beskrevet nedenfor.

Gray Level Size Zone Matrix

Gray Level Size Zone Matrix (GLSZM) kvantifiserer gråtonesoner i et bilde, der størrelsen til hver sone er antallet tilknyttede voksler som deler den samme gråtone-intensiteten [81]. I en GLSZ-matrise $P(i, j)$ er element (i, j) lik antallet soner med intensitetsverdi i og sonestørrelse j i et bilde. GLSZM er rotasjonsuavhengig, og det trengs kun en matrise for alle retninger i en ROI, da resultantmatrisen vil bli den samme uavhengig av hvilken retning bildematriksen blir lest fra. Konstruksjonen av en todimensjonal GLSZM vises i figur 3.4.4.

Bildematrikse				
5	2	5	4	4
3	3	3	1	3
2	1	1	1	3
4	2	2	2	3
3	5	3	3	2

		GLSZM				
		Sonestørrelse →				
		1	2	3	4	5
Intensitetsverdi ↓	1	0	0	0	1	0
	2	1	0	0	0	1
	3	1	0	1	0	1
	4	1	1	0	0	0
	5	3	0	0	0	0

Figur 3.4.4: GLSZM til høyre konstruert fra bildematriksen til venstre. De grønne og gule feltene i bildematriksen korresponderer til de gule og grønne feltene i GLSZM. I dette eksempelet er $P(5, 1)$ (gul) lik 3 fordi det finnes tre piksler med verdi 5 som ikke har noen naboer med samme intensitetsverdi, det vil si at de har sonestørrelse 1. Tilsvarende er $P(1, 4)$ (grønn) lik 1, fordi det eksisterer én kjede av fire sammenhengende piksler med intensitetsverdi 1 i bildematriksen, det vil si sonestørrelse 4. Adaptert fra [81].

Gray Level Run Length Matrix

Gray Level Run Length Matrix (GLRLM) kvantifiserer intensitetsrekker av gråtoner, som defineres som lengden av sammenhengende piksler som har samme intensitetsverdi [81]. I en GLRL-matrise $P(i, j | \Theta)$ beskriver element (i, j) hvor mange ganger en intensitetsrekke med gråtone i og lengde j oppstår i ROI langs vinkel θ . Konstruksjonen av en todimensjonal GLRLM vises i figur 3.4.5.

Bildematrikse				
5	2	5	4	4
3	3	3	1	3
2	1	1	1	3
4	2	2	2	3
3	5	3	3	2

		GLRLM				
		Lengde →				
		1	2	3	4	5
Intensitetsverdi ↓	1	1	0	1	0	0
	2	3	0	1	0	0
	3	4	1	1	0	0
	4	1	1	0	0	0
	5	3	0	0	0	0

Figur 3.4.5: GLRLM for $\theta = 0$ (horisontal retning) til høyre konstruert fra bildematriksen til venstre. De grønne og gule feltene i bildematriksen korresponderer til de gule og grønne feltene i GLRLM. I dette eksempelet har verdi $P(4, 2)$ (gul) verdien 1 fordi det finnes én rekke av to sammenhengende piksler med intensitetsverdi 4 i bildematriksen. På samme måte har $P(2, 3)$ (grønn) verdien 1, da det finnes én rekke av tre sammenhengende piksler med intensitetsverdi 2 i bildematriksen. Adaptert fra [81].

Den returnerte verdien av en GLRLM-egenskap er gjennomsnittet av egenskapen beregnet fra en GLRLM for alle 13 vinkler [81].

Neighbouring Gray Tone Difference Matrix

Neighbouring Gray Tone Difference Matrix (NGTDM) kvantifiserer forskjellen mellom en gråtone og den gjennomsnittlige gråtonen blant nabovokslene innen en avstand δ [81]. Summen av absolutte forskjeller for gråtone i lagres i NGTDM-matriksen, og defineres som

$$s_i = \begin{cases} \sum^{n_i} |i - \bar{A}_i| & \text{for } n_i \neq 0 \\ 0 & \text{for } n_i = 0 \end{cases} \quad (3.1)$$

der \bar{A}_i er den gjennomsnittlige gråtonen til nabolaget og n_i er antall voksler med gråtone i [81]. p_i er gråtonesannsynligheten, og definert som

$$p_i = \frac{n_i}{N_v} \quad (3.2)$$

hvor N_v er det totale antallet voksler i bildematriksen [81]. Konstruksjonen av en NGTDM fra et todimensjonalt bilde vises i figur 3.4.6.

Bildematrixe

1	2	5	2
3	5	1	3
1	3	5	5
3	1	1	1

NGTDM

i	n _i	p _i	s _i
1	6	0,375	13,350
2	2	0,125	2,000
3	4	0,250	2,630
4	0	0,000	0,000
5	4	0,250	10,075

Figur 3.4.6: NGTDM til høyre konstruert fra bildematrixen til venstre for $\delta = 1$. s_i er summen av den gjennomsnittlige nabopikselverdien til alle piksler med intensitetsverdi i . p_i er sannsynligheten for at en piksel har intensitetsverdi i og n_i er antall piksler i bildematrixen med intensitetsverdi i . Adaptert fra [81].

I denne oppgaven ble $\delta = 1$ brukt for NGTDM, som betyr at det kun ble tatt hensyn til de nærmeste naboene med direkte tilknytning til senterpiksel.

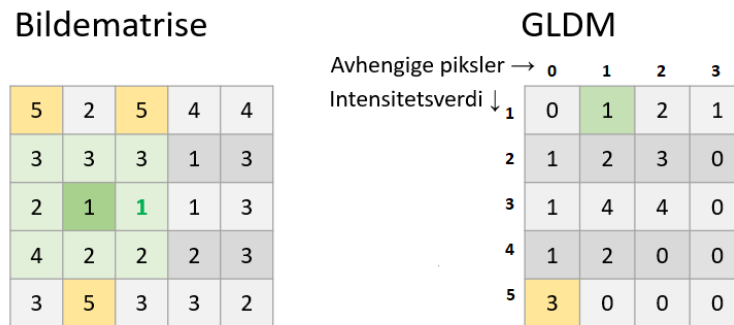
Gray Level Dependence Matrix

Gray Level Dependence Matrix (GLDM) kvantifiserer gråtoneavhengighet i et bilde. Gråtoneavhengighet er definert som antallet sammenkoblede vokslar innen en avstand δ fra en sentervoksel, gitt i antall vokslar som er avhengig av sentervokselen [81]. En nabovoksel med gråtone j er avhengig av sentervokselen med gråtone i dersom

$$|i - j| \leq \alpha$$

der α er en terskeverdi for avhengighet. I en GLD-matrix $P(i, j)$ beskriver element (i, j) antall ganger en voksel med gråtone i med j avhengige nabovokslar dukker opp i et bilde. Konstruksjonen av en todimensjonal GLDM vises i figur 3.4.7.

I denne oppgaven ble $\delta = 1$ brukt for GLDM, som betyr at det kun ble tatt hensyn til de nærmeste naboene med direkte tilknytning til senterpiksel. Verdien av α var lik 0 for beregningen av GLDM-egenskapene.



Figur 3.4.7: GLDM for $\alpha = 0$ og $\delta = 1$ til høyre konstruert fra bildematriksen til venstre. De grønne og gule feltene i bildematriksen korresponderer til de gule og grønne feltene i GLDM. I dette eksempelet er $P(5, 0)$ (gul) lik 3 fordi det finnes tre piksler med intensitetsverdi 5 som ikke har noen avhengige piksler i sitt nabolag i bildematriksen. Tilsvarende er $P(1, 1)$ (grønn) lik 1 fordi det finnes én piksel med intensitetsverdi 1 som har én avhengig piksel i nabolaget i bildematriksen. Merk at for GLDM telles det fra 0 i horisontal retning, mens for de resterende matrisene telles det fra 1. Det betyr at også intensitetsverdiene uten avhengige naboer inkluderes i GLDM. Adaptert fra [81].

Gray Level Co-occurrence Matrix

Gray Level Co-occurrence Matrix (GLCM) beskriver en andreordens felles sannsynlighetsfunksjon av en bilderegion begrenset av masken (ROI) [81]. I en GLC-matrise $P(i, j | \delta, \theta)$ representerer elementet $P(i, j)$ antall ganger kombinasjonen av intensitetsnivåer i og j oppstår i to piksler i bildet separert av en avstand på δ piksler langs en vinkel θ . Konstruksjonen av en todimensjonal GLCM vises i figur 3.4.8.

I denne oppgaven ble $\delta = 1$ brukt for GLCM, som betyr at det kun ble tatt hensyn til de nærmeste naboene med direkte tilknytning til senterpiksel. Den returnerte verdien av en GLCM-egenskap er gjennomsnittet av egenskapen beregnet fra en GLCM i alle 13 vinkler [81].

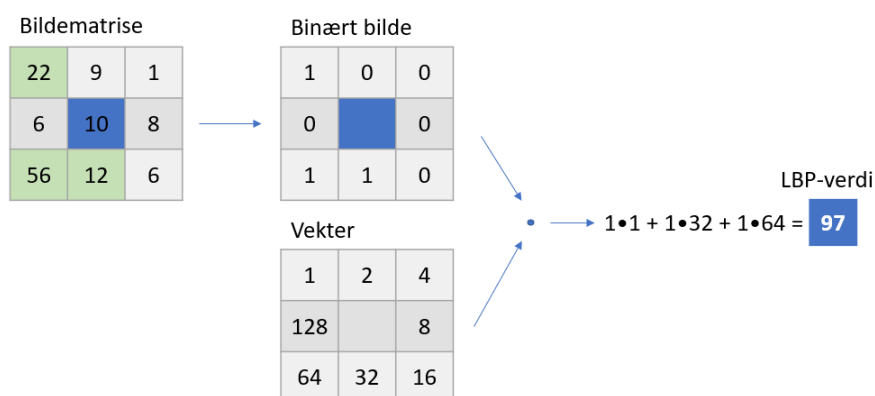
Bildematrix				
1	2	5	2	3
3	2	1	3	1
1	3	5	5	2
1	1	1	1	2
1	2	4	3	5

GLCM						
		Intensitetsverdi →				
		1	2	3	4	5
Intensitetsverdi ↓	1	6	4	3	0	0
	2	4	0	2	1	3
	3	3	2	0	1	2
	4	0	1	1	0	0
	5	0	3	2	0	2

Figur 3.4.8: GLCM for $\theta = 0^\circ$ og $\delta = 0$ til høyre konstruert fra bildematrixen til venstre. De grønne og gule feltene i bildematrixen korresponderer til de gule og grønne feltene i GLCM. $\theta = 0^\circ$ tilsier at det letes etter kombinasjoner rett til venstre og rett til høyre for en senterpiksel. I dette eksempelet er $P(5, 2)$ (gul) tildelt verdien 3 fordi en voksel med intensitetsverdi 2 ligger rett til høyre eller venstre for en voksel med intensitetsverdi 5 tre ganger i bildematrixen. Tilsvarende er $P(2, 4)$ (grønn) tildelt verdien 1 fordi vokslene med intensitetsverdiene 2 og 4 kun befinner seg ved siden av hverandre et sted i bildematrixen. Adaptert fra [81].

Local Binary Pattern

Local Binary Pattern (LBP) brukes til å beskrive lokale teksturmønstre i et bilde [82]. LBP jobber i blokkstørrelser på 3×3 , der senterpikselen brukes som en terskel for nabopikselene og LBP-verdien til senterpikselen genereres ved å først kode om blokken til et binært bilde hvor nabopikslene med høyere verdi enn senterpikselen gis verdien 1, og øvrige pikslene gis verdien 0. Den binære matrixen leses som et binært tall, i en sirkel som starter til venstre for senterpikselen og mot klokka. Dette er sammenfallende med å gange det binære bildet elementvis med en vektmatrix bestående av verdier som konverterer binære verdier til desimalverdier, illustrert i figur 3.4.9. LBP-verdien til senterpikselen er summen av disse desimalverdiene. LBP-egenskapen bygges opp av et histogram av LBP-verdiene, som regnes ut for hver piksel i bildet [82].



Figur 3.4.9: En 3x3 blokk fra en bildematrix konverteres til et binært bilde ved at pikslene med høyere verdi enn senterpikselen tildeles verdien 1, og øvrige piksler tildeles verdien 0. Videre konverteres det binære bildet til desimalverdier ved å multiplisere elementvis med vektmatrisen. LBP-verdien er summen av disse desimalverdiene. Adaptert fra [82].

Rotasjon av bildematriksen fører til ulik tolkning av LBP-verdier, og det finnes 10 unike rotasjonsavhengige mønstre som gir ulike verdier [83]. I denne oppgaven hentes det derfor ut 10 ulike LBP-egenskaper fra bildene, én for hver avhengige rotasjon.

3.4.3 Inskaper

Inskaper, tidligere biorad, ble utviklet av tidligere masterstudent Severin Langberg [84] og videreutviklet av ph.d.-kandidat Bao Ngoc Huynh og tidligere masterstudenter Ahmed Albuni [85] og Nasibeh Mohammadi [86]. Koden til inskaper er tilgjengelig fra: <https://github.com/NMBU-Data-Science/insker>. Inskaper trekker ut formegenskaper, førsteordens statistiske egenskaper og tekstoregnskaper fra bilder og tilhørende masker [70]. Inskaper kan kjøres med både fiksert binstørrelse eller fiksert binantall. Inndataen til inskaper er bilder og masker, binparameter, og markering av hvilke egenskaper som skal trekkes ut. Utdataen er en datamatrix som inneholder de ulike egenskapene til svulsten for hver pasient.

Inskaper hentet ut egenskapene definert i tillegg B fra PET- og CT-bildene med de forhåndsdefinerte diskretiseringsparameterene: 4, 8 og 16 bins for PET, og 16 og 32 bins for CT. Dette resulterte i fem datamatrixer som ble tatt med videre til egenskapsutvelging. Disse datamatrixene blir videre omtalt som egenskapsgrupper.

3.5 Arbeidsflyt

For hver av de tre responsene fulgte en arbeidsflyt som hadde som mål å finne prediksjonsmodellen med høyest ytelse. Arbeidsflyten er illustrert i figur 3.5.1.



Figur 3.5.1: Flytskjema som illustrerer prosessen som gjennomføres for å velge ut relevante egenskaper, bygge en prediksjonsmodell og teste denne.

3.6 Egenskapsutvelging

Egenskapsutvelgingen tok for seg utvelgelsen av egenskaper blant de kliniske dataene, PET-parameterene og bildeegenskapene. Etter uthenting av egenskaper besto datasettet av vesentlig flere egenskaper enn pasienter. Når en modell har for mange parametere øker sannsynligheten for at modellen blir overtilpasset, fordi den blir veldig kompleks [9]. Det var derfor hensiktsmessig å eliminere enkelte egenskaper før modellering for å gjøre beregningene mer effektive og modellene mer generaliserende. I tillegg kunne enkelte egenskaper påvirke prediksjonen i negativ retning, og det var viktig å utelukke disse. Målet med egenskapsutvelgingen var

å finne kombinasjonen av egenskaper som ga den mest korrekte prediksjonen av behandlingsutfall.

3.6.1 Korrelasjon mellom egenskaper fra ulike grupper

Korrelasjon mellom egenskaper fra ulike grupper utgjorde sammen med konstruksjon av datasett den preliminare egenskapsutvelgingen, steg 1, i figur 3.5.1. Korrelasjonen mellom egenskapene fra de ulike egenskapsgruppene hentet ut med imskaper ble regnet ut gjennom en RV-matrise (relativ varians) implementert i pakken Hoggorm [74]. Dersom en gruppe hadde høyere korrelasjon enn 99% med en annen gruppe, ble den gruppen av de to med den høyeste diskretiseringsparameteren eliminert.

3.6.2 Konstruksjon av datasett

Egenskapene fra gruppene som ikke ble eliminert i forrige underseksjon ble slått sammen til et datasett, sammen med de kliniske egenskapene og PET-parameterene. Eventuelle dupliserte verdier i de ulike gruppene ble fjernet. Denne sammenlåingen dannet datamatriksen som var utgangspunktet for videre egenskapsutvelging.

3.6.3 Repetert elastisk nett teknikk

Repetert elastisk nett teknikk (RENT) er en egenskapsutvelger for binær klassifisering og regresjon [71], som utgjorde steg 2 i arbeidsflyten fra figur 3.5.1. RENT trener et ensemble av generaliserte lineære modeller ved bruk av elastisk nett-regularisering. Elastisk nett-regularisering er et kompromiss mellom L1- og L2-regularisering som bruker L1-parameteren til å generere sparsomhet og L2-parameteren til å kontrollere at ikke for mange egenskaper velges bort [9].

RENT trener et ensemble av unike modeller basert på unike undersett av pasienter fra datasettet brukt til trening. Fordi hver modell er trent på ulike undersett vil alle modellene være litt ulike, og regulariseringen vil velge ulike egenskaper for hver modell [71]. I tillegg undersøker RENT hvor konsistent elastisk nett velger egenskaper på tvers av de unike modellene ved å analysere fordelingen av vektstørrelser for hver egenskap.

Fordi gjennomsnittlig ytelse til ensemblet varierer for ulike kombinasjoner av L1- og L2-parametere predefineres et intervall verdier av regulariseringsparameteren C og L1-forhold ($L1_ratio$). Kombinasjonen av disse parameterene som gir den mest nøyaktige prediksjonen kan identifiseres gjennom 5-foldet kryssvalidering (seksjon 3.7.1), eller velges manuelt. For denne parameterkombinasjonen regner RENT ut et ensemble av modeller som vil brukes til egenskapsutvelging [71].

Andre parametere som kan tilpasses i RENT er antall modeller, K , som bestemmer hvor mange modeller ensemblet skal bestå av, $testsize_range$ som tillater mer tilfeldighet i modellen ved å variere størrelsen på undersettene og $random_state$ som setter en verdi for å reproducere resultater. Parameteren $autoEnetParSel$ avgjør om den beste parameterkombinasjon av C og $L1_ratio$ hentes ut automatisk gjennom 5-foldet kryssvalidering ($autoEnetParSel = True$), eller om RENT trener K modeller for hver kombinasjon av hyperparameterene C og $L1_ratio$ ($autoEnetParSel = False$). Som standard inkluderes egenskapene som de er formatert i datasettet under egenskapsutvelgingen, ved at parameteren $poly$ er satt til OFF . Dersom variabelen $poly$ er satt til ON benyttes også kvadratet av egenskapene og vekselvirkninger (produkt) mellom alle egenskapene i modellen. Dette er gunstig om dataen er ulineær, men krever mye kapasitet og bør brukes på et allerede redusert datasett [71]. I egenskapsutvelgelsen ble logistisk regresjon brukt som klassifiseringsalgoritme for modellene, og MCC brukt som metrikk i modellevalueringen. En oversikt over de ulike verdiene som er brukt for de ulike parameterene i RENT finnes i tabell 3.7.

Tabell 3.7: Oversikt over hvilke verdier som er brukt for de ulike parameterene i RENT.

Parameter	Verdi
C	0.1 , 1, 10, 100
$L1 - ratio$	0, 0.1, 0.25, 0.5, 0.75, 0.9
K	100
$testsize_range$	(0.25, 0.25)
$random_state$	none
$autoEnetParSel$	False
$poly$	ON, OFF
$scoring$	MCC
$classifier$	logreg

Ved hjelp av parameterene τ_1 , τ_2 og τ_3 kan det reguleres hvor aggressivt RENT skal velge ut egenskaper. Disse parameterene defineres som følgende [71]:

- τ_1 : Kun egenskaper som blir valgt en høyere andel ganger enn denne verdien blant de K modellene blir selektert. I denne oppgaven ble τ_1 variert mellom 0.2 og 0.9 i steg på 0.1.
- τ_2 : Bestemmer hvor stor andel av egenskapenes vekter som må ha samme fortegn, for å sikre konsistens i modellen og eliminere egenskaper av liten betydning. I denne oppgaven hadde τ_2 alltid samme verdi som τ_1 , for å sørge for at flere egenskaper faktisk ble inkludert når det første kriteriet ble mindre restriktivt.
- τ_3 : Tester om egenskapsvektene er konsistent høye med lav varians på tvers av alle modeller, og korresponderer til statistisk t-test. I denne oppgaven brukes $\tau_3 = 0.975$, som tilsvarer et signifikansnivå på $\alpha = 0.5$, det vil si at $\tau_3 = 1 - 0.5 \cdot \alpha$.

En egenskap ble kun valgt dersom den oppfylte alle tre kriteriene [71].

3.6.4 Valg av egenskaper

I denne oppgaven ble det hentet ut ulike undersøtt av egenskaper for videre analyse. De ulike undersettene ble valgt ut ved å variere τ_1 og τ_2 mellom 0.2 og 0.9 i intervaller på 0.1, som resulterte i totalt åtte undersøtt for hver respons. I første omgang ble det kun hentet ut egenskaper uten vekselvirkninger. Valg av egenskaper utgjorde det tredje steget i arbeidsflyten i figur 3.5.1.

3.6.5 Korrelasjon mellom valgte egenskaper

Før egenskapene valgt ut av RENT ble brukt til predikering, ble korrelasjonen mellom de ulike utvalgte egenskapene beregnet. Dersom korrelasjonen mellom to egenskaper var høyere enn 0.985 ble egenskapen som ble valgt færrest ganger i

RENT ekskludert fra undersettet. Korrelasjonskoeffisienten brukt i denne oppgaven var Pearsons korrelasjonskoeffisient, definert som

$$\rho_{A,B} = \frac{\text{cov}(A,B)}{\sigma_A \sigma_B} \quad (3.3)$$

der $\text{cov}(A,B)$ er kovariansen mellom to ulike egenskaper A og B , og σ_A og σ_B er standardavviket til henholdsvis egenskap A og egenskap B . Denne elimineringen av egenskaper ble utført på alle de åtte undersettene, og utgjorde steg 4 i arbeidsflyten i figur 3.5.1.

3.7 Etablering av rammeverk for prediksjon

Prediksjonsdelen av oppgaven startet med å lage et rammeverk for hver av de åtte klassifiseringsalgoritmene som skulle brukes til å predikere behandlingsutfall. Dette utgjorde steg 5 i figur 3.5.1. Rammeverket besto av et standardiseringssteg sammen med den aktuelle klassifiseringsalgoritmen tilpasset med standardparametere. I denne oppgaven ble klassifiseringsalgoritmene logistisk regresjon, random forest, KNN, SVM, AdaBoost, QDA, passive aggressive og LightGBM brukt på hvert av de åtte undersettene.

3.7.1 Standardisering av data

Standardiseringssteget skalerte alle egenskapene slik at gjennomsnittsverdien til hver egenskap i datamatriksen brukt til trening var 0 og standardavviket 1 [9]. Standardiseringen ble gjort ved likningen

$$\mathbf{x}'_j = \frac{\mathbf{x}_j - \mu_j}{\sigma_j} \quad (3.4)$$

hvor \mathbf{x}_j er vektoren som inneholder verdien av en egenskap for alle pasientene, μ_j er gjennomsnittsverdien til egenskapen i treningssettet og σ_j er standardavviket til egenskapen i treningssettet. Hensikten med standardiseringen er å gjøre det enklere for modellen å forstå og løse problemet når alle alle egenskaper bidrar likt, og modellen vil kunne konvergere raskere mot en løsning [9]. Skaleringen ble utført i forhold til datasettet brukt til trening både for treningssettet og testsettet, og måtte utføres på nytt for hver kombinasjon av undersett av egenskaper og undersett av pasienter. I denne oppgaven ble standardisering gjort i hver iterasjon

under kryssvalidering (steg 6 i figur 3.5.1, forklart i seksjon 3.7.1) før en klassifiseringsalgoritme utførte en prediksjon og før trening og testing av de tilpassede modellene.

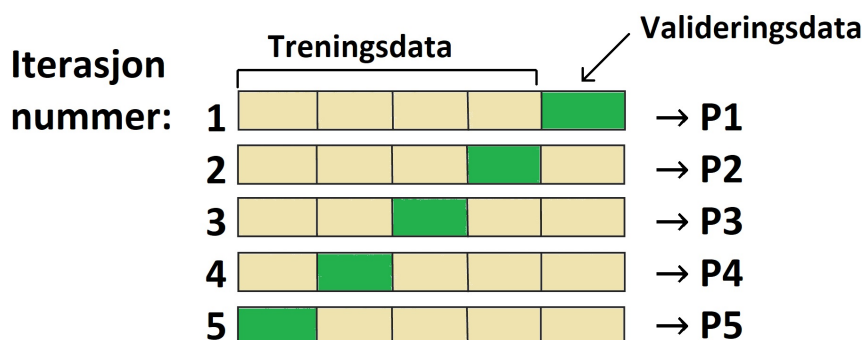
I denne oppgaven ble pakken `StandardScaler` fra `scikit-learn` benyttet til standardisering [55].

3.8 Validering av modeller

Steg 6 i figur 3.5.1 gikk ut på å finne de beste kombinasjonene av hyperparametere for de ulike klassifiseringsalgoritmene, og finne ut hvordan de ulike klassifiseringsalgoritmene ytet på de ulike undersettene av egenskaper. For å finne hyperparameterene som ga de beste resultatene for hver enkelt modell ble kryssvalidering og *grid search* brukt [55]. For å velge ut hvilke modeller som hadde størst potensiale til å predikere best på det eksterne datasettet ble klassifiseringsalgoritmene med tilpassede hyperparameterene på nytt validert gjennom en kryssvalidering.

3.8.1 Kryssvalidering

Kryssvalidering utføres ved at datasettet brukt til trening deles inn i et gitt antall folder. Deretter trenes modellen på alle foldene unntatt en, mens den siste folden benyttes til testing [9]. Prediksjonen måles opp mot sannheten ved hjelp av de valgte metrikkene, og resultatet lagres i en vektor for hver prediksjon. Prosedyren gjentas, men nå benyttes en annen fold til testing og resterende til trening. Dette gjentas helt til alle foldene har blitt benyttet til testing en gang. Deretter regnes gjennomsnittet av de lagrede resultatene ut for de valgte metrikkene, og disse verdiene gir et mer robust bilde på hvordan modellen vil yte på usett data enn om man kun utfører én test på en tilfeldig splittelse av datasettet [9]. I denne oppgaven ble 5-foldet kryssvalidering brukt, som er illustrert i figur 3.8.1.



Figur 3.8.1: Datasettet brukt til trening deles opp i fem folder, og hver fold benyttes etter tur til foldet kryssvalidering av en modell trent på de fire andre foldene. I dette eksempelet benyttes den grønne folden til testing og resterende til trening. Resultatet som rapporteres er gjennomsnittet av prediksjonene for hver iterasjon, som i dette eksempelet vil være gjennomsnittet av $P1 - P5$. Adaptert fra [9].

For å få enda mer robusthet i resultatene kan kryssvalideringen repeteres flere ganger. I denne oppgaven brukes pakken `RepeatedStratifiedKFold` fra `scikit-learn` [55].

3.8.2 Grid search-kryssvalidering

I en grid search-kryssvalidering søker algoritmen gjennom en tabell av ulike verdier for ulike hyperparametere, og gjennomfører en kryssvalidering for hver parameterkombinasjon [9]. Parameterkombinasjonen som ga høyest gjennomsnittlig verdi for MCC i kryssvalideringen blir lagret, slik at denne kombinasjonen kan brukes videre i modelleringen. I denne oppgaven brukes pakken `GridSearchCV` fra `scikit-learn` [55].

I valideringen av modellene ble det gjort en Grid Search for hver av de åtte klassifiseringsalgoritmene, logistisk regresjon, random forest, KNN, SVM, AdaBoost, QDA, passive aggressive og LightGBM, for hvert undersøkt av egenskaper. Kryssvalideringen som ble brukt for å validere de ulike parameterkombinasjonene var 5-foldet stratifisert kryssvalidering repetert 10 ganger.

Deretter ble disse modellene validert på nytt ved bruk av 5-foldet stratifisert kryssvalidering repetert 100 ganger, det vil si at en modellene med de utvalgte hyperparameterne ble trent og testet totalt 500 ganger for å oppnå robuste resultater.

For hver repetisjon ble datasettet stokket om slik at de nye inndelingene ikke var like som i tidligere repetisjoner. Stratifiseringen ble gjort med hensyn til responsvariabelen, og sikret dermed at klassedistribusjonene var tilnærmet like i hver fold. Dette motvirket ubalanse i valideringen. Kryssvalideringen av modellene ble gjort ved bruk av F1-score for begge klasser, nøyaktighet, MCC og ROC AUC. De ulike gjennomsnittlige metrikkene for hver klassifiseringsalgoritme for hvert undersett ble lagret i en tabell med totalt 64 ulike kombinasjoner av undersett av egenskaper og klassifiseringsalgoritmer.

3.9 Modeller med vekselvirkninger

I steg 7 i arbeidsflyten fra figur 3.5.1 ble *poly* i RENT endret fra *OFF* til *ON*, og egenskapsutvelging og validering av modeller ble repetert. Dette ble gjort for å se om modeller som også tok hensyn til vekselvirkninger mellom egenskaper kunne oppnå en høyere ytelse enn de tidligere modellene.

Istedenfor å bruke det fullstendige datasettet var det undersettet hentet ut med τ_1 og τ_2 lik 0.2 for DFS og OS og 0.5 for LRC som ble benyttet i RENT. De nye undersettene av egenskaper ville nå kunne bestå av de opprinnelige egenskapene, i tillegg til kvadratet av egenskaper og produktet mellom ulike egenskaper.

3.10 Etablere en baseline

Det siste steget før modellene ble testet på det eksterne datasettet, steg 8 i figur 3.5.1, besto av å etablere en *baseline*. En baseline fungerer som et sammenlikningsgrunnlag for fremtidige modeller, og setter en grense det forventes at de fremtidige modellene yter bedre enn. Baseline kan settes som modellytelsen man ville oppnådd ved tilfeldig gjetning, eller implementeres ved hjelp av en klassifiseringsmodell, som ble gjort i denne oppgaven. Målet for de fremtidige modellene er å slå baseline-ytelsen.

Et av målene med denne oppgaven var å vise at det var mulig å gjøre flere korrekte prediksjoner ved å inkludere radiomics-egenskaper enn ved å kun benytte klinisk data. Baselineytelsen tilsvarte derfor ytelsen til den modellen kun trent på kliniske data med høyest ytelse. Disse ytelsene ble funnet ved samme fremgangsmåte som for modellene trent på det fulle datasettet, men det ble ikke sett på modeller

med vekselvirkning mellom egenskaper. Det vil si at etableringen av en baseline ble utført gjennom steg 2-6 i figur 3.5.1, før modellene ble testet på det eksterne datasettet.

3.11 Testing på eksternt datasett

I steg 9 i figur 3.5.1 ble de tilpassede klassifiseringsalgoritmene for de ulike undersettene testet på det hittil urørte eksterne datasettet fra MAASTRO-klinikken. Resultatene ble validert med F1-score for begge klasser, nøyaktighet, MCC og ROC AUC.

Steg 2-9 i figur 3.5.1 ble repetert for hver av de tre responsene.

Kapittel 4

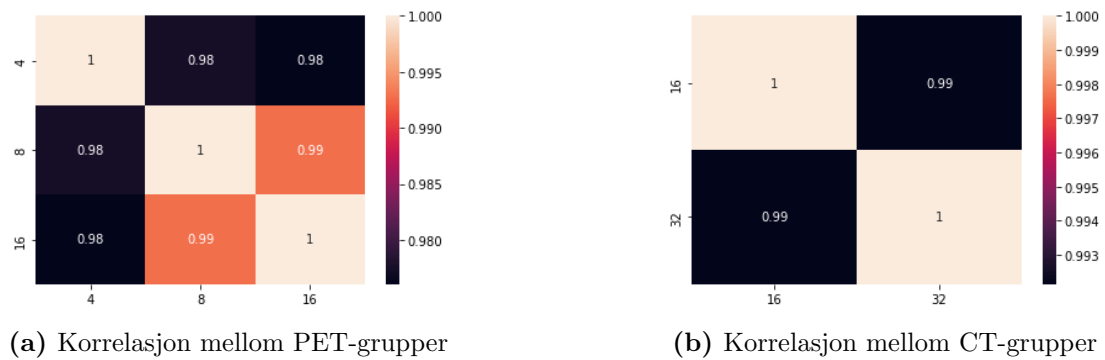
Resultater

I dette kapittelet presenteres resultatene som fremkommer fra eksperimentene beskrevet i kapittel 3 om materiale og metode.

4.1 Preliminær egenskapsutvelging

4.1.1 Korrelasjon mellom egenskaper fra ulike grupper

For hver verdi av diskretiseringsparameteren binantall hentes de 117 form-, førsteordens statistiske og tekstoregenskapene beskrevet i tillegg B ut for hvert bilde av hver pasient. Dette resulterer i totalt 5 egenskapsgrupper, *PET_16*, *PET_8*, *PET_4*, *CT_32* og *CT_16*. Mange av egenskapene er identiske eller tilnærmet identiske i de ulike gruppene, og enkelte grupper har derfor høy korrelasjon seg imellom. Korrelasjonen mellom egenskapene i de ulike gruppene for PET og CT er illustrert i figur 4.1.1.



Figur 4.1.1: Korrelasjonen mellom egenskapsgruppene hentet ut fra PET-bilder diskretisert til 16, 8 og 4 bins er vist i figur a). Korrelasjonen mellom gruppene hentet ut fra CT-bilder diskretisert til 32 og 16 bins er vist i figur b).

Korrelasjonen mellom PET-gruppene hentet ut fra bildene diskretisert til 16 bins og 8 bins var høyere enn 99%. Derfor ble gruppen egenskaper hentet ut fra bildene diskretisert til 16 bins fjernet. Korrelasjonen mellom PET-gruppene hentet ut fra bildene diskretisert til 8 og 4 bins var også veldig høy, men da en eliminering av *PET_8* gruppen ville ført til at bildeteksturen ble veldig grov, ble begge beholdt for å bevare mer av finteksturen i bildene, og for å unngå å eliminere egenskaper som kunne inneholde viktig informasjon. Korrelasjonen mellom de to CT-gruppene var høyere enn 99%, og derfor ble gruppen egenskaper hentet ut fra bildene diskretisert til 32 bins fjernet.

4.1.2 Konstruksjon av datasett

Etter at de gjenværende egenskapsgruppene ble slått sammen med de kliniske egenskapene og PET-parameterene ble eventuelle duplikatverdier fjernet. Til dette ble funksjonen *drop_duplicates* fra pandas brukt [69]. Den komplette datamatriksen som ble brukt til trening besto av 309 egenskaper. En oppsummering av de ulike egenskapene i datasettet finnes i tabell 4.1.

Tabell 4.1: Oversikt over de ulike typene egenskaper i datasettet brukt til trening.

Type egenskap	Antall egenskaper
Form	14
Førsteordens, PET	19
Førsteordens, CT	17
PET-tekstur	160
CT-tekstur	85
PET-parametere	3
Klinisk	11
Totalt	309

4.2 Baseline-modeller

For hver av de tre responsene ble det etablert en baseline-modell, som fungerte som et sammenlikningsgrunnlag for fremtidige modeller. Disse modellene ble valgt ut ved å bruke RENT på kun de kliniske egenskapene, finne modellen med høyest ytelse under kryssvalidering, og til slutt teste denne på det eksterne datasettet. Dette ble repetert for hver av de tre responsene. Baseline-ytelsen målt i MCC for hver respons er gitt i tabell 4.2.

Tabell 4.2: Oversikt over baseline-ytelse gitt i MCC for de ulike responsene. De listede egenskapene er undersettet av kliniske egenskaper valgt av RENT som oppnådde høyest MCC under kryssvalidering. Kryssvalidering gir ytelsen modellen oppnådde under stratifisert 5-foldet kryssvalidering repetert 100 ganger på treningssettet, mens ekstern validering gir ytelsen modellen oppnådde ved testing på det eksterne datasettet. *Kl.alg.: klassifiseringsalgoritme, LR: logistisk regresjon

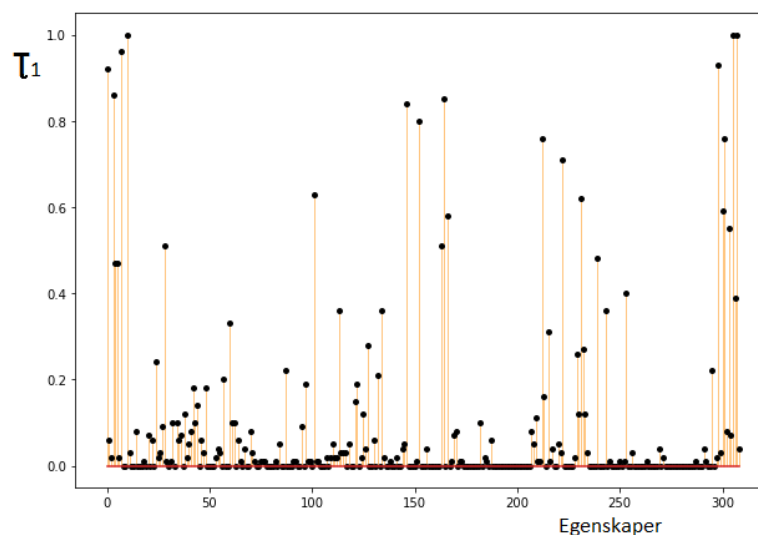
Respons	Egenskaper	Kl.alg.*	Kryssval.	Ekstern val.
DFS	age	LR*	0.382 ± 0.136	0.172
	cavum_oris			
	oropharynx			
	hpv_related			
	charlson			
	pack_years			
	uicc8_III-IV			
OS	age	LR*	0.500 ± 0.136	0.201
	cavum_oris			
	hpv_related			
	pack_years			
	uicc8_III-IV			
LRC	age	KNN	0.210 ± 0.165	0.182
	female			
	cavum_oris			
	larynx			
	histgrade_high			
	hpv_related			
	charlson			
	pack_years			
	uicc8_III-IV			

4.3 DFS

I denne seksjonen presenteres resultatene som fremkom fra steg 2-9 i figur 3.5.1 i detalj for responsen DFS. For responsene OS og LRC presenteres resultatene etter samme oppsett i henholdsvis seksjon 4.4 og 4.5.

4.3.1 Egenskapsutvelging

I denne oppgaven ble RENT brukt som egenskapsutvelger [71]. For DFS ble $C = 0.1$ og $L1_ratio = 0.5$ valgt som den optimale parameterkombinasjonen for elastisk-nett-regulariseringen i RENT. Fra de $K = 100$ modellene trent på ulike undersett av pasienter, ble det laget et frekvensplott som viser hvor ofte de ulike egenskapene ble valgt. Dette frekvensplottet er vist i figur 4.3.1.



Figur 4.3.1: Frekvensplott som viser hvor ofte de ulike egenskapene ble valgt av modellene i RENT for responsen DFS. Egenskap 1-14 er formegenskaper, 15-116 er CT-egenskaper, 117-295 er PET-egenskaper, 296-298 er PET-parametere og 299-309 er kliniske egenskaper. τ_1 beskriver andelen av modeller i RENT som valgte den aktuelle egenskapen.

For DFS var det mange egenskaper som aldri eller sjelden ble valgt. Toppene til venstre og høyre i figur 4.3.1 viser at de oftest valgte egenskapene var formegenskaper og kliniske egenskaper. PET-egenskapene ser også ut til å bli valgt oftere

enn CT-egenskapene.

Undersettene av egenskaper ble selektert basert på ulike verdier av τ_1 og τ_2 , som er terskelverdier for henholdsvis hvor stor andel av modellene i RENT som må velge egenskapen, og hvor stor andel av prøvene som må ha samme fortegn for at egenskapen skal bli valgt. Dette ble gjort for å teste ulike kombinasjoner av egenskaper, og for å finne ut hvor mange egenskaper som kunne lukes ut uten at det gikk utover modellenes ytelse. Verdiene av τ_1 og τ_2 ble holdt like for utvelgelsen av undersett, og varierte fra 0.2 til 0.9 i steg på 0.1. De ulike undersettene har fått navn etter hvilke τ -verdier som ble brukt som kriterie for utvelgelse. Egenskapene inkludert i de ulike undersettene er vist i figur 4.3.2.

4.3.2 Egenskapsutvelging med vekselvirkningsegenskaper

Egenskapene fra undersett 0.2 fra figur 4.3.2 ble kjørt på nytt gjennom RENT, men med parameteren *poly* skiftet fra *OFF* til *ON*, som gjorde det mulig for RENT å velge både opprinnelige egenskaper, kvadratet av egenskaper og produktet mellom egenskaper. Undersettet ble valgt fordi antallet egenskaper ikke var for stort med tanke på beregningskostnad, samtidig som vekselvirkning mellom flere egenskaper ble tatt i betraktning sammenlignet med om undersett 0.7 hadde blitt valgt. Undersett 0.2 består av 41 egenskaper, dermed vil RENT kunne velge blant 1722 egenskaper med og uten vekselvirkninger, fra likningen

$$B = A + A^2 \tag{4.1}$$

der A er antall egenskaper i undersett 0.2 og B er antall egenskaper i det nye datasettet med vekselvirkninger inkludert. Fra dette settet av egenskaper ble det hentet ut nye undersett av egenskaper for τ -verdier mellom 0.2 og 0.9 i steg på 0.1, og $\tau_1 = \tau_2$. Disse undersettene ble navngitt som τ -kriteriet for utvelgelse etterfulgt av en v . Vekselvirkningsmodellene er mer kompliserte enn modellene uten vekselvirkning, og kun de tre modellene med høyest ytelse blir presentert i denne oppgaven. Undersettet med vekselvirkningsegenskaper som ga den høyeste ytelsen gitt i MCC var undersett 0.9v, og er presentert i tabell 4.3.

Egenskap	τ_1, τ_2							
	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
shape_Elongation	x	x	x	x	x	x	x	x
shape_Maximum3DDiameter	x	x	x	x	x	x	x	x
shape_Sphericity	x	x	x	x	x	x	x	x
age	x	x	x	x	x	x	x	x
hpv_related	x	x	x	x	x	x	x	x
pack_years	x	x	x	x	x	x	x	x
shape_MajorAxisLength	x	x	x	x	x	x	x	x
glszm_ZoneEntropyPET_8	x	x	x	x	x	x	x	
glrlm_HighGrayLevelRunEmphasisPET_8	x	x	x	x	x	x	x	
glrlm_ShortRunLowGrayLevelEmphasisPET_8	x	x	x	x	x	x	x	
shape_MajorAxisLength	x	x	x	x	x	x		
LBP_102PET_8	x	x	x	x	x	x		
glszm_GrayLevelVariancePET_4	x	x	x	x	x	x		
oropharynx	x	x	x	x	x	x		
glcm_MCC_d_1CT_16	x	x	x	x	x			
glszm_SmallAreaHighGrayLevelEmphasisPET_4	x	x	x	x	x			
first_order_SkewnessCT_16	x	x	x	x				
glrlm_HighGrayLevelRunEmphasisPET_8	x	x	x	x				
glrlm_ShortRunHighGrayLevelEmphasisPET_8	x	x	x	x				
ngtdm_Coarseness_d_1PET_8	x	x	x	x				
cavum_oris	x	x	x	x				
larynx	x	x	x	x				
shape_Maximum2DDiameterColumn	x	x	x					
shape_Maximum2DDiameterRow	x	x	x					
ngtdm_Coarseness_d_1PET_4	x	x	x					
glrlm_ShortRunEmphasisCT_16	x	x						
LBP_201CT_16	x	x						
glszm_GrayLevelNonUniformityNormalizedPET_8	x	x						
LBP_201PET_8	x	x						
glrlm_HighGrayLevelRunEmphasisPET_4	x	x						
glrlm_LowGrayLevelRunEmphasisPET_4	x	x						
charlson	x	x						
first_order_MinimumCT_16	x							
glrlm_RunLengthNonUniformityNormalizedCT_16	x							
glcm_Correlation_d_1CT_16	x							
first_order_RangePET_8	x							
first_order_VariancePET_8	x							
glrlm_ShortRunLowGrayLevelEmphasisPET_8	x							
glszm_SizeZoneNonUniformityNormalizedPET_4	x							
glszm_SmallAreaLowGrayLevelEmphasisPET_4	x							
SUVpeak	x							

Figur 4.3.2: Egenskapene i de ulike undersettene valgt ut ved bruk av RENT for respon- sen DFS, basert på ulike verdier av terskelverdiene $\tau_1 = \tau_2$. Undersettet av egenskaper markert i grønt ga høyest MCC under kryssvalidering. Egenskaper markert i rødt ble fjernet grunnet høy korrelasjon med andre egenskaper.

Tabell 4.3: Undersett 0.9v for responsen DFS, som ga høyest gjennomsnittlig verdi av MCC under kryssvalidering av modellene med vekselvirkningsegenskaper. Dette undersettet er basert på $\tau_1 = \tau_2 = 0.9$ på egenskapssettet med vekselvirkninger mellom egenskaper. De gule cellene representerer formegenskaper, de grønne cellene representerer CT-egenskaper, de blå cellene representerer PET-egenskaper og de rosa cellene representerer kliniske egenskaper. Egenskapen brukt i modelleringen er produktet av egenskap 1 og 2, $Egenskap1 \cdot Egenskap2$, der to egenskaper er oppgitt.

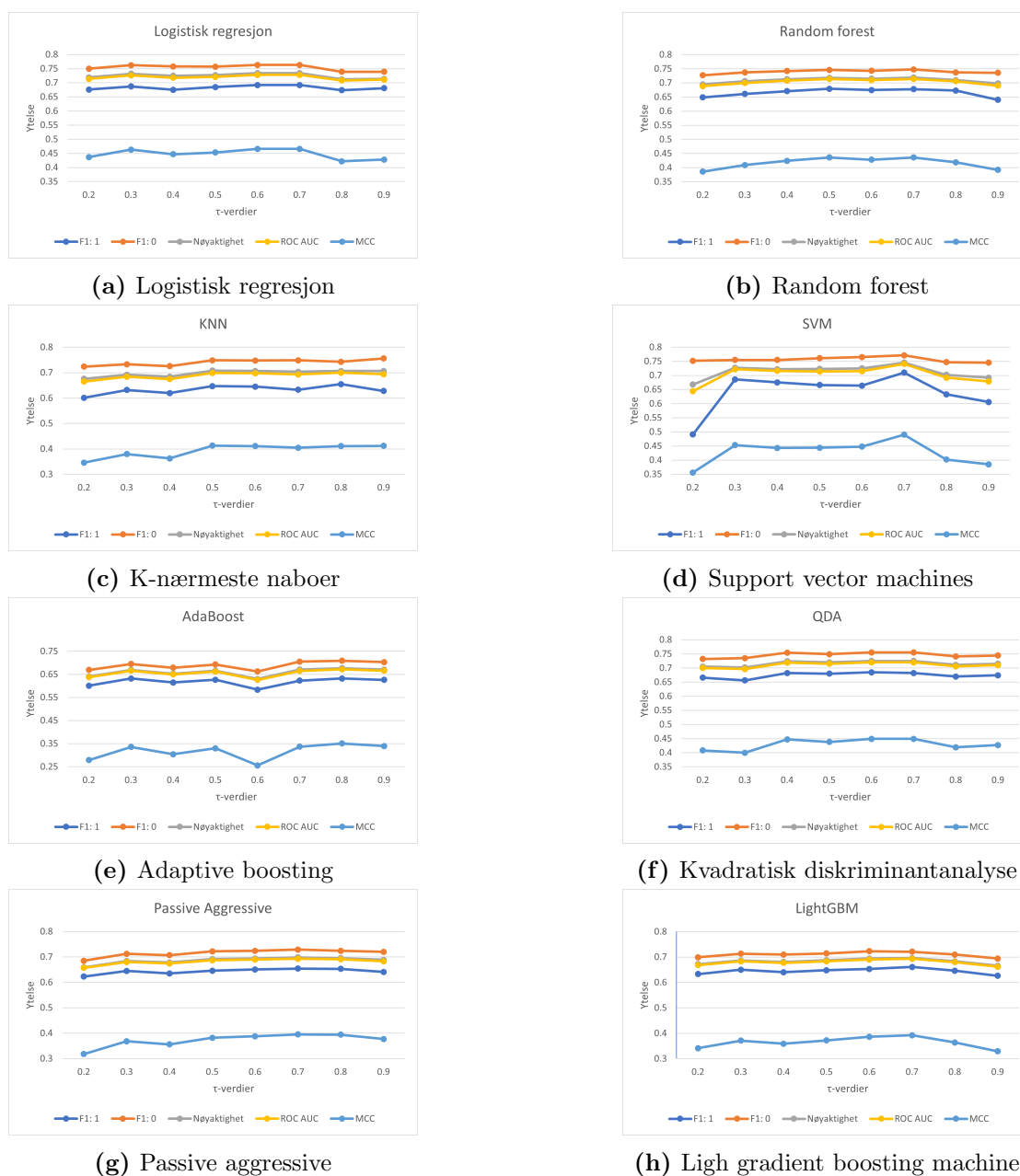
Egenskap 1	Egenskap 2
shape_Sphericity	-
shape_Elongation	shape_Sphericity
shape_Elongation	glrlm_ShortRunEmphasisCT_16
shape_Elongation	LBP_201CT_16
shape_Elongation	glszm_GrayLevelNonUniformityNormalizedPET_8
shape_Sphericity	glrlm_RunLengthNonUniformityNormalizedCT_16
shape_Sphericity	glrlm_ShortRunEmphasisCT_16
shape_Sphericity	glrlm_HighGrayLevelRunEmphasisPET_8
glrlm_RunLengthNonUniformityNormalizedCT_16	hpv_related
glcm_MCC_d_1CT_16	glszm_ZoneEntropyPET_8
glcm_MCC_d_1CT_16	glrlm_ShortRunLowGrayLevelEmphasisPET_8
glszm_ZoneEntropyPET_8	glrlm_ShortRunLowGrayLevelEmphasisPET_8
glrlm_ShortRunHighGrayLevelEmphasisPET_8	hpv_related
glrlm_ShortRunLowGrayLevelEmphasisPET_8	age
ngtdm_Coarseness_d_1PET_8	LBP_102PET_8
LBP_102PET_8	ngtdm_Coarseness_d_1PET_4
glszm_GrayLevelVariancePET_4	age
glszm_SmallAreaLowGrayLevelEmphasisPET_4	pack_years

4.3.3 Kryssvalidering

Hvert av de åtte undersettene uten vekselvirkningsegenskaper, presentert i figur 4.3.2, og de åtte undersettene med vekselvirkningsegenskaper var igjennom en grid search for hver av de åtte ulike klassifiseringsalgoritmene; logistisk regresjon, random forest, KNN, SVM, AdaBoost, QDA, passive aggressive og LightGBM. Deretter ble hyperparameterkombinasjonene som ga den høyeste ytelsen for MCC testet på nytt i en 5-foldet stratifisert kryssvalidering repetert 100 ganger. Her ble metrikkene F1-score for begge klasser, nøyaktighet, ROC AUC og MCC rapportert. De gjennomsnittlige resultatene fra kryssvalideringene for undersettene uten vekselvirkningsegenskaper er visualisert i figur 4.3.3. SVM med τ -verdier på 0.7 ga den høyeste modellytelsen for DFS. Dette var relativt restriktive τ -verdier, og undersettet av egenskaper besto kun av 14 egenskaper, som vist i figur 4.3.2. De tre kombinasjonene av klassifiseringsalgoritme og undersett av egenskaper uten vekselvirkningsegenskaper som viste høyest gjennomsnittlig ytelse under kryssvalideringen er gitt i tabell 4.4. En fullstendig oversikt over ytelsen til alle modellene uten vekselvirkningsegenskaper under kryssvalidering finnes i tillegg C.

Tabell 4.4: De tre modellene uten vekselvirkningsegenskaper for responsen DFS med høyest ytelse basert på 5-foldet stratifisert kryssvalidering repetert 100 ganger. Første linje for hver modell gir den gjennomsnittlige ytelsen, mens den andre linjen gir standardavviket til modellen. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av egenskaper brukt i kryssvalideringen.

Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
SVM	0.7	0.71	0.771	0.745	0.741	0.490
		0.075	0.057	0.061	0.062	0.124
Logistisk regresjon	0.7	0.692	0.763	0.734	0.728	0.466
		0.081	0.059	0.065	0.066	0.132
Logistisk regresjon	0.3	0.687	0.762	0.732	0.726	0.463
		0.085	0.060	0.066	0.068	0.136



Figur 4.3.3: Visualisering av den gjennomsnittlige ytelsen for DFS-modellene gitt i F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, de ulike kombinasjonene av undersett av egenskaper, gitt ved τ -verdien, hvor $\tau_1 = \tau_2$, og klassifiseringsalgoritme a) til f) oppnådde under 5-foldet stratifisert kryssvalidering repetert 100 ganger.

De to modellene med høyest gjennomsnittlig ytelse var begge basert på undersettet der τ_1 og τ_2 var lik 0.7, som viser at det var mulig å utføre gode prediksjoner selv på et undersett av egenskaper der en stor andel egenskaper ble ekskludert. Logistisk regresjon hadde en stabil ytelse for τ -verdier mellom 0.2 og 0.7, vist i figur 4.3.3 a), og det var mulig å ekskludere et stort antall egenskaper uten fall i modellytelse. Usikkerheten i modellene var noe høyere for de logistiske regresjonsmodellene enn for SVM-modellen.

De tre modellene med vekselvirkningsegenskaper som hadde høyest gjennomsnittlig verdi av MCC under kryssvalidering for DFS er gitt i tabell 4.5.

Tabell 4.5: De tre modellene med vekselvirkningsegenskaper for responsen DFS med høyest ytelse basert på 5-foldet stratifisert kryssvalidering repetert 100 ganger. Første linje for hver modell gir den gjennomsnittlige ytelsen, mens den andre linjen gir standardavviket til modellen. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av vekselvirkningsegenskaper brukt i kryssvalideringen.

Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
Logistisk regresjon	0.9v	0.752	0.803	0.782	0.778	0.564
		0.072	0.054	0.059	0.060	0.120
QDA	0.7v	0.746	0.805	0.780	0.775	0.561
		0.075	0.052	0.059	0.061	0.121
Logistisk regresjon	0.8v	0.746	0.797	0.776	0.772	0.552
		0.072	0.055	0.060	0.061	0.121

Modellene med vekselvirkningsegenskaper oppnådde en MCC som var høyere enn modellene uten vekselvirkningsegenskaper under kryssvalidering, med en usikkerhet som var litt lavere. Disse modellene var også baserte på undersett av egenskaper som inneholdt flere egenskaper enn modellene uten vekselvirkningsegenskaper med høyest ytelse.

4.3.4 Testing på eksternt datasett



Figur 4.3.4: Visualisering av ytelsen, gitt i F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, de ulike kombinasjonene av undersett av egenskaper, gitt ved τ -verdien, hvor $\tau_1 = \tau_2$, og klassifiseringsalgoritme a) til f) oppnådde ved testing på eksternt datasett for responen DFS.

Samtlige modeller bygget på OUS-datasettet ble testet på det usette eksterne datasettet fra MAASTRO-klinikken. Det vil si at det ble utført én prediksjon av behandlingsutfallet til MAASTRO-pasientene, før denne prediksjonen ble sammenlignet med det sanne behandlingsutfallet. Ytelsen til de ulike klassifiseringsalgoritmene på det eksterne datasettet ble validert med F1-score for begge klasser, nøyaktighet, ROC AUC og MCC. Ytelsen til modellene på undersettene uten vekselvirkningsegenskaper er visualisert i figur 4.3.4.

De tre modellene presentert i tabell 4.4, var modellene uten vekselvirkningsegenskaper som hadde høyest ytelse under kryssvalidering for OUS-datasettet. Ytelsen til disse tre modellene på det eksterne datasettet er gitt i tabell 4.6 a). De tre modellene som hadde høyest ytelse på det eksterne datasettet, uavhengig av hvordan disse modellene presterte under kryssvalidering, er presentert i tabell 4.6 b). En fullstendig oversikt over ytelsen til de ulike modellene på det eksterne datasettet finnes i tillegg C.

Tabell 4.6: Tabell a) viser ytelsen til de tre modellene uten vekselvirkningsegenskaper for responsen DFS som hadde høyest ytelse under kryssvalidering testet på eksternt datasett. Tabell b) viser de tre modellene uten vekselvirkningsegenskaper med høyest ytelse under testing på eksternt datasett. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av egenskaper brukt til prediksjon. Modellene med ytelse over baseline-ytelsen på $MCC = 0.172$ er markert i grønt.

a)

Klassifiseringsalgoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
SVM	0.7	0.667	0.377	0.566	0.529	0.064
Logistisk regresjon	0.7	0.672	0.313	0.556	0.510	0.023
Logistisk regresjon	0.3	0.710	0.333	0.596	0.544	0.111

b)

Klassifiseringsalgoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
AdaBoost	0.2	0.733	0.478	0.646	0.609	0.244
Random forest	0.3	0.737	0.462	0.646	0.605	0.242
LightGBM	0.2	0.723	0.471	0.636	0.600	0.222

Tabell 4.6a) viser at ingen av de tre modellene uten vekselvirkningsegenskaper med høyest ytelse under kryssvalidering presterte bedre enn baseline-modellen på det eksterne datasettet. Baseline-ytelsen var på $MCC = 0.172$ for DFS. Fra tabell 4.6b) kommer det frem at enkelte andre modeller, som ikke gjorde det like bra under kryssvalideringen, hadde en ytelse høyere enn baseline-ytelsen på det eksterne datasettet.

De tre modellene presentert i tabell 4.5, var modellene med vekselvirkningsegenskaper som hadde høyest ytelse under kryssvalidering. Ytelsen til disse tre modellene på det eksterne datasettet er gitt i tabell 4.7. De tre modellene med vekselvirkningsegenskaper som hadde høyest ytelse på det eksterne datasettet, uavhengig av hvordan disse modellene presterte under kryssvalidering, er gitt i tabell 4.7b).

Tabell 4.7: Tabell a) viser ytelsen til de tre modellene med vekselvirkningsegenskaper for responsen DFS som hadde høyest ytelse under kryssvalidering testet på eksternt datasett. Tabell b) viser de tre modellene med vekselvirkningsegenskaper for responsen DFS med høyest ytelse basert på testingen på eksternt datasett. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av vekselvirkningsegenskaper brukt til prediksjon. Modellene med ytelse høyere enn baseline-ytelsen på $MCC = 0.172$ er markert i grønt.

a)

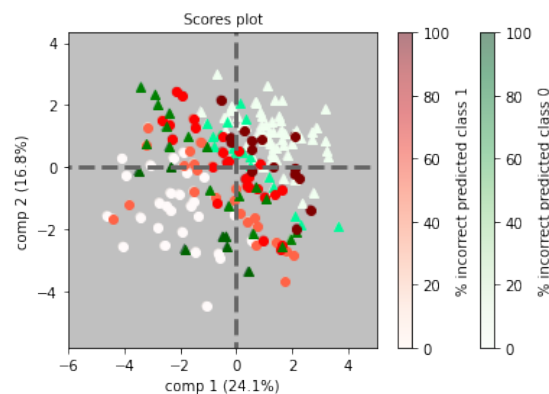
Klassifiseringsalgoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
Logistisk regresjon	0.9v	0.730	0.393	0.626	0.577	0.190
QDA	0.7v	0.710	0.333	0.596	0.544	0.111
Logistisk regresjon	0.8v	0.739	0.400	0.636	0.586	0.215

b)

Klassifiseringsalgoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
Passive aggressive	0.6v	0.757	0.414	0.657	0.603	0.270
KNN	0.2v	0.729	0.493	0.646	0.612	0.246
Passive aggressive	0.8v	0.733	0.478	0.646	0.609	0.244

Tabell 4.7a) viser at to av de tre modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering presterte bedre enn baseline-modellen på det eksterne datasettet. Fra tabell 4.7b) kommer det frem at enkelte andre modeller, som ikke gjorde det like bra under kryssvalideringen, hadde en enda høyere ytelse enn modellene i tabell 4.7a) på det eksterne datasettet.

En prinsipalkomponentanalyse (PCA) utført på undersettet av egenskaper uten vekselvirkninger med høyest ytelse under kryssvalidering for DFS er illustrert i figur 4.3.5. Plottet viser at det var en glidende overgang mellom de to klassene, og pasientene som oftest ble feilpredikerte var de som befant seg i den midtre delen av PCA-plottet.



Figur 4.3.5: PCA av undersettet av egenskaper som ga høyest ytelse under kryssvalidering. Sirkelene representerer pasientene med sann klasse 1, og trekantene representerer pasientene med sann klasse 0. Fargeintensiteten symboliserer andelen ganger den aktuelle pasienten ble feilpredikert av modellene i RENT.

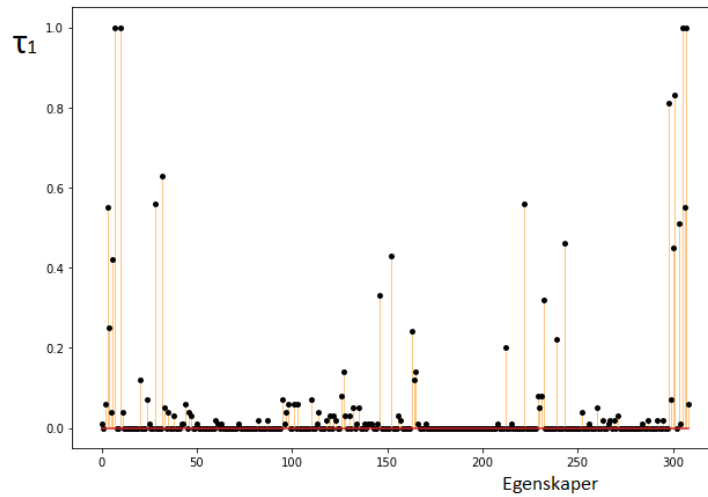
4.4 OS

I dette delkapittelet presenteres egenskapsutvelging med og uten vekselvirkningsegenskaper, kryssvalidering og testing på eksternt datasett for responsen OS.

4.4.1 Egenskapsutvelging

For OS ble $C = 0.1$ og $L1_ratio = 0.75$ valgt som den optimale parameterkombinasjonen for elastisk-nett-regulariseringen i RENT. Frekvensplottet basert på de

$K = 100$ modellene trent på ulike undersett av pasienter i RENT er vist i figur 4.4.1.



Figur 4.4.1: Frekvensplot som viser hvor ofte de ulike egenskapene ble valgt av modellene i RENT for responsen OS. Egenskap 1-14 er formegenskaper, 15-116 er CT-egenskaper, 117-295 er PET-egenskaper, 296-298 er PET-parametere og 299-309 er kliniske egenskaper. τ_1 beskriver andelen av modeller i RENT som valgte den aktuelle egenskapen.

Mange egenskaper ble aldri eller sjelden ble valgt. Toppene til venstre og høyre i figur 4.4.1 viser at de oftest valgte egenskapene var formegenskaper og kliniske egenskaper, mens bildeegenskapene blir valgt sjeldnere for OS enn for DFS.

Egenskap	τ_1, τ_2							
	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
shape_Maximum3DDiameter	x	x	x	x	x	x	x	x
shape_Sphericity	x	x	x	x	x	x	x	x
hpv_related	x	x	x	x	x	x	x	x
pack_years	x	x	x	x	x	x	x	x
age	x	x	x	x	x	x	x	x
oropharynx	x	x	x	x	x	x	x	
glszm_GrayLevelNonUniformityNormalizedCT_16	x	x	x	x	x			
shape_MajorAxisLength	x	x	x	x				
first_order_SkewnessCT_16	x	x	x	x				
glszm_GrayLevelVariancePET_4	x	x	x	x				
larynx	x	x	x	x				
charlson	x	x	x	x				
shape_Maximum2DDiameterSlice	x	x	x					
glrlm_HighGrayLevelRunEmphasisPET_8	x	x	x					
glrlm_LowGrayLevelRunEmphasisPET_4	x	x	x					
cavum_oris	x	x	x					
glszm_ZoneEntropyPET_8	x	x						
glszm_SmallAreaLowGrayLevelEmphasisPET_4	x	x						
shape_Maximum2DDiameterColumn	x							
glrlm_ShortRunHighGrayLevelEmphasisPET_8	x							
LBP_102PET_8	x							
glrlm_HighGrayLevelRunEmphasisPET_4	x							

Figur 4.4.2: Egenskapene i de ulike undersettene valgt ut ved bruk av RENT for responsen OS, basert på ulike verdier av terskelverdiene $\tau_1 = \tau_2$. Undersettet markert i grønt ga høyest MCC under kryssvalidering.

4.4.2 Egenskapsutvelgning med vekselvirkningsegenskaper

Egenskapene fra undersett 0.2 fra figur 4.4.2 ble kjørt på nytt gjennom RENT for å inkludere vekselvirkningsegenskaper. Undersett 0.2 består av 22 egenskaper, dermed vil RENT kunne velge blant 506 egenskaper med vekselvirkninger, fra likning 4.1. Det ble fra dette settet av egenskaper hentet ut nye undersett av egenskaper for τ -verdier mellom 0.2 og 0.9 i steg på 0.1, og $\tau_1 = \tau_2$, navngitt som τ -kriteriet for utvelgelse etterfulgt av en v. De tre modellene med høyest ytelse under kryssvalidering presenteres i tabell 4.10. Undersettet med vekselvirkningsegenskaper som ga den høyeste ytelsen gitt i MCC var undersett 0.5v, og er presentert i tabell 4.8.

Tabell 4.8: Undersett 0.5v for responsen OS, som ga høyest gjennomsnittlig verdi av MCC under kryssvalidering av modellene med vekselvirkningsegenskaper. Dette undersettet er basert på $\tau_1 = \tau_2 = 0.9$ på egenskapssettet med vekselvirkninger mellom egenskaper. De gule cellene representerer formegenskaper, de grønne cellene representerer CT-egenskaper, de blå cellene representerer PET-egenskaper og de rosa cellene representerer kliniske egenskaper. Egenskapen brukt i modelleringen er produktet av egenskap 1 og 2, *Egenskap 1* · *Egenskap 2*, der to egenskaper er oppgitt.

Egenskap 1	Egenskap 2
shape_MajorAxisLength	pack_years
shape_Maximum2DDiameterColumn	pack_years
shape_Maximum2DDiameterSlice	pack_years
shape_Maximum3DDiameter	glszm_GrayLevelVariancePET_4
shape_Maximum3DDiameter	age
shape_Maximum3DDiameter	pack_years
shape_Sphericity	shape_Sphericity
shape_Sphericity	first_order_SkewnessCT_16
shape_Sphericity	glszm_GrayLevelNonUniformityNormalizedCT_16
shape_Sphericity	hpv_related
glszm_ZoneEntropyPET_8	glrlm_LowGrayLevelRunEmphasisPET_4
glszm_ZoneEntropyPET_8	age
glrlm_ShortRunHighGrayLevelEmphasisPET_8	oropharynx
glrlm_ShortRunHighGrayLevelEmphasisPET_8	hpv_related
glszm_GrayLevelVariancePET_4	glszm_SmallAreaLowGrayLevelEmphasisPET_4
glszm_GrayLevelVariancePET_4	age
glszm_GrayLevelVariancePET_4	pack_years
glszm_SmallAreaLowGrayLevelEmphasisPET_4	larynx
glszm_SmallAreaLowGrayLevelEmphasisPET_4	pack_years
charlson	pack_years

4.4.3 Kryssvalidering

Hvert undersett av egenskaper med og uten vekselvirkninger fikk tilpasset hyperparametere for hver av de åtte klassifiseringsalgoritmene, og ble testet på nytt i en 5-foldet stratifisert kryssvalidering repetert 100 ganger. De gjennomsnittlige resultatene fra kryssvalideringene for undersettene uten vekselvirkninger er visualisert i figur 4.4.3.

Figur 4.4.3 viser at QDA med τ -verdier på 0.6 ga den høyeste modellytelsen for OS. Dette var ganske restriktive τ -verdier, og dette undersettet av egenskaper besto kun av 7 egenskaper, som vist i figur 4.4.2. De tre kombinasjonene av klassifiseringsalgoritmer og undersett av egenskaper uten vekselvirkninger som ga høyest ytelse under kryssvalideringen er gitt i tabell 4.9.

Tabell 4.9: De tre modellene uten vekselvirkningsegenskaper for responsen OS med høyest ytelse basert på 5-foldet stratifisert kryssvalidering repetert 100 ganger. Første linje for hver modell gir den gjennomsnittlige ytelsen, mens den andre linjen gir standardavviket til modellen. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av egenskaper brukt i kryssvalideringen.

Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
QDA	0.6	0.688	0.826	0.778	0.754	0.527
		0.091	0.046	0.059	0.066	0.130
QDA	0.9	0.69	0.824	0.776	0.754	0.525
		0.089	0.048	0.060	0.066	0.131
QDA	0.3	0.703	0.812	0.771	0.759	0.522
		0.078	0.049	0.057	0.061	0.121



Figur 4.4.3: Visualisering av den gjennomsnittlige ytelsen, gitt i F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, de ulike kombinasjonene av undersett av egenskaper, gitt ved τ -verdien, hvor $\tau_1 = \tau_2$, og klassifiseringsalgoritme a) til f) oppnådde under 5-foldet stratifisert kryssvalidering repetert 100 ganger for responsen OS.

De tre modellene med høyest gjennomsnittlig ytelse var alle basert på klassifise-

ringsalgoritmen QDA. Det var svært liten forskjell i ytelsen til de tre modellene, som viser at det er mulig å ekskludere et stort antall egenskaper uten fall i modellytelse. Usikkerheten i modellene var tilnærmet lik, og på samme nivå som usikkerheten i modellene for DFS.

De tre modellene med vekselvirkningsegenskaper som hadde høyest gjennomsnittlig verdi av MCC under kryssvalidering er gitt i tabell 4.10.

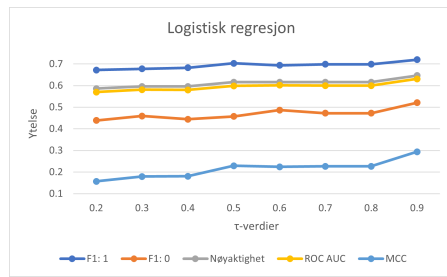
Tabell 4.10: De tre modellene med vekselvirkningsegenskaper for responsen OS med høyest ytelse basert på 5-foldet stratifisert kryssvalidering repetert 100 ganger. Første linje for hver modell gir den gjennomsnittlige ytelsen, mens den andre linjen gir standardavviket til modellen. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av vekselvirkningsegenskaper brukt i kryssvalideringen.

Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
QDA	0.5v	0.769	0.844	0.814	0.813	0.621
		0.070	0.049	0.056	0.059	0.116
QDA	0.2v	0.764	0.846	0.814	0.808	0.617
		0.074	0.052	0.059	0.061	0.121
QDA	0.4v	0.762	0.843	0.812	0.807	0.612
		0.074	0.050	0.058	0.062	0.121

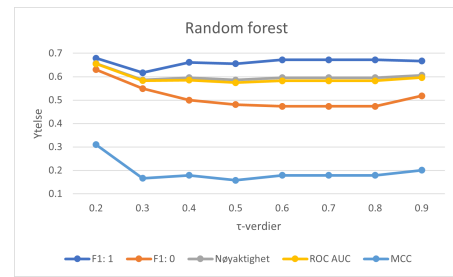
Modellene med vekselvirkningsegenskaper oppnådde en MCC som var høyere enn modellene uten vekselvirkningsegenskaper under kryssvalidering, med en usikkerhet som var litt lavere. Disse modellene var også baserte på undersett av egenskaper som inneholdt flere egenskaper enn for modellene uten vekselvirkningsegenskaper med høyest ytelse.

4.4.4 Testing på eksternt datasett

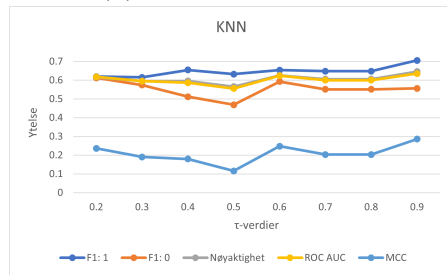
Samtlige modeller bygget på OUS-datasettet ble testet på det usette eksterne datasettet fra MAASTRO-klinikken. Modellytelsen ble validert med F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, og er visualisert i figur 4.4.4.



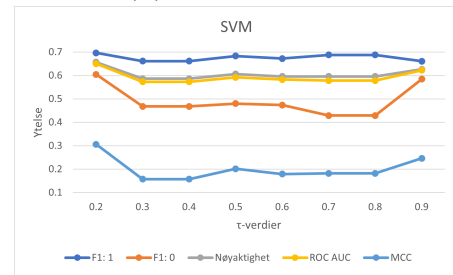
(a) Logistisk regresjon



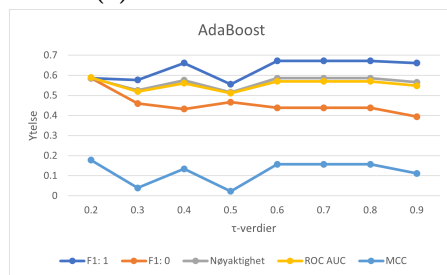
(b) Random forest



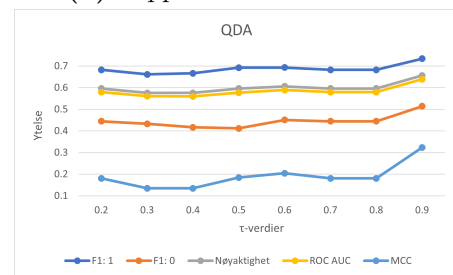
(c) K-nærmeste naboer



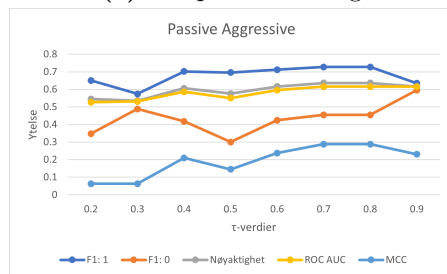
(d) Support vector machines



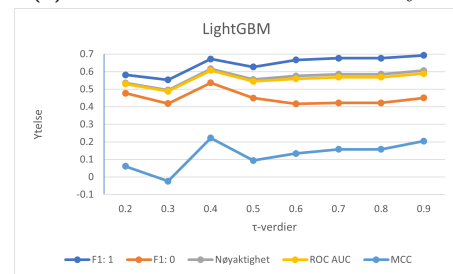
(e) Adaptive boosting



(f) Kvadratisk diskriminantanalyse



(g) Passive aggressive



(h) Ligh gradient boosting machine

Figur 4.4.4: Visualisering av ytelsen, gitt i F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, de ulike kombinasjonene av undersett av egenskaper, gitt ved τ -verdien, hvor $\tau_1 = \tau_2$, og klassifiseringsalgoritme a) til f) oppnådde ved testing på eksternt data-sett for responsen OS.

De tre modellene presentert i tabell 4.9, var modellene uten vekselvirkningsegen-

skaper som hadde høyest ytelse under kryssvalidering for OUS-datasettet. Ytelsen til disse tre modellene på det eksterne datasettet er gitt i tabell 4.11a). De tre modellene uten veskelvirkningsegenskaper som hadde høyest ytelse på det eksterne datasettet, uavhengig av hvordan disse modellene presterte under kryssvalidering, er presentert i tabell 4.11b).

Tabell 4.11: Tabell a) viser ytelsen til de tre modellene uten vekselvirkningsegenskaper for responsen OS som hadde høyest ytelse under kryssvalidering testet på eksternt datasett. Tabell b) viser de tre modellene uten veskelvirkningsegenskaper med høyest ytelse basert på testingen på usett eksternt datasett. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av egenskaper brukt til prediksjon. Modellene med ytelse over baseline-ytelsen på $MCC = 0.201$ er markert i grønt.

a)							
Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC	
QDA	0.6	0.693	0.451	0.606	0.589	0.204	
QDA	0.9	0.734	0.514	0.657	0.639	0.324	
QDA	0.3	0.661	0.432	0.576	0.561	0.134	
b)							
Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC	
QDA	0.9	0.734	0.514	0.657	0.639	0.324	
Logistisk regresjon	0.9	0.72	0.521	0.646	0.631	0.294	
LightGBM	0.7	0.727	0.455	0.636	0.616	0.288	

Tabell 4.11a) viser at to av de tre modellene med høyest ytelse under kryssvalidering presterte bedre enn baseline-modellen på det eksterne datasettet. Baseline-ytelsen var på $MCC = 0.201$ for OS. Fra tabell 4.11b) kommer det frem at enkelte andre modeller, som ikke gjorde det like bra under kryssvalideringen, hadde en enda høyere ytelse enn modellene i tabell 4.4.4 på det eksterne datasettet.

De tre modellene presentert i tabell 4.10, var modellene med vekselvirkningsegenskaper som hadde høyest ytelse under kryssvalidering. Ytelsen til disse tre modellene på det eksterne datasettet er gitt i tabell 4.12a). De tre modellene med

vekselvirkningsegenskaper som hadde høyest ytelse på det eksterne datasettet, uavhengig av hvordan disse modellene presterte under kryssvalidering, er gitt i tabell 4.12b).

Tabell 4.12: Tabell a) viser ytelsen til de tre modellene med vekselvirkningsegenskaper for responsen DFS som hadde høyest ytelse under kryssvalidering på eksternt datasett er gitt i tabell a). Tabell b) viser de tre modellene med vekselvirkningsegenskaper med høyest ytelse basert på testingen på usett eksternt datasett. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av vekselvirkningsegenskaper brukt til prediksjon. Modellene med ytelse over baseline-ytelsen på $MCC = 0.201$ er markert i grønt.

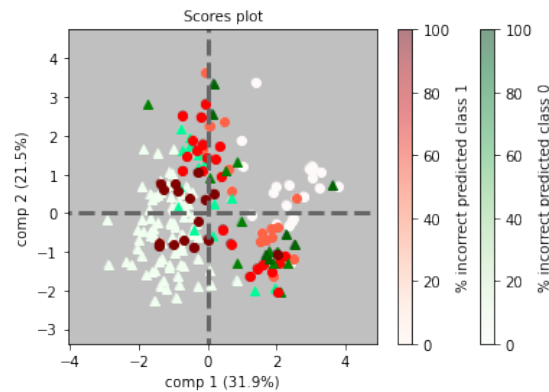
a)							
Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC	
QDA	0.5v	0.709	0.479	0.626	0.609	0.251	
QDA	0.2v	0.703	0.457	0.616	0.598	0.229	
QDA	0.4v	0.713	0.464	0.626	0.608	0.255	

b)							
Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC	
Logistisk regresjon	0.5v	0.734	0.514	0.657	0.639	0.324	
Logistisk regresjon	0.4v	0.730	0.528	0.657	0.640	0.318	
Logistisk regresjon	0.3v	0.720	0.521	0.646	0.631	0.294	

Tabell 4.12a) viser at alle tre modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering presterte bedre enn baseline-modellen på det eksterne datasettet. Fra tabell 4.12b) kommer det frem at enkelte andre modeller, som ikke gjorde det like bra under kryssvalideringen, hadde en enda høyere ytelse enn modellene presentert i tabell 4.12a) på det eksterne datasettet.

En PCA utført på undersettet av egenskaper uten vekselvirkninger med høyest ytelse under kryssvalidering for OS er illustrert i figur 4.4.5. Plottet viser at det var en glidende overgang mellom de to klassene, og pasientene som oftest ble

feilpredikerte befant seg i midten av plottet. Andelen feilpredikerte pasienter med sann klasse 1 var høyere enn andelen feilpredikerte klasse 0.



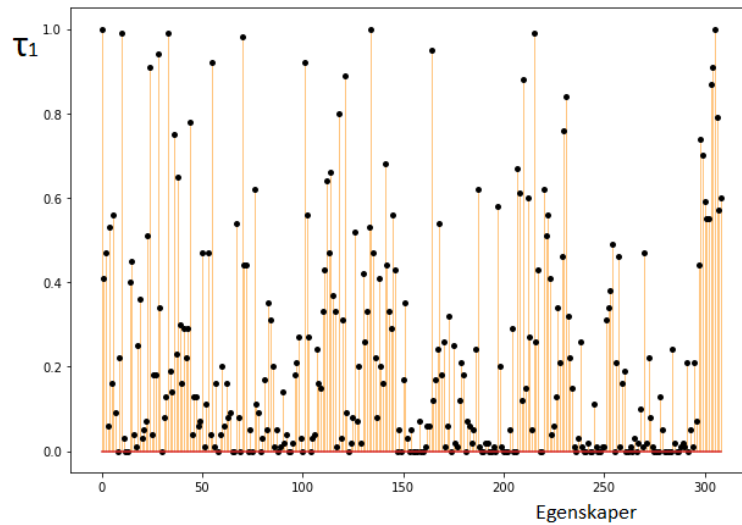
Figur 4.4.5: PCA av undersettet vs egenskaper som ga høyest ytelse under kryssvalidering. Sirkene representerer pasientene med sann klasse 1, og trekantene representerer pasientene med sann klasse 0. Fargeintensiteten symboliserer andelen ganger den aktuelle pasienten ble feilpredikert av modellene i RENT.

4.5 LRC

I dette delkapittelet presenteres egenskapsutvelging med og uten vekselvirknings-egenskaper, kryssvalidering og testing på eksternt datasett for responsen LRC.

4.5.1 Egenskapsutvelging

For LRC ble $C = 0.1$ og $L1_ratio = 0.9$ valgt som den optimale parameterkombinasjonen for elastisk-nett-regulariseringen i RENT. Frekvensplottet basert på de $K = 100$ modellene trent på ulike undersett av pasienter i RENT er vist i figur 4.5.1.



Figur 4.5.1: Frekvensplot som viser hvor ofte de ulike egenskapene ble valgt av modellene i RENT for responsen LRC. Egenskap 1-14 er formegenskaper, 15-116 er CT-egenskaper, 117-295 er PET-egenskaper, 296-298 er PET-parametere og 299-309 er kliniske egenskaper. τ_1 beskriver andelen av modeller i RENT som valgte den aktuelle egenskapen.

For LRC var det færre egenskaper som sjelden eller aldri ble valgt enn for DFS og OS, og flere som ble valgt ofte. Figur 4.3.1 viser at egenskapene som ble valgt oftest for LRC var en kombinasjon av formegenskaper, bildeegenskaper og kliniske egenskaper.

Undersettene av egenskaper ble valgt ut basert på ulike verdier av τ_1 og τ_2 , hvor $\tau_1 = \tau_2$ og varieres fra 0.2 til 0.9 i steg på 0.1. Egenskapene inkludert i undersett 0.5 til 0.9 er vist i figur 4.5.2. Undersett 0.2 inneholdt 108 egenskaper og en oversikt over undersett 0.2 til 0.4 finnes i tillegg D.

Egenskap	τ_1, τ_2							
	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
shape_Elongation	x	x	x	x	x	x	x	x
shape_Sphericity	x	x	x	x	x	x	x	x
first_order_MinimumCT_16	x	x	x	x	x	x	x	x
first_order_SkewnessCT_16	x	x	x	x	x	x	x	x
glszm_GrayLevelVarianceCT_16	x	x	x	x	x	x	x	x
glrlm_RunEntropyCT_16	x	x	x	x	x	x	x	x
gldm_DependenceNonUniformityNormalized_d_1CT_16	x	x	x	x	x	x	x	x
glcm_MCC_d_1CT_16	x	x	x	x	x	x	x	x
glszm_GrayLevelNonUniformityNormalizedPET_8	x	x	x	x	x	x	x	x
glrlm_ShortRunLowGrayLevelEmphasisPET_8	x	x	x	x	x	x	x	x
LBP_201PET_8	x	x	x	x	x	x	x	x
hpy_related	x	x	x	x	x	x	x	x
first_order_KurtosisPET_8	x	x	x	x	x	x	x	x
LBP_O21PET_8	x	x	x	x	x	x	x	x
glszm_SmallAreaHighGrayLevelEmphasisPET_4	x	x	x	x	x	x	x	x
larynx	x	x	x	x	x	x	x	x
glszm_LargeAreaHighGrayLevelEmphasisCT_16	x	x	x	x	x	x	x	x
glszm_ZoneEntropyCT_16	x	x	x	x	x	x	x	x
first_order_EnergyPET_8	x	x	x	x	x	x	x	x
glszm_SmallAreaEmphasisPET_4	x	x	x	x	x	x	x	x
histgrade_high	x	x	x	x	x	x	x	x
glszm_LowGrayLevelZoneEmphasisCT_16	x	x	x	x	x	x	x	x
glrlm_RunEntropyCT_16	x	x	x	x	x	x	x	x
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16	x	x	x	x	x	x	x	x
LBP_120CT_16	x	x	x	x	x	x	x	x
LBP_210CT_16	x	x	x	x	x	x	x	x
glszm_SizeZoneNonUniformityPET_8	x	x	x	x	x	x	x	x
glcm_ClusterTendency_d_1PET_8	x	x	x	x	x	x	x	x
glcm_SumSquares_d_1PET_8	x	x	x	x	x	x	x	x
shape_Maximum2DDiameterColumn	x	x	x	x	x	x	x	x
shape_Maximum2DDiameterSlice	x	x	x	x	x	x	x	x
ngtdm_Strength_d_1CT_16	x	x	x	x	x	x	x	x
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16	x	x	x	x	x	x	x	x
glcm_MaximumProbability_d_1CT_16	x	x	x	x	x	x	x	x
ngtdm_Contrast_d_1PET_8	x	x	x	x	x	x	x	x
glcm_Imc1_d_1PET_8	x	x	x	x	x	x	x	x
glszm_GrayLevelNonUniformityPET_4	x	x	x	x	x	x	x	x
age	x	x	x	x	x	x	x	x
female	x	x	x	x	x	x	x	x
oropharynx	x	x	x	x	x	x	x	x
charlson	x	x	x	x	x	x	x	x

Figur 4.5.2: Egenskapene i de ulike undersettene valgt ut ved bruk RENT for responsen LRC, basert på ulike verdier av terskelverdiene $\tau_1 = \tau_2$. Undersettet markert i grønt ga høyest MCC under kryssvalidering. Egenskaper markert i rødt ble fjernet grunnet høy korrelasjon med andre egenskaper.

4.5.2 Egenskapsutvelging med vekselvirkningsegenskaper

Egenskapene fra undersett 0.5 fra figur 4.5.2 ble kjørt på nytt gjennom RENT. Undersett 0.5 består av 41 egenskaper, dermed vil RENT kunne velge blant 1722 egenskaper med vekselvirkninger, fra likning 4.1. Det ble det hentet ut nye undersett av egenskaper for τ -verdier mellom 0.2 og 0.9 i steg på 0.1, og $\tau_1 = \tau_2$,

navngitt som τ -kriteriet for utvelgelse etterfulgt av en v. De tre modellene med høyest ytelse presenteres i tabell 4.14. Undersettet med vekselvirkningsegenskaper som ga den høyeste ytelsen gitt i MCC var undersøtt 0.5v, som er gitt i tillegg E.

4.5.3 Kryssvalidering

Hvert undersøtt av egenskaper med og uten vekselvirkninger fikk tilpasset hyperparametere for hver av de åtte klassifiseringsalgoritmene, og ble testet på nytt i en 5-foldet stratifisert kryssvalidering repetert 100 ganger. De gjennomsnittlige resultatene fra kryssvalideringene for undersettene uten vekselvirkningsegenskaper er visualisert i figur 4.5.3. Figur 4.5.3 viser at SVM med τ -verdier på 0.5 ga den høyeste ytelsen for LRC under kryssvalidering. Dette undersøttet besto av 40 egenskaper, og τ -verdiene var således mindre restriktive for LRC enn for DFS og OS. De tre kombinasjonene av klassifiseringsalgoritmer og undersøtt av egenskaper uten vekselvirkninger som ga høyest ytelse under kryssvalideringen er gitt i tabell 4.13.

Tabell 4.13: De tre modellene uten vekselvirkningsegenskaper for responsen LRC med høyest ytelse basert på 5-foldet stratifisert kryssvalidering repetert 100 ganger. Første linje for hver modell viser den gjennomsnittlige ytelsen, mens den andre linjen viser standardavviket til modellen. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av egenskaper brukt i kryssvalideringen.

Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
SVM	0.5	0.555	0.861	0.789	0.707	0.428
		0.135	0.043	0.062	0.086	0.171
Logistisk regresjon	0.9	0.477	0.875	0.800	0.661	0.394
		0.148	0.032	0.050	0.075	0.169
Logistisk regresjon	0.5	0.513	0.861	0.785	0.681	0.39
		0.140	0.040	0.058	0.083	0.170



Figur 4.5.3: Visualisering av den gjennomsnittlige ytelsen, gitt i F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, de ulike kombinasjonene av undersett av egenskaper, gitt ved τ -verdien, hvor $\tau_1 = \tau_2$, og klassifiseringsalgoritme a) til f) oppnådde under 5-foldet stratifisert kryssvalidering repetert 100 ganger for responsen LRC.

Modellen med høyest gjennomsnittlig ytelse var basert på SVM og undersettet der

τ_1 og τ_2 var lik 0.5. Modellen med nest høyest ytelse, logistisk regresjon basert på undersettet der τ_1 og τ_2 var lik 0.9, har vesentlig lavere ytelse, som viser at for LRC er det ugunstig å ekskludere for mange egenskaper. Usikkerheten i LRC-modellene er høyere enn for DFS- og OS-modellene.

De tre modellene med vekselvirkningsegenskaper som hadde høyest gjennomsnittlig verdi av MCC under kryssvalidering er gitt i tabell 4.14.

Tabell 4.14: De tre modellene med vekselvirkningsegenskaper for responsen DFS med høyest ytelse basert på 5-foldet stratifisert kryssvalidering repetert 100 ganger. Første linje for hver modell viser den gjennomsnittlige ytelsen, mens den andre linjen viser standardavviket til modellen. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av vekselvirkningsegenskaper brukt i kryssvalideringen.

Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
SVM	0.5v	0.715	0.918	0.873	0.803	0.647
		0.108	0.031	0.046	0.071	0.132
Logistisk regresjon	0.5v	0.714	0.915	0.869	0.807	0.640
		0.114	0.033	0.051	0.077	0.142
SVM	0.4v	0.709	0.913	0.867	0.803	0.632
		0.119	0.035	0.052	0.078	0.148

Modellene med vekselvirkningsegenskaper oppnådde en MCC som var høyere enn modellene uten vekselvirkningsegenskaper under kryssvalidering, med en lavere usikkerhet. Disse modellene var også baserte på undersett av egenskaper som inneholdt flere egenskaper enn modellene uten vekselvirkningsegenskaper med høyest ytelse.

4.5.4 Testing på eksternt datasett

Samtlige modeller bygget på OUS-datasettet ble testet på det usette eksterne datasettet fra MAASTRO-klinikken. Modellytelsen ble validert med F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, og er visualisert i figur 4.5.4.



Figur 4.5.4: Visualisering av ytelsen, gitt i F1-score for begge klasser, nøyaktighet, ROC AUC og MCC, de ulike kombinasjonene av undersett av egenskaper, gitt ved τ -verdien, hvor $\tau_1 = \tau_2$, og klassifiseringsalgoritme a) til f) oppnådde ved testing på eksternt data-sett for responsen LRC.

De tre modellene presentert i tabell 4.13, var modellene uten vekselvirkningsegen-

skaper som hadde høyest ytelse under kryssvalidering for OUS-datasettet. Ytelsen til disse tre modellene på det eksterne datasettet er gitt i tabell 4.15a). De tre modellene uten veskelvirkningsegenskaper som hadde høyest ytelse på det eksterne datasettet, uavhengig av hvordan disse modellene presterte under kryssvalidering, er presentert i tabell 4.15b).

Tabell 4.15: Tabell a) viser ytelsen til de tre modellene uten vekselvirkningsegenskaper for responsen LRC som hadde høyest ytelse under kryssvalidering på eksternt datasett. Tabell b) viser de tre modellene uten veskelvirkningsegenskaper med høyest ytelse basert på testingen på usett eksternt datasett. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av egenskaper brukt til prediksjon.

a)							
Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC	
SVM	0.5	0.241	0.686	0.556	0.471	-0.051	
Logistisk regresjon	0.9	0.373	0.618	0.525	0.565	0.108	
Logistisk regresjon	0.5	0.281	0.657	0.535	0.490	-0.016	

b)							
Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC	
AdaBoost	0.2	0.348	0.803	0.700	0.578	0.151	
QDA	0.9	0.389	0.651	0.556	0.584	0.140	
SVM	0.6	0.207	0.864	0.768	0.542	0.137	

Tabell 4.15a) viser at ingen av de tre modellene med høyest ytelse under kryssvalidering presterte bedre enn baseline-modellen på det eksterne datasettet. Baseline-ytelsen var på $MCC = 0.182$ for LRC. Fra tabell 4.15b) kommer det frem at ingen av de resterende modellene heller hadde en ytelse over baseline-ytelsen på det eksterne datasettet.

De tre modellene presentert i tabell 4.14, var modellene med vekselvirkningsegenskaper som hadde høyest ytelse under kryssvalidering. Ytelsen til disse tre modellene på det eksterne datasettet er gitt i tabell 4.16a). De tre modellene med

vekselvirkningsegenskaper som hadde høyest ytelse på det eksterne datasettet, uavhengig av hvordan disse modellene presterte under kryssvalidering, er gitt i tabell 4.16b).

Tabell 4.16: Tabell a) viser ytelsen til de tre modellene med vekselvirkningsegenskaper for responsen LRC som hadde høyest ytelse under kryssvalidering testet på eksternt datasett. Tabell b) viser de tre modellene med vekselvirkningsegenskaper med høyest ytelse basert på testingen på usett eksternt datasett. Undersettet spesifiserer hvilke verdier av $\tau_1 = \tau_2$ som definerer undersettet av vekselvirkningsegenskaper brukt til prediksjon.

a)

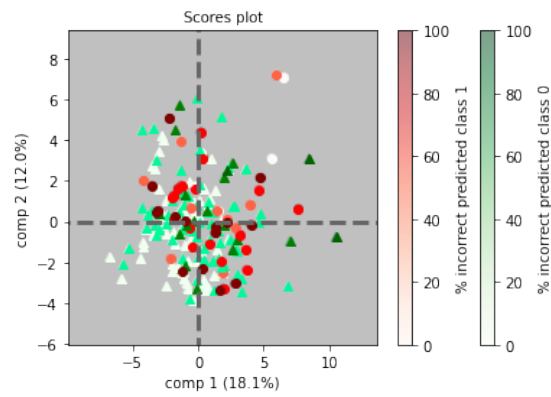
Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
SVM	0.5v	0.357	0.746	0.636	0.571	0.125
Logistisk regresjon	0.5v	0.328	0.701	0.586	0.539	0.066
SVM	0.4v	0.250	0.704	0.576	0.484	-0.028

b)

Kl.algoritme	Undersett	F1:1	F1: 0	Nøyaktighet	ROC AUC	MCC
KNN	0.6v	0.373	0.734	0.626	0.581	0.140
Logistisk regresjon	0.5v	0.328	0.701	0.586	0.539	0.066
LightGBM	0.9v	0.361	0.715	0.606	0.568	0.116

Tabell 4.16 a) viser at ingen av de tre modellene med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering presterte bedre enn baseline-modellen på det eksterne datasettet. Fra tabell 4.16b) kommer det frem at heller ingen av de øvrige modellene hadde en ytelse høyere enn baseline-ytelsen på $MCC = 0.182$.

En PCA utført på undersettet av egenskaper uten vekselvirkninger med høyest ytelse under kryssvalidering for LRC er illustrert i figur 4.5.5. Plottet viser at det ikke finnes noe tydelig skille mellom klassene, og at de ser ut til å være vanskelige å separere.



Figur 4.5.5: PCA av undersettet av egenskaper som ga høyest ytelse under kryssvalidering. Sirklene representerer pasientene med sann klasse 1, og trekantene representerer pasientene med sann klasse 0. Fargeintensiteten viser andelen ganger den aktuelle pasienten ble feilpredikert av RENT.

Kapittel 5

Diskusjon

Formålet med oppgaven var å bruke radiomics-egenskaper beregnet fra PET- og CT-avbildninger sammen med klinisk data til å predikere behandlingsutfallet for pasienter med hode- og halskreft. Samtidig var det et mål å avdekke hvilke egenskaper som var de viktigste for prediksjonen av de tre ulike behandlingsutfallene; sykdomsfri overlevelse (DFS), generell overlevelse (OS) og lokalregionalt tilbakefall (LRC). Dette kapittelet tar for seg eventuelle utfordringer knyttet til de ulike aspektene av oppgaven, og ser på mulige løsninger på disse problemene.

5.1 Datasettet

Datasettet fra OUS besto av 197 pasienter behandlet for hode- og halskreft i tidsrommet 2007-2013. HPV-status ble plukket ut som en av de viktigste egenskapene for å predikere behandlingsutfallet, men for hele 52 av pasientene var HPV-status ukjent. Med relativt få pasienter i datasettet ville det være ugunstig å fjerne enda flere, og derfor ble pasientene med ukjent HPV-status tildelt en egen kategori. For det eksterne datasettet brukt til testing var HPV-status til alle pasientene kjent. Et alternativ til å tildele en ny klasse til pasientene med ukjent HPV-status kunne vært å trene en modell som predikerte denne. Flere studier har gjort dette tidligere, og oppnådd gode resulater [13]. Eksempelvis oppnådde Bogowicz et al. [87]. en AUC på 0.85 for treningssettet og 0.78 for valideringssettet. Her ble en modell basert på CT-radiomics-egenskaper og logistisk regresjon brukt.

Datasettet kunne også inneholde pasienter som døde av andre årsaker enn kreft i

løpet av oppfølgingsperioden, som kan ha hatt en innvirkning på modellene [63]. En annen faktor som spiller inn er onkologenes inntegning av svulstene. Inntegning av svulstene vil kunne variere mellom onkologer, og Giraud et al. [88] viste at det også er en signifikant forskjell mellom inntegningen ved ulike behandlingsfasiliteter, som kan ha bidratt til at modellene trent på OUS-datasettet ikke fungerte like godt på datasettet fra MAASTRO. Mori et al. [89] undersøkte dette for kreft i bukspyttkjertelen, og konkluderte med at CT-radiomics-egenskaper var robuste mot usikkerhet i inntegning. Belli et al. [90] viste at PET-radiomics-egenskapene også var robuste mot ulike inntegninger, ved å sammenligne egenskaper uthentet fra bilder med semi-automatisk svulstinntegning med egenskapene uthentet fra bilder med manuell svulstinntegning.

OUS-datasettet har tidligere vært brukt i to andre masteroppgaver ved Norges miljø- og biovitenskapelige universitet. Midtfjord [91] predikerte behandlingsutfall for lokalregionalt tilbakefall og progresjonsfri overlevelse, mens Langberg [84] identifiserte biomarkører for sykdomsfri overlevelse. Både Midtfjord og Langberg brukte andre egenskapsutvelgere enn RENT, og hadde ikke et eksternt datasett å validere modellene på.

5.2 RENT som egenskapsutvelger

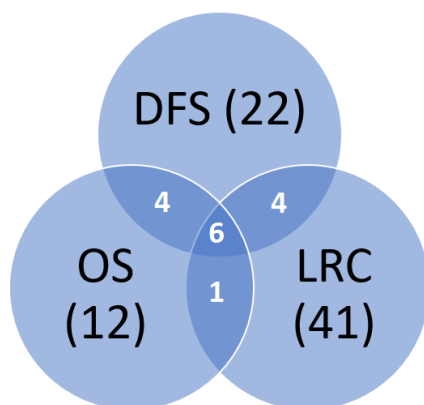
I denne oppgaven ble egenskapene i de ulike undersettene selektert av RENT. En av fordelene med RENT er at det også tas hensyn til stabiliteten i egenskapsutvelgingen, mens andre egenskapsutvelgere fokuserer utelukkende på den prediktive ytelsen til modellene [71]. Stabiliteten ivaretas ved at egenskapsutvelgingen repeteres for et ensemble av modeller, og ved at det er mulig å justere tre ulike kriterier for utvelgelse. Ved å bruke RENT er det også mulig å identifisere og samle informasjon om objekter i dataen som ofte blir feilpredikerte [71]. For å ta hensyn til ubalanserte klassedistribusjoner bruker RENT både F1-score og MCC som evalueringsmetrikk.

RENT er en ny egenskapsutvelger, og er lite utprøvd sammenliknet med mer etablerte egenskapsutvelgere. Selv om RENT har vist god ytelse på tvers av flere datasett [71], er det fortsatt et behov for mer testing før det kan konkluderes med om RENT er en konkurransedyktig egenskapsutvelger. Enkelte av egenskapene som RENT valgte ut i denne oppgaven viste seg å være høyt korrelerte. For å fjerne overflødige egenskaper var det derfor nødvendig med et eget steg som fjernet høyt korrelerte egenskaper, som vist i figur 3.5.1. En annen ulempe med RENT er at den beregningsmessige kostnaden er høy. RENT er per nå ikke tilgjengelig for klassifi-

sering av multiple klasser, som kan løses ved å bruke «en mot alle»-tilnærmingen [71]. Dette vil igjen medføre en høy beregningsmessig kostnad.

5.3 Evaluering av utvalgte egenskaper

Fra frekvensplottene, figur 4.3.1, 4.4.1 og 4.5.1, kommer det tydelig frem at antallet egenskaper som velges er veldig ulikt for de tre responsene, men også at et stort antall egenskaper aldri blir selektert. Dette underkapittelet fokuserer på egenskapene som ble valgt av RENT minst 50% av gangene, det vil si når $\tau_1 > 0.5$. Definisjonen av egenskapene er hentet fra dokumentasjonen til PyRadiomics [81]. Enkelte av de utvalgte egenskapene var viktige for flere av responsene. Fellestrekkene mellom egenskapene valgt for de ulike responsene er vist i figur 5.3.1.



Figur 5.3.1: Diagrammet viser fellestrekk mellom egenskapene som ble valgt i minst 50% av tilfellene for de ulike responsene.

For DFS var det 22 egenskaper som ble valgt ut mer enn halvparten av gangene. Undersøttet uten vekselvirkninger som ga den høyeste ytelsen under kryssvalidering inkluderte fem formegenskaper, fire kliniske egenskaper og fem PET-radiomics-egenskaper. De fire kliniske egenskapene var også representerte i baseline-modellen. Ingen CT-egenskaper eller PET-parametere ble valgt av over 50% av modellene i RENT. I sin studie predikerte Moan et al. [63] DFS ved hjelp av kliniske egenskaper og PET/CT-parametere, og fant heller ikke at PET-parametere var signifikante for prediksjon av DFS. Alle de kliniske egenskapene valgt i denne oppgaven viste også en signifikant sammenheng med DFS i Moan et al. sin studie.

For OS var det 12 egenskaper som ble valgt ut av mer enn halvparten av modelle-

ne. Undersettet uten vekselvirkninger som ga den høyeste ytelsen under kryssvalidering besto av to formegenskaper, fire kliniske egenskaper og en CT-egenskap. *Oropharynx* ble valgt som en av de kliniske egenskapene, men var ikke inkludert i baseline-modellen, der *cavum_oris* og *wicc_III – IV* ble valgt ut som viktigere for prediksjonen. Ingen PET-parametere ble valgt av mer enn 50% av modellene i RENT for OS, som ikke stemmer over ens med Moan et al. sin studie [63], der flere PET-parametere viste sammenheng med overlevelse. Vallieres et al. [12] fikk den beste prediksjonen av OS ved å kun benytte kliniske egenskaper, etterfulgt av en modell som benyttet en kombinasjon av radiomics-egenskaper fra PET og kliniske egenskaper. I denne oppgaven var ingen PET-radiomics-egenskaper inkludert i undersettet som ga den beste prediksjonen.

Fire egenskaper ble valgt av over 50% av modellene for både DFS og OS, men ikke for LRC. Disse var formegenskapene *shape_Maximum3DDiameterCT_16*, og *shape_MajorAxisLength*, som begge er mål på svulstens størrelse, den kliniske egenskapen *pack_years* og teksturegenskapen *glszm_GrayLevelVariancePET_4*, som forklarer variansen i gråtonene for sonene i PET-bildene diskretisert til 4 bins.

For LRC var det 41 egenskaper som ble valgt ut av mer enn halvparten av modellene i RENT. Undersettet av egenskaper som ga den høyeste ytelsen under kryssvalidering besto av fire formegenskaper, 13 PET-egenskaper, 16 CT-egenskaper og seks kliniske egenskaper. For LRC ble ikke egenskapen *pack_years* selektert, enda denne var blant de mest valgte egenskapene for DFS og OS, og var inkludert i baseline-modellen for LRC. De seks kliniske egenskapene i undersettet som ledet til beste prediksjon var alle inkluderte i baseline-modellen for LRC. En kombinasjon av kliniske egenskaper og radiomics-egenskaper ga en bedre prediksjonen av behandlingsutfall for LRC enn undersettene med høyere andel radiomics-egenskaper, noe som også ble poengtert av Vallieres et al. [12].

Fire egenskaper ble valgt for både DFS og LRC, men ikke for OS. Disse var *shape_Elongation*, *glcm_MCC_d_1CT_16* som er et mål på kompleksiteten til teksturen, *glrlm_ShortRunLowGrayLevelEmphasisPET_8* som er en felles sannsynlighetsdistribusjon for kortere intensitetsrekker med lavere gråtoneverdier, og *glszm_SmallAreaHighGrayLevelEmphasisPET_4* som er en felles sannsynlighetsdistribusjon for mindre soner med høyere gråtoneverdier. Med unntak av *shape_elongation* var alle disse teksturegenskaper, og at disse ikke ble valgt for OS stemmer over ens med at bildeegenskaper sjelden ble valgt av RENT for denne responsen. Egenskapen *charlson* ble valgt for OS og LRC, men ikke for DFS.

Seks egenskaper ble valgt for alle tre responsene: *shape_Sphericity*, *hpv_related*, *first_order_SkewnessCT_16_age*, *oropharynx* og *larynx*. Dette betyr at for mo-

dellene var pasientens HPV-status og alder, svulstens rundhet og skjevhet i histogrammet fra CT-bildene viktige prediktorer for alle responsene, i tillegg til om kreften var lokalisert i strupehodet eller munnsvelget. Moan et al. [63] undersøkte de samme tre responsene på samme datasett, men inkluderte kun kliniske egenskaper og PET/CT-parametere. I Moans studie ble HPV-status og *oropharynx* trukket frem som kliniske egenskaper viktige for alle tre responsene, i tillegg til *cavum_oris* og *wicc8_III – IV*. I Midtjords [91] oppgave ble flere formegenskaper trukket ut som viktige for prediksjonen av LRC, men *sphericity* var ikke en av dem. Langberg [84] derimot, plukket ut *sphericity* som en viktig egenskap for prediksjonen av DFS.

5.4 Evaluering av resultater

5.4.1 DFS

Modellen som hadde høyest gjennomsnittlig ytelse under kryssvalideringen for DFS var SVM på undersettet basert på τ_1 og τ_2 lik 0.8, med en MCC på 0.490 ($AUC = 0.741$). Når denne modellen ble testet på det eksterne datasettet oppnådde den kun en MCC på 0.064 ($AUC = 0.529$), som viser en vesentlig reduksjon i ytelse fra kryssvalidering til ekstern validering. Ingen av modellene uten vekselvirkninger presterte over baseline-ytelsen for DFS, som var på $MCC = 0.172$. F1-score antydte at modellen gjorde en god jobb med å predikere klasse 1 (tilbakefall) for det eksterne datasettet (F1: 1 = 0.667), men slet mer med klasse 0 (sykdomsfri overlevelse, F1: 0 = 0.377). Under kryssvalideringen gjorde modellen en jevn god jobb med å predikere klassene, men var best på å klassifisere klasse 0. En mulig årsak til dette er at klassedistribusjonen er annerledes i det eksterne datasettet.

Vekselvirkningsmodellen med høyest ytelse var logistisk regresjon på undersettet basert på τ_1 og τ_2 lik 0.9 med vekselvirkninger, 0.9v. Denne modellen hadde en gjennomsnittlig MCC på 0.564 ($AUC = 0.778$) med en usikkerhet som var tilnærmet det samme som for modellen uten vekselvirkningsegenskaper. På det eksterne datasettet hadde denne vekselvirkningsmodellen en MCC på 0.190 ($AUC = 0.577$). I motsetning til modellen uten vekselvirkning presterte denne over baseline-ytelsen på. Også vekselvirkningsmodellene presentert i tabell 4.5 var vesentlig dårligere til å predikere klasse 0. Vekselvirkningsegenskapene er mer komplekse enn de ordinære egenskapene, og den marginale økningen i ytelse sammenliknet med baseline-modellen går på bekostning av tolkbarhet.

I sin oppgave oppnådde Langberg [84] en AUC på 0.764 for prediksjonen av DFS på OUS-datasettet for pasientene med negativ HPV-status. For pasientene med positiv HPV-status oppnådde han en AUC på 0.690. Resultatene under kryssvalidering i denne studien tilsvarer den samme ytelsen Langberg fikk for de HPV-negative pasientene, for hele pasientgruppen.

5.4.2 OS

Modellen som hadde høyest gjennomsnittlig ytelse under kryssvalideringen for OS var QDA på undersettet basert på τ_1 og τ_2 lik 0.6, med en MCC på 0.527 ($AUC = 0.754$). Når denne modellen ble testet på det eksterne datasettet oppnådde den kun en MCC på 0.204 ($AUC = 0.589$), mens modellen som gjorde det tilnærmet like godt under kryssvalidering, QDA på undersettet basert på τ_1 og τ_2 lik 0.2, oppnådde en MCC på 0.324 ($AUC = 0.639$) på det eksterne datasettet. Undersettet basert på $\tau_1 = \tau_2 = 0.6$ inneholdt sju egenskaper, mens undersettet basert på $\tau_1 = \tau_2 = 0.2$ inneholdt 22 egenskaper. Forskjellen i ytelsen på det eksterne datasettet for disse undersettene skyldes trolig at en eller flere av egenskapene som har blitt eliminert i undersettet basert på $\tau_1 = \tau_2 = 0.6$ inneholdt informasjon som var viktig for prediksjonen av MAASTRO-pasientene. Ingen av modellene uten vekselvirkningsegenskaper presterte over baseline-ytelsen, som var på 0.207 (MCC) for OS. Også her var modellene best til å predikere klasse 0 (overlevelse) under kryssvalideringen, mens for den usette dataen falt prediksjonen av klasse 0 kraftig. Prediksjonen av klasse 1 (død) var tilnærmet like god under kryssvalidering og eksternt validering.

Vekselvirkningsmodellen med høyest ytelse for OS var QDA på undersettet med vekselvirkningsegenskaper basert på τ_1 og τ_2 lik 0.5. På det eksterne datasettet oppnådde den beste vekselvirkningsmodellen en MCC på 0.251 ($AUC = 0.609$), som er vesentlig høyere enn baseline. Vekselvirkningsmodellene kan være vanskelige å tolke, men for OS oppnådde de vesentlig høyere resultat under både kryssvalidering og validering på eksternt datasett enn de enklere modellene. Vallieres et al. [12] fant at for OS var det en modell bestående av kun kliniske egenskaper som hadde høyest ytelse ved validering på et eksternt datasett, som oppnådde en AUC på 0.78. I denne oppgaven hadde modellen med høyest AUC en ytelse på 0.609, som er vesentlig lavere. Dette kan skyldes ulik klassedistribusjon i datasett brukt til trening og testing eller ulikheter i egenskapene i de ulike datasettene.

5.4.3 LRC

Modellen som hadde høyest gjennomsnittlig ytelse under kryssvalidering for LRC var SVM på undersettet basert på τ_1 og τ_2 lik 0.5, med en MCC på 0.428 ($AUC = 0.707$). Denne modellen hadde en usikkerhet på 0.171, som var høyere enn de beste modellene for DFS og OS. Når denne modellen ble testet på det eksterne datasettet oppnådde den kun en MCC på -0.051 ($AUC = 0.471$), som er dårligere enn tilfeldig gjetning. Modellen med den nest beste ytelsen under kryssvalidering presentert i tabell 4.13, logistisk regresjon på undersettet basert på τ_1 og τ_2 lik 0.9, oppnådde derimot en MCC på 0.108 ($AUC = 0.565$). Dette er allikevel et stykke under baseline-ytelsen på $MCC = 0.182$ for LRC. Fra F1-score for begge klasser kommer det frem at modellen gjør en god jobb med å predikere klasse 0, men sliter mer med klasse 1, altså motsatt av for DFS og OS. Dette er også tilfelle under kryssvalideringen, der prediksjonen av klasse 1 ligger rundt 0.5 for alle de tre beste modellene. For LRC representerer klasse 1 lokalregionalt tilbakefall. PCA-plottet i figur 4.5.5 viser at det ikke er noe tydelig skille mellom pasientene i hver klasse, som forklarer hvorfor de var vanskelige å predikere korrekt. I denne modellene ble ingen PET-parametere selektert, mens for Moan et al. [63] var en MTV-egenskap signifikant for prediksjonen av LRC.

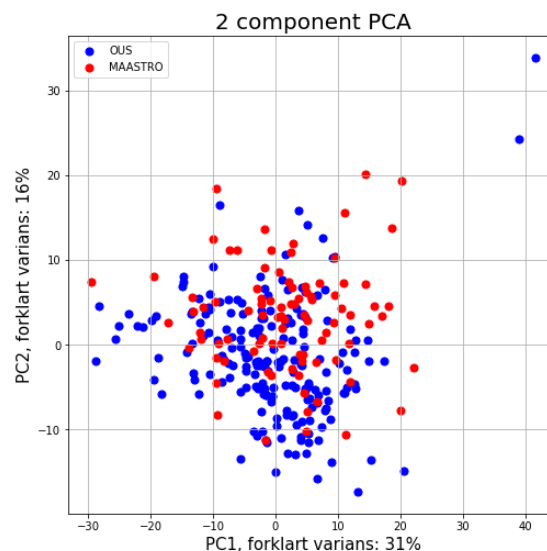
Vekselvirkningsmodellen med høyest ytelse for LRC var SVM på undersettet basert på τ_1 og τ_2 lik 0.5 med vekselvirkningsegenskaper. For MCC hadde denne modellen en usikkerhet på 0.132, som er vesentlig lavere enn modellene uten vekselvirkningsegenskaper. På det eksterne datasettet oppnådde den beste vekselvirkningsmodellen en MCC på 0.125 ($AUC = 0.571$), som er så vidt høyere enn den beste modellen uten vekselvirkning, men fortsatt et godt stykke under baseline-ytelsen. Også disse modellene er vesentlig dårligere på å predikere klasse 1, men tilnærmet like gode til å predikere klasse 0 som under kryssvalidering. For LRC var det dermed ingen modeller som hadde en ytelse høyere enn baseline-ytelsen på det eksterne datasettet. Vallieres et al. [12] oppnådde en AUC på 0.69 for LRC ved testing på eksternt datasett, som er vesentlig høyere enn høyeste AUC i denne studien, som var på 0.581. Dette kan muligens spores tilbake i ulikheter i datasettet brukt til trening og testing, som diskuteres i seksjon 5.5.

I sin oppgave oppnådde Midtfjord [91] en AUC på 0.67 for LRC under kryssvalidering. Her var det hovedsakelig formegenskaper som var viktige for prediksjonen av LRC. Midtfjord oppnådde høyest ytelse med egenskapsutvelgeren ReliefF. Den høye ytelsen oppnådd under kryssvalidering i denne oppgaven kan tyde på at RENT valgte ut egenskaper som inneholdt en større andel relevant informasjon for prediksjonen av LRC. Vallieres et al. [12] pekte ut tre radiomics-egenskaper som

spesielt signifikante for prediksjon av LRC, deriblant *glszm_LowGrayLevelZoneEmphasisCT_16*, som i denne oppgaven ble valgt ut av RENT i over 60% av tilfellene. Denne egenskapen er et mål på distribusjonen av lavere sonestørrelser i CT-bildene diskretisert til 16 bins.

5.5 Ulikheter i datasett brukt til trening og testing

For å forstå hvorfor ytelsen til modellene falt kraftig fra kryssvalidering til validering på det eksterne datasettet var det viktig å identifisere eventuelle ulikheter mellom datasettene. En PCA ble utført på et datasett som inkluderte samtlige pasienter fra begge datasett og samtlige egenskaper. Denne analysen er visualisert i figur 5.5.1.

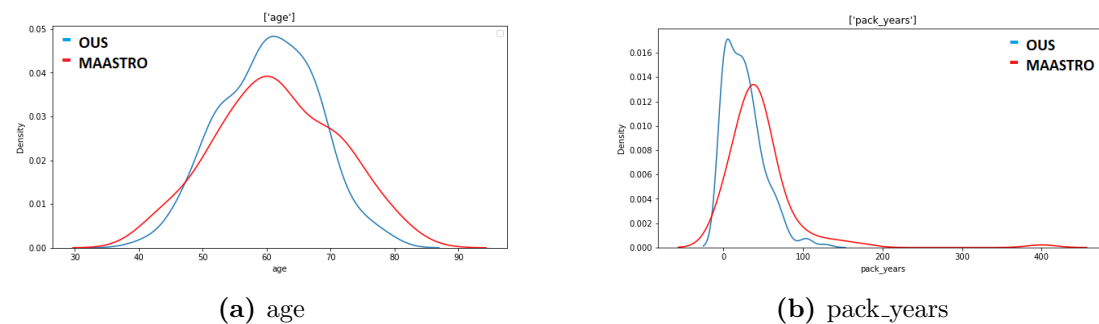


Figur 5.5.1: PCA utført på et datasett sammensatt av alle pasienter fra OUS og MAASTRO. Pasienter fra OUS er markert i blått, mens pasienter fra MAASTRO er markert i rødt.

Fra figur 5.5.1 var det ikke mulig å skille pasientene fra de ulike behandlingstilstandene fra hverandre da distribusjonene av pasienter overlappet, men det vistes en tendens til at pasientene fra OUS trakk seg mot de lave verdiene av prinsipalkomponent 1 og prinsipalkomponent 2, mens pasientene fra MAASTRO hadde høyere

verdier for begge prinsipalkomponenter. Det finnes enkelte utliggere, som de to OUS-pasientene øverst i høyre hjørne, som strider veldig i mot disse tendensene. Prinsipalkomponent 1 og 2 forklarer kun 47% av variansen i dataen, og analysen viser ikke den fulle sannheten selv om den danner et grunnlag for videre analyse.

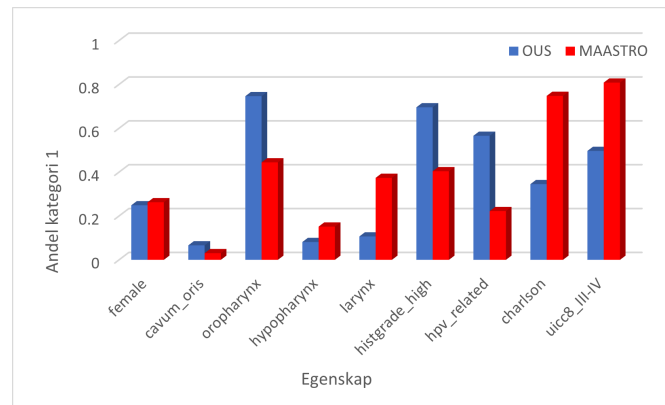
Figur 5.5.2 viser distribusjonen av de kontinuerlige egenskapene *age* og *pack_years* for datasettene fra OUS og MAASTRO. *age* ble valgt ut som egenskap i alle baseline-modellene, og i alle modellene som hadde høyest ytelse for hver respons. *pack_years* var også en mye valgt egenskap, men var ikke inkludert i den beste modellen for LRC.



Figur 5.5.2: Distribusjonen av variablene *age* og *pack_years* for OUS-datasettet (blått) og MAASTRO-datasettet (rødt).

Gjennomsnittsalderen til pasientene ved OUS og MAASTRO var tilnærmet den samme, men bredden av aldre var noe større for MAASTRO - datasettet. Distribusjonen av *pack_years* viser at pasientene fra MAASTRO i gjennomsnitt hadde røyket mer enn pasientene fra OUS, og at det også var flere som hadde røyket. En av pasientene fra MAASTRO-datasettet hadde røyket vesentlig mer enn de andre pasientene. Denne utliggeren kunne blitt håndtert gjennom transformering av dataen, for eksempel ved Box-Cox-transformering eller Yeo-Johnson-transformering [92], [93].

Figur 5.5.3 viser gjennomsnittsverdiene til de ulike binære variablene; det vil si prosentandelen pasienter tildelt kategori 1 for de ulike egenskapene.

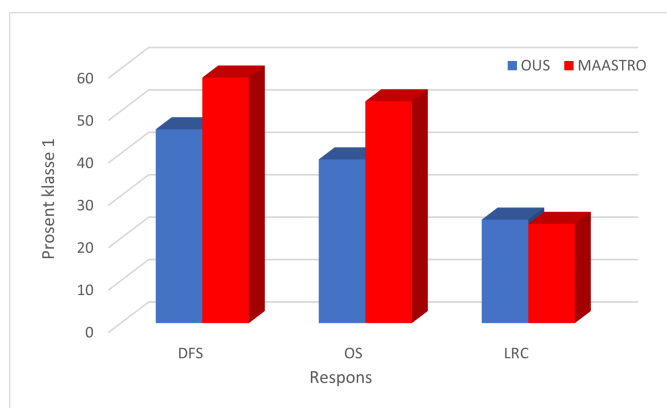


Figur 5.5.3: Gjennomsnittsverdien til de binære kliniske egenskapene i OUS-datasett (blå) og MAASTRO-datasett (rød).

cavum_oris, *oropharynx*, *hypopharynx* og *larynx* forteller hvor i hodet eller halsen kreften er lokalisert. Blant pasientene behandlet ved OUS hadde nesten 80% kreft i munnsvelget (*oropharynx*). Bruixola et al. [13] viste at denne krefttypen er ofte knyttet til HPV-infeksjon, men den kan også skyldes tobakk eller andre årsaker [15]. I MAASTRO-datasettet var det flere pasienter som hadde kreft i strupen (*larynx*). Denne krefttypen er sterkt knyttet til tobakksbruk, som kan spores tilbake til forskjellene i distribusjon av *pack_years*, og også påvirke klassedistribusjonene [94].

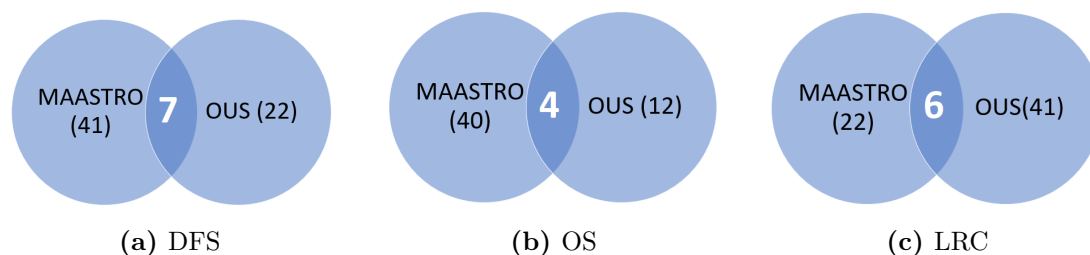
Andelen HPV-positive pasienter var høyere for OUS, noe som kan ha en sammenheng med den høye andelen pasienter med kreft i *oropharynx*. Dette viser ikke hele bildet av sannheten, da en stor andel pasienter manglet HPV-status i OUS-datasettet. Tilsvarende var tilfellet for *uicc8_III – IV*, der flere pasienter lå innenfor stadium 3-4 for MAASTRO-datasettet, mens rundt en fjerdedel av OUS-pasientene ikke hadde registrert stadium. En annen årsak til at flere pasienter hadde lavere kreftstadium for OUS-datasettet kan være den høye andelen pasienter med kreft i *oropharynx*. Kreft i *oropharynx* har bedre prognose, og dermed blir også inndelingen av kreftstadium annerledes, forklart av O’Sullivan et al. [95].

I seksjon 5.4 ble muligheten for at den lave ytelsen ved ekstern validering kunne skyldes ulikhet i klassedistribusjonene for treningsdataen og testdataen presentert. Figur 5.5.4 viser prosentandelen pasienter tildelt klasse 1 i OUS-datasettet og MAASTRO-datasettet.



Figur 5.5.4: Klassedistribusjonen for OUS-data (blå) og MAASTRO-data (rød). Plottet viser prosentandelen pasienter tildelt klasse 1, som representerer tilbakefall for DFS, død for OS og lokalregionalt tilbakefall for LRC.

For å se nærmere på ulikheten i egenskapene for de to settene ble RENT brukt på MAASTRO-datasettet, for å sammenlikne de valgte egenskapene med egenskapene som ble valgt for OUS-datasettet. Figur ?? viser hvor mange av de samme egenskapene som ble valgt fra MAASTRO- og OUS-datasettene mer enn 50% av gangene.



Figur 5.5.5: Oversikt over antall egenskaper som ble valgt i minst 50% av tilfellene for hver av de tre ulike responsene for OUS- og MAASTRO-datasettene.

Felles egenskaper for DFS var formegenskapene *shape_Elongation* og *shape_Sphericity*, teksturegenskapene *glrlm_ShortRunLowGrayLevelEmphasisPET_8* og *glszm_GrayLevelVariancePET_4* og de kliniske egenskapene *cavum_oris*, *larynx* og *pack_years*. For OS var felles formegenskap *shape_Sphericity* og felles kliniske egenskaper *hpv_related*, *age* og *oropharynx*. For LRC var de felles egenskapene formegenskapen *shape_Sphericity*, de kliniske egenskapene *female* og *charlson* og teksturegenskapene *glszm_GrayLevelVariance*, *LBP_021PET_8*, *LBP_210CT_16*. *shape_sphericity*, svulstens rundhet, står frem som en viktig egenskap for prediksjø-

nen av alle responsene for begge datasett. Definisjon av de ulike formegenskapene og tekstoregenskapene finnes i tillegg B.

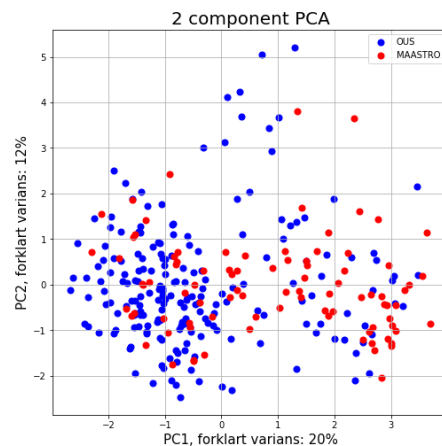
For DFS og OS valgte RENT vesentlig flere egenskaper mer enn 50% av gangene for MAASTRO-datasettet enn for OUS-datasettet. Egenskapene som ble valgt felles for DFS var to formegenskaper, to PET-egenskaper og tre kliniske egenskaper. Egenskapene *shape_Maximum3DDiameter*, *hpv_related*, *age* og *pack_years* som alle ble valgt av over 90% av modellene for OUS-datasettet ble ikke valgt for MAASTRO-datasettet. Av de 41 valgte egenskapene var kun tre formegenskaper og fem kliniske egenskaper. Blant de 22 egenskapene valgt fra OUS-datasettet var fem egenskaper formegenskaper og seks egenskaper kliniske. Bildeegenskapene spilte dermed en viktigere rolle for MAASTRO-datasettet enn for OUS-datasettet.

Egenskapene som ble valgt felles for OS var en formegenskap og tre kliniske egenskaper. Egenskapene *shape_Maximum3DDiameter* og *pack_years* ble valgt av mer enn 90% av modellene for OUS-datasettet, men av færre enn 50% av modellene for MAASTRO-datasettet. Fra MAASTRO-datasettet ble det valgt tre formegenskaper, fem kliniske egenskaper og resten bildeegenskaper for LRC. Fra OUS-datasettet var det like mange formegenskaper og kliniske egenskaper som ble valgt, men siden det her var mange færre egenskaper som ble valgt totalt, ville også for OS bildeegenskapene utspille en viktigere rolle for MAASTRO-datasettet. Formegenskaper ble trukket ut av fra begge datasett for alle responser, som viser at formegenskaper er viktige egenskaper for prediksjon av behandlingsutfall. Dette fant også Langberg [84] og Midtjord [91] i sine oppgaver.

For LRC viser egenskapsutvelgingen fra MAASTRO-settet motsatt tendens enn for DFS og OS, og vesentlig flere egenskaper enn for OUS-settet ble valgt av RENT. Egenskapene som ble valgt felles var en formegenskap, to kliniske egenskaper og tre bildeegenskaper. Kun to av de ni mest valgte egenskapene fra OUS-settet ble valgt av mer enn 50% av modellene for MAASTRO-settet.

Det er stor forskjell mellom hvilke egenskaper som fungerer som gode prediktorer for de to ulike datasettene. Dette er en årsak til at ytelsen ble så mye lavere ved testing på det eksterne datasettet enn under kryssvalidering. Forskjellen i viktigheten til bildeegenskapene kan ha sammenheng med at det er to ulike scannere som har blitt brukt hos de ulike behandlingsinstitusjonene. Radiomics-egenskaper er ikke reproducerbare på tvers av scannere fra ulike leverandører, som forklart av TuShu et al. [96]. Denne uttalelsen bygger på Meyer et al. [97] sin studie, som viste at de fleste radiomics-egenskapene var sterkt knyttet til anskaffelse og rekonstruksjon av CT-bildene.

For å forstå hvor ulike datasettene er, ble det trent en modell som inkluderte alle pasientene fra begge datasettene, der pasientene fra OUS ble tildelt klasse 0, og pasientene fra MAASTRO ble tildelt klasse 1. En PCA ble utført på dette samlede datasettet, med egenskapene som ble valgt av over 50% av modellene i RENT for både OUS-datasettet og MAASTRO-datasettet; det vil si egenskapene i midten av venndiagrammene i figur ???. Dette ble gjort for å avdekke om det var store forskjeller i egenskapene som ble ansett som viktige for begge datasett. Resultatet fra PCA-studien er vist i figur 5.5.6.



Figur 5.5.6: PCA basert på egenskapene som ble valgt av over 50% av modellene i RENT for en eller flere responser for begge datasett. OUS-pasientene er markert i blått, og MAASTRO-pasientene i rødt.

Resultatene fra analysen viser at pasientene hovedsakelig fordeler seg i to grupperinger. Ingen av grupperingene inneholder pasienter kun fra et sykehus, men grupperingen til venstre inneholder en vesentlig større andel pasienter behandlet ved OUS, mens grupperingen til høyre inneholder en vesentlig større andel pasienter behandlet ved MAASTRO.

Det er vanskelig å vite hvor stor del av forskjellen mellom ytelse under kryssvalidering og ekstern validering som stammer fra ulikheter i datasettene, og hvor stor del som kan skyldes at modellene er overtilpasset til datasettet brukt til trening. En løsning for å finne ut dette vil være å teste modellene på datasett fra andre behandlingsinstitusjoner eller på et datasett bestående av nye pasienter behandlet ved OUS. Ulempen ved sistnevnte vil være at de nye pasientene er behandlet i nyere tid, som også kan ha innvirkning på modellenes ytelse. En annen mulighet er å legge til en andel MAASTRO-pasienter til treningsdatasettet, og inkludere en andel OUS-pasienter i testdatasettet, for så å trene modellene på nytt og se om de

er mer generaliserende ovenfor det nye datasettet.

5.6 Diskretisering av bildene

I denne oppgaven ble det utført en relativ diskretisering av bildene til 32 og 16 bins for CT-bildene, og 16, 8 og 4 bins for PET-bildene. Zwanenburg et al. [10] forklarer at ulike diskretiseringsnivåer kan ha ulik pålitelighet og reproduserbarhet, så vel som de to ulike diskretiseringsmetodene, relativ og absolutt diskretisering. Begge metoder har sine fordeler, og ved å bruke absolutt diskretisering ville de uthentede egenskapene blitt annerledes.

Lijun et al. [98] fant ut at diskretisering av bildene spilte en større rolle for egenskapene enn segmenteringen av bildene for oropharynx-kreft. Valg av diskretiseringsparameter vil gi ulike utslag, og i denne oppgaven ble det derfor hentet ut egenskaper med flere parametere for både PET og CT. Disse ulike egenskapsgruppene viste seg å være høyt korrelerte, og egenskapsgruppene med den høyeste diskretiseringsparameteren kunne elimineres uten et stort tap av informasjon. Figur 4.1.1 viste en sterk korrelasjon også mellom PET-bildene hentet ut med diskretiseringsparameteren 8 og 4, men her ble begge beholdt for å bevare både grovtekstur og fintekstur i bildene. Det ulike antallet bins valg for PET og CT har sammenheng med bredden av verdier som skal representeres i bildet. I denne oppgaven var WW for CT satt til 350 HU, som betyr at 350 verdier skal representeres av det samme antallet gråtoner som diskretiseringsparameteren. De høyeste SUV-verdiene representert i PET-bildene i henholdsvis treningsdataen og testdataen var på 29.9 og 33.1, altså er bredden av verdier som skal representeres mindre for PET-bildene enn for CT-bildene.

5.7 Overoptimistiske resultater i kryssvalidering

Selv om ulikhetene i datasettene er store, er det også grunn til å tro at resultatene etter kryssvalidering var noe overoptimistiske. Under egenskapsutvelgingen så RENT alle pasientene i treningssettet, og disse ble gjenbrukt i modellutvelgelsen. Det innebærer at testsettet i hver fold under kryssvalidering opprinnelig var en del av treningssettet egenskapsutvelgingen ble utført på. Dette kunne vært løst ved å bruke *holdout*-kryssvalidering, som forklart av Rashka og Mirjalili [9].

Ideelt sett burde en del av treningsdatasettet blitt holdt tilbake, og brukt som et valideringssett før modellene ble testet på det eksterne datasettet. Ulempen med dette er at treningssettet besto av få pasienter i utgangspunktet, og ytelsen til modellene var veldig varierende for ulike inndelinger av undersett av pasienter brukt til trening og testing. Når et tilfeldig utvalg av pasienter ble brukt som valideringssett varierte ytelsen til modellen mellom en AUC på 0.4 til 0.9 for de ulike utvalgene. Det betyr at modellytelsen var veldig avhengig av hvilke pasienter som ble trukket ut, og man kunne risikert at modellen ble både underoptimistisk og overoptimistisk ved å bruke denne tilnærmingen. Derfor ble kryssvalidering benyttet for å få et bilde av modellytelsen før testing på eksternt datasett. Optimalt sett burde egenskapsutvelgingen også skjedd innenfor kryssvalideringsløkken for hver fold, men innenfor tiden tilgjengelig til arbeid med oppgaven ble dette for tidkrevende.

Det var derfor forventet at ytelsen til modellene ville falle noe fra kryssvalidering til validering på eksternt datasett. Til tross for dette, er forskjellene i datasettene og klassedistribusjonene så store at de trolig var hovedårsak til det store fallet i ytelse.

5.8 Forslag til videre arbeid

Som diskutert i seksjon 6.1, ville det vært mulig å benytte en predikert HPV-status for pasientene der denne manglet. Da HPV-status konsekvent ble plukket ut som en av de viktigste egenskapene i prediksjonen av behandlingsutfall, vil dette muligens føre til en høyere ytelse for modellene, spesielt i testing på eksternt data der alle pasienters HPV-status er kjent. Videre vil det være interessant å studere pasientene som ofte ble feilklassifiserte, og prøve å lage modeller som er bedre til å skille disse.

De to datasettene hadde veldig ulik distribusjon av lokasjon til svulsten. Da egenskaper knyttet til svulstens plassering ofte ble valgt ut som viktige for prediksjonen, vil det være interessant å se om modeller tilpasset hver enkelt svulstlokasjon vil føre til en høyere modellytelse.

Innenfor tidsrommet satt av til denne oppgaven ble det ikke tid til videre tolkning av de utvalgte egenskapene. For å få en endra bedre forståelse av hvilke egenskaper som er viktige indikatorer for prediksjon av behandlingsutfall og hvorfor disse er viktige, vil det være interessant å studere de utvalgte egenskapene grundigere. Denne oppgaven så også på modeller basert på vekselvirkninger mellom ulike

egenskaper. Videre ville det vært interessant å studere de utvalgte vekselvirknings-egenskapene nærmere for å forstå sammenheng mellom disse, og hvilken betydning de kan ha i prediksjonen av behandlingsutfall.

Bruk av holdout-kryssvalidering, som diskutert i seksjon 5.7, kunne ledet til andre undersøkt av egenskaper. Det vil være interessant å se hvordan de utvalgte egenskapene ville endret seg, og om resultatene under kryssvalidering og ekstern validering ville ligget nærmere hverandre med denne fremgangsmåten.

I tillegg til binære responsvariabler inneholder OUS-datasettet en *time to event*-variabel for hver respons, som forteller hvor lang tid det tok før hendelsen (tilbakefall/død) inntraff. Med denne informasjonen er det mulig å lage en regresjonsmodell basert på det samme datasettet. Å ta hensyn til time-to-event variabler vil kunne styrke modellene, som vist av Keek et al. [99]. Vallieres et al. [12] lagde både en klassifiseringsmodell og en regresjonsmodell i sin studie. Inkluderingen av regresjonsmodeller kan gjøre personalisering av behandling enda enklere og mer nøyaktig.

Det har blitt gjort forsøk med å bruke dyplæring til å predikere behandlingsutfall for hode- og halskreft. Studier utført av blant annet Diamant et al. [100] har avdekket et stort potensiale for å bruke dyplæring til prediksjon av behandlingsutfall. For videre arbeid vil det være interessant å utvikle en dyplæringsmodell til prediksjon på dette datasettet, og eventuelt se om en kombinasjon av modeller kan avdekke større deler av informasjonen i datasettene.

Kapittel 6

Konklusjon

I denne oppgaven ble ulike modeller for prediksjon av behandlingsutfall for hode- og halskreftpasienter utviklet. Oppgaven tar for seg egenskapsuthenting fra medisinske bilder ved bruk av innskaper [70], egenskapsutvelgelse ved bruk av repetert elastisk nett teknikk [71] og tilpasning, trening og testing av åtte ulike klassifiseringsalgoritmer på 16 ulike undersett av egenskaper. De 16 undersettene besto av åtte undersett uten vekselvirkningsegenskaper og åtte undersett med vekselvirkningsegenskaper.

Modellene ble trent på et datasett fra Oslo universitetssykehus for tre ulike responser; sykdomsfri overlevelse, generell overlevelse og lokalregionalt tilbakefall. Deretter ble modellene testet på et eksternt datasett fra MAASTRO-klinikken i Nederland. Følgende resultater ble oppnådd:

For sykdomsfri overlevelse oppnådde modellen uten vekselvirkningsegenskaper med høyest ytelse en MCC på 0.490 under kryssvalidering. På det eksterne datasettet oppnådde samme modell en MCC på 0.064. Modellen med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering oppnådde en MCC på 0.564, og en ytelse på 0.190 på det eksterne datasettet. Egenskapene som sto frem som de viktigste for prediksjonen av sykdomsfri overlevelse var en kombinasjon av kliniske egenskaper, formegenskaper og radiomics-egenskaper fra PET-bilder.

For generell overlevelse oppnådde modellen uten vekselvirkningsegenskaper med høyest ytelse en MCC på 0.527 under kryssvalidering. På det eksterne datasettet oppnådde samme modell en MCC på 0.204. Modellen med nest høyest ytelse under kryssvalidering, med en MCC på 0.525, oppnådde en ytelse på 0.324 på det eksterne datasettet. Modellen med vekselvirkningsegenskaper med høyest ytelse

under kryssvalidering oppnådde en MCC på 0.621, og en ytelse på 0.251 på det eksterne datasettet. Egenskapene som sto frem som de viktigste i prediksjonen av generell overlevelse var kliniske egenskaper og formegenskaper.

For lokalregionalt tilbakefall oppnådde modellen uten vekselvirkningsegenskaper med høyest ytelse en MCC på 0.428 under kryssvalidering. På det eksterne datasettet oppnådde samme modell en MCC på -0.051. Modellen med vekselvirkningsegenskaper med høyest ytelse under kryssvalidering oppnådde en MCC på 0.647, og en ytelse på 0.125 på det eksterne datasettet. For lokalregionalt tilbakefall sto radiomics-egenskapene frem som de viktigste egenskapene for prediksjon av behandlingsutfall.

For sykdomsfri overlevelse og generell overlevelse hadde de beste modellene høyere ytelse enn de beste modellene kun basert på kliniske data. Andre studier har allikevel oppnådd høyere ytelse [12]. Modellene med vekselvirkningsegenskaper oppnådde gjennomsnittlig høyere ytelse enn de enklere modellene under både kryssvalidering og testing på eksternt datasett. Disse modellene er vanskeligere å tolke enn modellene med de opprinnelige egenskapene, og det finnes et potensiale for forbedring av modellene ved å studere vekselvirkningsegenskapene nærmere.

Ytelsen til modellene sank kraftig fra kryssvalidering til ekstern validering, og det ble avdekket store ulikheter i datasettet brukt til trening av modellen og det eksterne datasettet brukt til testing. For å vite om RENT er et egnet verktøy for utvelging av egenskaper til prediksjon av behandlingsutfall for pasienter med hode- og halskreft vil det være nødvendig å teste modellen på flere datasett fra andre behandlingstilstander.

Bibliografi

- [1] B. Roald, T. Sauer og O. Klepp, *Kreft i Store medisinske leksikon* på snl.no, (lest 04.02.22). URL: <https://snl.no/kreft>.
- [2] Kreftregisteret, *Kreft i Norge [Internett]*, (lest 04.02.22). URL: <https://www.kreftregisteret.no/Temasider/om-kreft/>.
- [3] Kreftforeningen, *Hva er kreft? [Internett]*, (lest 14.01.22). URL: <https://kreftforeningen.no/om-kreft/hva-er-kreft/>.
- [4] Folkehelseinstituttet, *Tall fra Dødsårsaksregisteret [Internett]*, (lest 04.02.22). URL: <https://www.fhi.no/hn/helseregistre-og-registre/dodsarsaksregisteret/tall-fra-dodsarsaksregisteret-for-2020/>.
- [5] Kreftforeningen, *Undersøkelser [Internett]*, (lest 21.01.22). URL: <https://kreftforeningen.no/om-kreft/undersokelser>.
- [6] —, *Hode- og halskreft [Internett]*, (lest 17.01.22). URL: <https://kreftforeningen.no/om-kreft/kreftformer/hode-og-halskreft/>.
- [7] Oslo Universitetssykehus, *Hode- og halskreft [Internett]*, (lest 21.01.22). URL: <https://oslo-universitetssykehus.no/behandlinger/hode-og-halskreft>.
- [8] Kreftforeningen, *Ny kreftsykdom (sekundær kreft) [Internett]*, (lest 23.03.22). URL: <https://kreftforeningen.no/om-kreft/senskader-voksne/ny-kreftsykdom-sekundaer-kreft/>.
- [9] S. Raschka og V. Mirjalili, *Python machine learning : machine learning and deep learning with Python, scikit-learn, and TensorFlow*, eng, Packt: Birmingham, England, 2017.
- [10] A. Zwanenburg, «Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis,» eng, *European journal of nuclear medicine and molecular imaging*, årg. 46, nr. 13, s. 2638–2655, 2019, ISSN: 1619-7070.

- [11] S. Keek, F. Wesseling, H. Woouff mfl., «A Prospectively Validated Prognostic Model for Patients with Locally Advanced Squamous Cell Carcinoma of the Head and Neck Based on Radiomics of Computed Tomography Images,» eng, *Cancers*, årg. 13, nr. 13, s. 3271, 2021, ISSN: 2072-6694.
- [12] M. Vallières, E. Kay-Rivest, L. J. Perrin mfl., «Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer,» eng, *Scientific reports*, årg. 7, nr. 1, s. 10 117–10 117, 2017, ISSN: 2045-2322.
- [13] G. Bruixola, E. Remacha, A. Jiménez-Pastor mfl., «Radiomics and radiogenomics in head and neck squamous cell carcinoma: Potential contribution to patient management and challenges,» eng, *Cancer treatment reviews*, årg. 99, s. 102 263–102 263, 2021, ISSN: 0305-7372.
- [14] National Cancer Institute, *Head and Neck Cancers [Internett]*, (lest 04.02.22). URL: <https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet>.
- [15] Kreftregisteret, *Alkohol og tobakk hovedårsak til hode og halskreft [Internett]*, (lest 17.01.22). URL: <https://www.kreftregisteret.no/Generelt/Nyheter/Alkohol-og-tobakk-og-ikke-HPV-smitte-er-hovedarsak-til-hode-og-halskreft/>.
- [16] O. Klepp, *Metastase i Store medisinske leksikon* på snl.no, (lest 17.01.22). URL: <https://sml.snl.no/metastase>.
- [17] Kreftforeningen, *Behandling [Internett]*, (lest 21.01.22). URL: <https://kreftforeningen.no/om-kreft/behandling/>.
- [18] O. Klepp og T. K. Olsen, *Kreftbehandling i Store medisinske leksikon* på snl.no, (lest 21.01.22). URL: <https://sml.snl.no/kreftbehandling>.
- [19] H. Fritzsche og M. Phillips, *Electromagnetic radiation i Encyclopedia Britannica* på britannica.com, (lest 16.02.22). URL: <https://www.britannica.com/science/electromagnetic-radiation/The-electromagnetic-spectrum>.
- [20] J.-P. Hansen og J. Linder, *Foton i Store norske leksikon* på snl.no, (lest 16.02.22). URL: <https://snl.no/foton/>.
- [21] P. Hermanek, J. S. Rathore, V. Aloisi og S. Carmignato, «Principles of X-ray Computed Tomography,» i *Industrial X-Ray Computed Tomography*, S. Carmignato, W. Dewulf og R. Leach, red. Cham: Springer International Publishing, 2018, s. 25–67, ISBN: 978-3-319-59573-3. DOI: [10.1007/978-3-319-59573-3_2](https://doi.org/10.1007/978-3-319-59573-3_2). adresse: https://doi.org/10.1007/978-3-319-59573-3_2.
- [22] T. Holtebekk og J. Linder, *Røntgenstråler i Store norske leksikon* på snl.no, (lest 15.02.22). URL: <https://snl.no/rontgenstraler/>.

- [23] T. Bjørnstad og J. Linder, *Gammastråling* i *Store norske leksikon* på snl.no, (lest 15.02.22). URL: <https://snl.no/https://snl.no/gammastr%C3%A5ling/>.
- [24] A. Borthne, *Røntgenbilde* i *Store medisinske leksikon* på snl.no, (lest 23.03.22). URL: <https://sml.snl.no/rontgenbilde>.
- [25] T. Holtebekk og T. Bjørnstad, *Foton* i *Store norske leksikon* på snl.no, (lest 16.02.22). URL: <https://snl.no/radioaktivitet/>.
- [26] J. Lilley, *Nuclear physics : principles and applications*, eng, Wiley: Chichester, 2001.
- [27] R. A. Serway, *Physics for scientists and engineers : with modern physics : Vol. 2*, eng, Brooks/Cole-Thomson Learning: Belmont, Calif, 2004.
- [28] A. Borthne, *Gammakamera* i *Store medisinske leksikon* på snl.no, (lest 23.03.22). URL: <https://sml.snl.no/gammakamera>.
- [29] A. Granov, L. Tiutin og T. S. M.D., *Positron Emission Tomography*, eng, Springer: Berlin, Heidelberg, 2013.
- [30] J. S. Welsh, «Beta decay in science and medicine,» eng, *American journal of clinical oncology*, årg. 30, nr. 4, s. 437–439, 2007.
- [31] A. Kamal, *Nuclear Physics*, eng, Spinger: Berlin, Heidelberg, 2014.
- [32] T. Hoang, *Compton effect [Internett]*, (lest 08.02.22). URL: <https://radiopaedia.org/articles/compton-effect>.
- [33] The Editors of Encyclopaedia Britannica, *Rayleigh Scatering* i *Encyclopaedia Britannica* på britannica.com, (lest 17.02.22). URL: <https://www.britannica.com/science/Rayleigh-scattering>.
- [34] J. O. Eeg, *Antipartikkel* i *Store norske leksikon* på snl.no, (lest 08.02.22). URL: <https://snl.no/antipartikkel>.
- [35] A. Berger, «How Does It Work? Positron Emission Tomography,» eng, *BMJ: British Medical Journal*, årg. 326, nr. 7404, s. 1449–1449, 2003, ISSN: 0959-8138.
- [36] Kreftforeningen, *PET/CT [Internett]*, (lest 04.02.22). URL: <https://kreftforeningen.no/om-kreft/undersokelser/pet-ct/>.
- [37] M. B. og Arne Borthne, *PET* i *Store medisinske leksikon* på snl.no, (lest 25.01.22). URL: <https://sml.snl.no/PET>.
- [38] Kreftlex, *PET/CT-undersøkelse [Internett]*, (lest 07.02.22). URL: <https://kreftlex.no/Tykk-og-endetarmskreft/ProsedyreFolder/UTREDNING/Us-PET-undersokelse?lg=ks&containsFaq=True>.

- [39] A. Almuhaideb, N. Papathanasiou og J. Bomanji, «¹⁸F-FDG PET/CT imaging in oncology,» eng, *Annals of Saudi medicine*, årg. 31, nr. 1, s. 3–13, 2011, ISSN: 0256-4947.
- [40] A. Dixon, *F-18 fluorodeoxyglucose [Internett]*, (lest 07.02.22). URL: <https://radiopaedia.org/articles/f-18-fluorodeoxyglucose?lang=us>.
- [41] H. Redman, *F-18 Radionuclide Fact Sheet [Internett]*, (lest 07.02.22). URL: <https://ehs.stanford.edu/reference/f-18-radionuclide-fact-sheet>.
- [42] The American Cancer Society medical and editorial content team, *Understanding Radiation Risk from Imaging Tests [Internett]*, (lest 23.03.22). URL: <https://www.cancer.org/treatment/understanding-your-diagnosis/tests/understanding-radiation-risk-from-imaging-tests.html>.
- [43] P. Moskal, N. Zoń, T. Bednarski mfl., «A novel method for the line-of-response and time-of-flight reconstruction in TOF-PET detectors based on a library of synchronized model signals,» eng, *Nuclear instruments methods in physics research. Section A, Accelerators, spectrometers, detectors and associated equipment*, årg. 775, s. 54–62, 2015, ISSN: 0168-9002.
- [44] Stanford University, scintillator materials group, *What are scintillator materials? [Internett]*, (lest 10.02.22). URL: <https://web.stanford.edu/group/scintillators/scintillators.html>.
- [45] University of Washington, Division of Nuclear Medicine, *Introduction to PET Physics [Internett]*, (lest 14.02.22). URL: https://depts.washington.edu/imreslab/from%20old%20SITE/pet_intro/.
- [46] V. Oikonen, *Attenuation Correction [Internett]*, (lest 10.02.22). URL: <http://www.turkupetcentre.net/petanalysis/attenuation.html>.
- [47] F. H. Fahey, *Data Acquisition in PET Imaging [Internett]*, (lest 10.02.22). URL: <http://www.guillemet.org/irene/coursem/DataorganizationinPET.pdf>.
- [48] University of Queensland Biomedical Imaging, *UQx Bioimg101x 6.7.2 Attenuation Correction V2 YT1 [video]*, (sett 10.02.22). Link: <https://www.youtube.com/watch?v=FS7Q2pUoI8M>.
- [49] St Olavs Hospital, *CT-undersøkelse [Internett]*, (lest 25.01.22). URL: <https://www.helsenorge.no/undersokelse-og-behandling/ct-undersokelse/>.

- [50] National Institute of Biomedical Imaging and Bioengineering, *Computed Tomography (CT) [Internett]*, (lest 25.01.22). URL: <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct/>.
- [51] M. Lin Eugene og P. Alessio Adam, «What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT?» eng, *Journal of cardiovascular computed tomography*, årg. 3, nr. 6, s. 403–408, 2009, ISSN: 1934-5925.
- [52] R. Schofield, L. King, U. Tayal mfl., «Image reconstruction: Part 1 – understanding filtered back projection, noise and image acquisition,» eng, *Journal of cardiovascular computed tomography*, årg. 14, nr. 3, s. 219–225, 2020, ISSN: 1934-5925.
- [53] G.-X. Yuan, C.-H. Ho og C.-J. Lin, «Recent Advances of Large-Scale Linear Classification,» *Proceedings of the IEEE*, årg. 100, nr. 9, s. 2584–2603, 2012. DOI: [10.1109/JPROC.2012.2188013](https://doi.org/10.1109/JPROC.2012.2188013).
- [54] scikit-learn, *1.1. Linear Models [Internett]*, (lest 25.02.22). URL: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression.
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort mfl., «Scikit-learn: Machine Learning in Python,» *Journal of Machine Learning Research*, årg. 12, s. 2825–2830, 2011.
- [56] scikit-learn, *1.4. Support Vector Machines [Internett]*, (lest 25.02.22). URL: <https://scikit-learn.org/stable/modules/svm.html>.
- [57] LightGBM, *Features [internett]*, (lest 04.03.22). URL: <https://lightgbm.readthedocs.io/en/latest/Features.html>.
- [58] G. Ke, Q. Meng, T. Finley mfl., «Lightgbm: A highly efficient gradient boosting decision tree,» *Advances in neural information processing systems*, årg. 30, s. 3146–3154, 2017.
- [59] S. Hrouda-Rasmussen, *Quadratic Discriminant Analysis [Internett]*, (lest 02.03.22). URL: <https://towardsdatascience.com/quadratic-discriminant-analysis-ae55d8a8148a>.
- [60] Scikit-learn, *1.2. Linear and Quadratic Discriminant Analysis [Internett]*, (lest 02.03.22). URL: https://scikit-learn.org/stable/modules/lda_qda.html.

- [61] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz og Y. Singer, «Online Passive-Aggressive Algorithms,» *Journal of Machine Learning Research*, årg. 7, nr. 19, s. 551–585, 2006. adresse: <http://jmlr.org/papers/v7/crammer06a.html>.
- [62] D. Chicco og G. Jurman, «The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,» eng, *BMC genomics*, årg. 21, nr. 1, s. 6–6, 2020, ISSN: 1471-2164.
- [63] J. M. Moan, C. D. Amdal, E. Malinen, J. G. Svestad, T. V. Bogsrud og E. Dale, «The prognostic role of 18F-fluorodeoxyglucose PET in head and neck cancer depends on HPV status,» eng, *Radiotherapy and oncology*, årg. 140, s. 54–61, 2019, ISSN: 0167-8140.
- [64] A. R. Groendahl, I. S. Knudtsen, B. N. Huynh mfl., «A comparison of methods for fully automatic segmentation of tumors and involved nodes in PET/CT of head and neck cancers,» *Physics in Medicine & Biology*, årg. 66, nr. 6, s. 065012, mar. 2021. DOI: [10.1088/1361-6560/abe553](https://doi.org/10.1088/1361-6560/abe553). adresse: <https://doi.org/10.1088/1361-6560/abe553>.
- [65] A. R. Grøndahl, *Clinical data (dichotomised as in Moan et al (2019))*, Notatark tilhørende klinisk data fra Oslo Universitetssykehus.
- [66] G. Van Rossum og F. L. Drake Jr, *Python reference manual*. 1995.
- [67] *Anaconda Software Distribution*, versjon Vers. 2-2.4.0, 2020. adresse: <https://docs.anaconda.com/>.
- [68] C. R. Harris, K. J. Millman, S. J. van der Walt mfl., «Array programming with NumPy,» *Nature*, årg. 585, nr. 7825, s. 357–362, sep. 2020. DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). adresse: <https://doi.org/10.1038/s41586-020-2649-2>.
- [69] W. McKinney, «Data Structures for Statistical Computing in Python,» i *Proceedings of the 9th Python in Science Conference*, S. van der Walt og J. Millman, red., 2010, s. 56–61. DOI: [10.25080/ajora-92bf1922-00a](https://doi.org/10.25080/ajora-92bf1922-00a).
- [70] NMBU Data Science, *Imskaper [Internett]*, (lest 11.03.22). URL: <https://github.com/NMBU-Data-Science/imskaper..>
- [71] A. Jenul, S. Schrunner, B. N. Huynh og O. Tomic, «RENT: A Python Package for Repeated Elastic Net Feature Selection,» *Journal of Open Source Software*, årg. 6, nr. 63, s. 3323, 2021. DOI: [10.21105/joss.03323](https://doi.org/10.21105/joss.03323). adresse: <https://doi.org/10.21105/joss.03323>.
- [72] Brett, M. et al., *nipy/nibabel: 3.1.1.[programvare]*, URL: <https://doi.org/10.5281/zenodo.3924343> (2020).

- [73] J. D. Hunter, «Matplotlib: A 2D graphics environment,» *Computing in Science & Engineering*, årg. 9, nr. 3, s. 90–95, 2007. DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [74] O. Tomic, T. Graff, K. H. Liland og T. Næs, «hoggorm: a python library for explorative multivariate statistics,» *The Journal of Open Source Software*, årg. 4, nr. 39, 2019. DOI: [10.21105/joss.00980](https://doi.org/10.21105/joss.00980). adresse: <http://joss.theoj.org/papers/10.21105/joss.00980>.
- [75] O. Tomic, *hoggormplot [programvare]*, (lest 04.04.22) URL: <https://github.com/olivertomic/hoggormPlot>.
- [76] M. L. Waskom, «seaborn: statistical data visualization,» *Journal of Open Source Software*, årg. 6, nr. 60, s. 3021, 2021. DOI: [10.21105/joss.03021](https://doi.org/10.21105/joss.03021). adresse: <https://doi.org/10.21105/joss.03021>.
- [77] Microsoft Corporation, *Microsoft Excel*, versjon 2019 (16.0), 24. sep. 2018. adresse: <https://office.microsoft.com/excel>.
- [78] Redaksjonen, Helsebiblioteket.no, *Lungekreft – ikke-småcellet [internett]*, (lest 20.03.22). URL: <https://www.helsebiblioteket.no/pasientinformasjon/kreft/lungekreft-ikke-smacellet>.
- [79] N. Yang, «On the Importance of Proper Window and Level Settings in Temporal Bone CT Imaging,» eng, *Philippine journal of otolaryngology head neck surgery*, årg. 35, nr. 2, 2020, ISSN: 1908-4889.
- [80] A. Zwanenburg, M. Vallières, M. A. Abdalah mfl., «The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping,» eng, 2020, ISSN: 0033-8419.
- [81] pyradiomics, *Radiomic Features*, lest(14.03.22). URL: <https://pyradiomics.readthedocs.io/en/latest/features.html#module-radiomics>.
- [82] K. Chauhan, *Image Processing for Automated Diagnosis of Cardiac Diseases*. eng, Elsevier Science Technology: San Diego, 2021.
- [83] M. Pietikäinen, T. Ahonen, A. Hadid og G. Zhao, *Computer Vision Using Local Binary Patterns*, ger ; eng, 1. Aufl., ser. Computational Imaging and Vision. London: Springer Verlag London Limited, 2011, bd. 40, ISBN: 9780857297471.
- [84] G. S. R. E. Langberg, *Searching for biomarkers of disease-free survival in head and neck cancers using PET/CT radiomics [Masteroppgave]*, eng, 2019, Norwegian University of Life Sciences, Ås. adresse: <https://hdl.handle.net/11250/2641820>.

- [85] A. Albuni, *Development of a user-friendly radiomics framework [Masteroppgave]*, eng, 2020, Norwegian University of Life Sciences, Ås. adresse: <https://hdl.handle.net/11250/2721430>.
- [86] N. Mohammadi, *Radiomics using MR brain scans and RENT for identifying patients receiving ADHD treatment [Masteroppgave]*, eng, 2021, Norwegian University of Life Sciences, Ås. adresse: <https://hdl.handle.net/11250/2829804>.
- [87] M. Bogowicz, O. Riesterer, K. Ikenberg mfl., «Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma,» Elsevier: eng, 2017, ISSN: 0360-3016.
- [88] P. Giraud, S. Elles, S. Helfre mfl., «Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists,» eng, *Radiotherapy and oncology*, årg. 62, nr. 1, s. 27–36, 2002, ISSN: 0167-8140.
- [89] M. Mori, G. Benedetti, S. Partelli mfl., «Ct radiomic features of pancreatic neuroendocrine neoplasms (panNEN) are robust against delineation uncertainty,» eng, *Physica medica*, årg. 57, s. 41–46, 2019, ISSN: 1120-1797.
- [90] M. L. Belli, M. Mori, S. Broggi mfl., «Quantifying the robustness of [18F]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients,» eng, *Physica medica*, årg. 49, s. 105–111, 2018, ISSN: 1120-1797.
- [91] A. D. Midtjord, *Prediksjon av behandlingsutfall for hode- og halskreft ved bruk av radiomics av PET/CT-bilder [Masteroppgave]*, nor, 2018, Norwegian University of Life Sciences, Ås. adresse: <http://hdl.handle.net/11250/2570202>.
- [92] A. Sengupta, *Advances in multivariate statistical methods*, eng, World Scientific: New Jersey, 2009.
- [93] A. C. Atkinson, M. Riani og A. Corbellini, «The Box–Cox Transformation: Review and Extensions,» eng, *Statistical science*, årg. 36, nr. 2, s. 239, 2021, ISSN: 0883-4237.
- [94] F. Ø. Winther og J. Mork, *Strupekreft i Store medisinske leksikon* på snl.no, (lest 05.05.22). URL: <https://sml.snl.no/strupekreft>.
- [95] P. O’Sullivan Brian, M. Huang Shao Hui, M. Su Jie mfl., «Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study,» eng, årg. 17, nr. 4, s. 440–451, 2016, ISSN: 1470-2045.

- [96] S.-J. Tu, W.-Y. Chen og C.-T. Wu, «Uncertainty measurement of radiomics features against inherent quantum noise in computed tomography imaging,» eng, *European radiology*, årg. 31, nr. 10, s. 7865–7875, 2021, ISSN: 0938-7994.
- [97] M. Meyer, J. Ronald, F. Vernuccio mfl., «Reproducibility of CT radiomic features within the same patient: influence of radiation dose and CT reconstruction settings,» *Radiology*, årg. 293, nr. 3, s. 583–591, 2019.
- [98] L. Lu, W. Lv, J. Jiang mfl., «Robustness of Radiomic Features in [11C]Choline and [18F]FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization,» eng, *Molecular imaging and biology*, årg. 18, nr. 6, s. 935–945, 2016, ISSN: 1536-1632.
- [99] S. Keek, S. Sanduleanu, F. Wesseling mfl., «Computed tomography-derived radiomic signature of head and neck squamous cell carcinoma (peri)tumoral tissue for the prediction of locoregional recurrence and distant metastasis after concurrent chemo-radiotherapy,» eng, *PloS one*, årg. 15, nr. 5, e0232639–e0232639, 2020, ISSN: 1932-6203.
- [100] A. Diamant, A. Chatterjee, M. Vallières, G. Shenouda og J. Seuntjens, «Deep learning in head neck cancer outcome prediction,» eng, *Scientific reports*, årg. 9, nr. 1, s. 2764, 2019, ISSN: 2045-2322.

Tillegg A

Parametere

I dette tillegget gis en oversikt over de ulike parameterene søkt igjennom i Grid Search for de ulike klassifiseringsalgoritmene. For alle algoritmene unntatt LightGBM er definisjonene hentet fra Scikit-Learn [55].

A.1 Logistisk regresjon

En oversikt over de tilpassede hyperparameterene for logistisk regresjon er gitt i tabell A.1.

Tabell A.1: Tilpassede hyperparameterene for logistisk regresjon.

Parameter	Verdier
penalty Hvilken form for regularisering som benyttes	l1, l2, elasticnet, none
max_iter Maksimalt antall iterasjoner <i>solver</i> får til å konvergere	100, 500, 1000
solver	newton-cg, lbfgs, liblinear, sag, saga

Algoritme brukt i optimerings-
problem

C 0.1, 1, 10, 100

Reguleringsparameter

A.2 Random forest

En oversikt over de tilpassede hyperparameterene for random forest er gitt i tabell A.2.

Tabell A.2: Tilpassede hyperparameterene for random forest.

Parameter	Verdier
criterion Funksjonen som måler kvaliteten til splitten	gini, entropy
max_depth Maksimal dybde til tre	2, 4, 6, 8, 10
n_estimators Antall trær i skogen	20, 35, 50, 65, 80, 95, 110

A.3 K-nærmeste naboer

En oversikt over de tilpassede hyperparameterene for KNN er gitt i tabell A.3.

Tabell A.3: Tilpassede hyperparameterene for KNN.

Parameter	Verdier
-----------	---------

n_neighbors Antall naboer	2, 3, 4, 5, 6, 7, 8
algorithm Algoritme brukt til å beregne nermeste naboer	auto, ball_tree, kd_tree, brute
weights Vektfunksjon brukt i prediksjon	uniform, distance
leaf_size Bladstørrelse for ball_tree og kd_tree	10, 20, 30, 40, 50
metric Distansemetrikk	minkowski, euclidian
p Power parameter for <i>minkowski</i> -metrikk	1, 2

A.4 Support Vector Machines

En oversikt over de tilpassede hyperparameterene for SVM er gitt i tabell A.4.

Tabell A.4: Tilpassede hyperparametere for SVM.

Parameter	Verdier
C Reguleringsparameter	0.1, 1, 10, 100
kernel Spesifiserer kjernetypen som brukes i algoritmen	linear, poly, rbf, sigmoid, precomputed
gamma Kjernekoefisient for rbf, poly og sigmoid	scale, auto, 0.1, 0.3, 0.5, 0.7

A.5 AdaBoost

En oversikt over de tilpassede hyperparameterene for AdaBoost er gitt i tabell A.5.

Tabell A.5: Tilpassede hyperparameterere for AdaBoost.

Parameter	Verdier
n_estimators Maksimalt antall estimatorer boosting termineres ved	20, 35, 50, 65, 80, 95, 110
learning_rate Vekt lagt til hver klassifiseringsalgoritme ved hver boosting-iterasjon	0.5, 0.75, 1, 1.25, 1.5

A.6 Kvadratisk diskriminantanalyse

En oversikt over de tilpassede hyperparameterene for QDA er gitt i tabell A.6.

Tabell A.6: Tilpassede hyperparameterere for QDA.

Parameter	Verdier
reg_param Regulariserer kovariansestimaten	0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2

A.7 Passive aggressive

En oversikt over de tilpassede hyperparameterene for passive aggressive er gitt i tabell A.7.

Tabell A.7: Tilpassede hyperparameterer for passive aggressive.

Parameter	Verdier
C	0.1, 1, 10, 100
Regulariseringsparameter	
max_iter	100, 500, 1000, 1300
Maksimalt antall iterasjoner over treningsdataen	

A.8 Light gradient boosting machine

En oversikt over de tilpassede hyperparameterene for LightGBM er gitt i tabell A.8. Beskrivelsen av parameterene er hentet fra [57].

Tabell A.8: Tilpassede hyperparameterer for LightGBM.

Parameter	Verdier
boosting_type	gbdt, dart, goss, rf
Type boosting som brukes i algoritmen	
num-leaves	11, 21, 31, 41, 51
Maksimalt antall blader i treet	
n_estimators	20, 35, 50, 65, 80, 95, 110
Antall boosting-iterasjoner	
learning_rate	0.5, 0.75, 1, 1.25, 1.5

Vekt lagt til hver klassifiseringsalgoritme ved hver boosting-iterasjon

Tillegg B

Bildeegenskaper

I dette vedlegget vises en oversikt over alle bildeegenskapene hentet ut fra PET- og CT-bildene brukt i denne oppgaven. Definisjonene på de ulike egenskapene er hentet fra PyRadiomics-dokumentasjonen [81], med unntak av LBP-egenskapene.

B.1 Formegenskaper

For beregningen av formegenskaper er det definert en triangulær *mesh*. Beskrivelse av denne finnes i dokumentasjonen til PyRadiomics [81].

For de uthentede formegenskapene i tabell B.1 gjelder følgende [81]:

- N_v : antall voksler inkludert i ROI
- N_f : antall triangler som definerer Mesh.
- V : volum av mesh i mm^3
- A : overflateareal av mesh i mm^2

Tabell B.1: Oversikt over formegenskapene brukt i oppgaven.

Egenskap

Definisjon

Elongation	$elongation = \sqrt{\frac{\lambda_{minor}}{\lambda_{major}}}$
Forholdet mellom to lengste prinsipalkomponenter i ROI	
Flatness	$flatness = \sqrt{\frac{\lambda_{least}}{\lambda_{major}}}$
Forholdet mellom største og minste prinsipalkomponent i ROI	
LeastAxisLength	$LeastAxis = 4\sqrt{\lambda_{least}}$
Minste akselengde av ROI-omsluttende ellipse regnet ut med største prinsipalkomponent λ_{least}	
MajorAxisLength	$MajorAxis = 4\sqrt{\lambda_{major}}$
Største akselengde av en ROI-omsluttet ellipse kalkulert ved bruk av største prinsipalkomponent λ_{major}	
Maximum2DDiameterColumn	
Lengste parvise avstand (euklidisk) mellom to hjørner av svulsten i et plan utspent av radene og tverrsnittet	
Maximum2DDiameterRow	
Lengste parvise avstand (euklidisk) mellom to hjørner av svulsten i et plan utspent av kolonnene og tverrsnittet	
Maximum2DDiameterSlice	
Lengste parvise avstand (euklidisk) mellom to hjørner av svulsten i et plan utspent av radene og kolonnene	
Maximum3DDiameter	
Lengste parvise avstand (euklidisk) mellom to hjørner av svulsten	
MeshVolume	$V = \sum_{i=1}^{N_f} V_i,$
Volumet til ROI	V_i er volumet til en gitt voksel
MinorAxisLength	$MinorAxis = 4\sqrt{\lambda_{minor}}$

Nest lengste akselengde av ROI-omsluttende ellipse regnet ut ved den største prinsipalkomponenten λ_{minor}

Sphericity

$$sphericity = \frac{\sqrt[3]{36^2}}{A}$$

Mål på svulstens rundhet relativ til en kule

SurfaceArea

$$A = \sum_{i=1}^{N_f} A_i,$$

Overflateareaet til svulsten kalkulert av overflaten til hver triangel i mesh

A_i er overflateareaet til en gitt triangel

SurfaceVolumeRatio

$$SurfaceToVolumeRatio = \frac{A}{V}$$

Forholdet mellom overflateareal og volum

VoxelVolume

$$V_{voxel} = \sum_{k=1}^{N_v} V_k$$

Volumet til ROI ved volumet av en voksel ganget med antall vokslar i ROI

B.2 Førsteordens statistiske egenskaper

Førsteordens statistiske egenskaper beskriver distribusjonen av vokselintensiteter innen bilderegionen definert av masken.

For beregningen av førsteordens statistiske egenskaper i tabell B.2 gjelder [81]:

- X : sett av N_p vokslar inkludert i ROI
- $P(i)$: førsteordens histogram med N_g diskrete intensitetsnivåer, hvor N_g er antallet bins med verdier som ikke er 0.
- $p(i)$: er det normaliserte førsteordens histogrammet og lik $\frac{P(i)}{N_p}$

Tabell B.2: Oversikt over de førsteordens statistiske egenskapene brukt i oppgaven.

Egenskap	Definisjon
10Percentile	
Tiende persentil av X	
90Percentile	
Nittiende persentil av X	
Energy	$energy = \sum_{i=1}^{N_p} (X(i) + c)^2$
Mål på omfanget av vokselverdier i et bilde	
Entropy	$entropy = -\sum_{i=1}^{N_g} p(i) \log_2(p(i) +)$
Usikkerhet i bildeverdier	
InterquartileRange	$interquartilerange = P_{75} - P_{25}$
25. til 75. persentil av X	
Kurtosis	$kurtosis = \frac{\mu_4}{\sigma^4}$,
Mål på spissheten til distribusjonen	μ_4 er fjerde sentralmoment
Maximum	$maximum = \max(X)$
Høyeste intensitetsverdi i ROI	
MeanAbsoluteDeviation	$MAD = \frac{1}{N_p} \sum_{i=1}^{N_p} X(i) - \bar{X} $

Gjennomsnittlig avstand for alle intensitetsverdier fra gjennomsnittsverdi til bildevektor

Mean

$$mean = \frac{1}{N_p} \sum_{i=1}^{N_p} X(i)$$

Gjennomsnittlig intensitetsverdi i ROI

Median

Median intensitetsverdi i ROI

Minimum

$$minimum = \min(X)$$

Minste intensitetsverdi i ROI

Range

$$range = \max(X) - \min(X)$$

Intervallet av intensitetsverdier i ROI

RobustMeanAbsoluteDeviation

$$rMAD = \frac{1}{N_{10-90}} \sum_{i=1}^{N_{10-90}} |X_{10-90}(i) - \bar{X}_{10-90}|$$

Gjennomsnittlig avstand for alle intensitetsverdier fra gjennomsnittsverdi til bildevektor utregnet på undersettet av intensitetsverdier mellom 10. og 90. persentil

RootMeanSquared

$$RMS = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) + c)^2}$$

Kvadratrotten av gjennomsnittet til alle kvadratiske intensitetsverdier

Skewness

$$skewness = \frac{\mu_3}{\sigma^3},$$

Måler asymmetri i distribusjonen om gjennomsnittsverdi

μ_3 er tredje sentralmoment

TotalEnergy

$$totalenergy = V_{voxel} \sum_{i=1}^{N_p} (X(i) + c)^2$$

Energy-egenskapen skalert med volumet til vokselene

Uniformity

$$uniformity = \sum_{i=1}^{N_g} p(i)^2$$

Mål på sum of squares for hver intensitetsverdi

Variance

$$variance = \frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^2$$

Mål på spredningen av distribusjonen om gjennomsnittet

B.3 Teksturegenskaper

Gray Level Size Zone Matrix

For beregning av GLSZM-egenskapene i tabell B.3 gjelder følgende:

- N_g : antall diskrete intensitetsverdier i bildet
- N_s : antall diskrete sonestørrelser i bildet
- N_p : antall voksler i bildet
- N_z : antall soner i ROI
- $P(i, j)$: size zone matrix.
- $p(i, j)$: normalisert size zone matrix mhp antall soner.

Tabell B.3: Oversikt over GLSZM-egenskapene brukt i oppgaven.

Egenskap	Definisjon
GrayLevelNonUniformity Variabiliteten til intensitetsverdiene i bildet	$GLN = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_s} P(i,j))^2}{N_z}$
GrayLevelNonUniformityNormalized Normalisert versjon av GLN	$GLNN = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_s} P(i,j))^2}{N_z}$
GrayLevelVariance Variansen i gråtoner for sonene	$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i, j) (i - \mu)^2$
HighGrayLevelZoneEmphasis Distribusjonen av høyere gråtoneverdier	$HGLZE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j) i^2}{N_z}$
LargeAreaEmphasis Mål på distribusjonen av soner med stort areal	$LAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j) j^2}{N_z}$

LargeAreaHighGrayLevelEmphasis Proporsjonen i bildet av felles distribusjon av større soner med høyere gråtoneverdier	$LAHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j) i^2 j^2}{N_z}$
LargeAreaLowGrayLevelEmphasis Proporsjonen i bildet av felles distribusjon av større soner med lavere gråtoneverdier	$LALGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j) j^2}{i^2}}{N_z}$
LowGrayLevelZoneEmphasis Distribusjonen av lavere gråtone-soner	$LGLZE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2}}{N_z}$
SizeZoneNonUniformity Variabiliteten i sonevolumer i bildet	$SZN = \frac{\sum_{j=1}^{N_s} (\sum_{i=1}^{N_g} P(i,j))^2}{N_z}$
SizeZoneNonUniformityNormalized Normalisert verdi av SZN	$SZN = \frac{\sum_{j=1}^{N_s} (\sum_{i=1}^{N_g} P(i,j))^2}{N_z^2}$
SmallAreaEmphasis Mål på distribusjon av små size zoner	$SAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{j^2}}{N_z}$
SmallAreaHighGrayLevelEmphasis Proporsjonen i bildet av felles distribusjon av mindre soner med høyere gråtoneverdier	$SAHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j) i^2}{j^2}}{N_z}$
SmallAreaLowGrayLevelEmphasis Proporsjonen i bildet av felles sannsynlighet med mindre soner med lavere intensitetsverdier	$SALGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2 j^2}}{N_z}$
ZoneEntropy Usikkerheten i distribusjonen av sone størrelser og intensitetsverdier	$ZE = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j) \log_2(p(i,j) +)$
ZonePercentage Grovheten til teksturen målt i forholdet mellom antall soner og antall vokslar i ROI	$ZP = \frac{N_z}{N_p}$

Zone Variance

$$ZV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i, j)(j - \mu)^2$$

Variansen i sonestørrelser for sonene

Gray Level Run Length Matrix

For beregning av GLRLM-egenskapene i tabell B.4 gjelder følgende [81]:

- N_g : antall diskrete intensitetsverdier i bildet
- N_r : antall diskrete intensitetsrekker i bildet
- N_p : antall vokslar i bildet
- $N_r(\Theta)$: antall intensitetsrekker i bildet langs vinkel Θ
- $P(i, j | \Theta)$ er run length matrix for vilkårlig retning Θ
- $p(i, j | \Theta)$: normalisert run length matrix, lik $= \frac{P(i, j | \Theta)}{N_r(\Theta)}$
- er en vilkårlig tilfeldig verdi, her satt til $22 \cdot 10^{-16}$

Tabell B.4: Oversikt over GLRLM-egenskapene brukt i oppgaven.

Egenskap	Definisjon
GrayLevelNonUniformity Variabiliteten av intensitetsverdier i bildet	$GLN = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} P(i, j))^2}{N_r(\Theta)}$
GrayLevelNonUniformityNormalized Normalisert versjon av GLN	$GLNN = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} P(i, j))^2}{N_r(\Theta)^2}$
GrayLevelVariance Variansen i gråtoner for intensitetsrekke-	$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j \theta) (i - \mu)^2$
HighGrayLevelRunEmphasis Distribusjonen av høyere intensitets-	$LGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j \theta) i^2}{N_r(\Theta)}$
LongRunEmphasis Mål på distribusjonen av lange intensi-	$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j \theta) j^2}{N_r(\Theta)}$

LongRunHighGrayLevelEmphasis Felles sannsynlighet for lengre intensitetsrekker med høyere gråtoneverdier	$LRHGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j \Theta) i^2 j^2}{N_r(\Theta)}$
LongRunLowGrayLevelEmphasis Felles sannsynlighet for lengre intensitetsrekker med lavere gråtoneverdier	$LRLGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \Theta)}{i^2}}{N_r(\Theta)}$
LowGrayLevelRunEmphasis Distribusjonen av gråtoneverdier	$LGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \Theta)}{j^2}}{N_r(\Theta)}$
RunEntropy Usikkerhet i distribusjon av intensitetsrekker og intensitetsverdier	$RE = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j \Theta) \log_2(p(i,j \Theta) +)$
RunLengthNonUniformity Likheten av intensitetsrekker gjennom bildet	$RLN = \frac{\sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} P(i,j \Theta))^2}{N_r(\Theta)}$
RunLengthNonUniformityNormalized Normalisert versjon av RLN	$RLN = \frac{\sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} P(i,j \Theta))^2}{N_r(\Theta)^2}$
RunPercentage Grovheten til teksturen målt av forholdet mellom antall intensitetsrekker og antall vokslar	$RP = \frac{N_r(\Theta)}{N_p}$
RunVariance Variansen i intensitetsrekkelengde	$RV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j \Theta) j$
ShortRunEmphasis Mål på distribusjonen av korte intensitetsrekker	$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \Theta)}{j^2}}{N_r(\Theta)}$
ShortRunHighGrayLevelEmphasis Felles sannsynlighet for kortere intensitetsrekker med høyere gråtoneverdier	$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \Theta) i^2}{j^2}}{N_r(\Theta)}$
ShortRunLowGrayLevelEmphasis Felles sannsynlighet for kortere intensitetsrekker med lavere gråtoneverdier	$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \Theta)}{i^2 j^2}}{N_r(\Theta)}$

Neighbouring Gray Tone Difference Matrix

For beregning av NGTDM-egenskapene i tabell B.5 gjelder følgende [81]:

- n_i er antall voxler i X_{gl} med gråtone i
- $N_{v,p}$ totalt antall voxler i X_{gl}
- s_i summen av absolutte forskjeller for gråtone i
- p_i er gråtonesannsynligheten lik $\frac{n_i}{N_v}$
- N_g antall diskrete gråtoner
- N_{g_p} antall gråtoner der $p_1 \neq 0$

Tabell B.5: Oversikt over NGTDM-egenskapene brukt i oppgaven.

Egenskap	Definisjon
Busyness_d_1 Mål på endringen fra en piksel til dens nabo	$Busyness = \frac{\sum_{i=1}^{N_g} p_i s_i}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} i-j p_j}$
Coarseness_d_1 Mål på gjennomsnittlig forskjell mellom sentervoksel og dens nabolag	$Coarseness = \frac{1}{\sum_{i=1}^{N_g} p_i s_i}$
Complexity_d_1 Et bilde anses som komplekst når bildet er ikke-uniformt og det er mange raske endringer i gråtoneintensitet	$Complexity = \frac{1}{N_{v,p} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g}} i-j \frac{p_i s_i + p_j s_j}{p_i + p_j}$
Contrast_d_1 Mål på romlig intensitetsendring, avhengig av generell gråtonedynamisk range	$Contrast = \left(\frac{1}{g,p(N_{g,p}-1)} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_i p_j (i-j)^2 \right) \left(\frac{1}{N_{v,p} \sum_{i=1}^{N_g} s_i} \right)$

Strength_d_1

$$Strength = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p_i + p_j)(i-j)^2}{\sum_{i=1}^{N_g} s_i}$$

Mål på primitivene i et bilde

Gray Level Dependence Matrix

For beregning av GLDM-egenskapene i tabell B.6 gjelder følgende [81]:

- N_g antall diskrete intensitetsverdier i bildet.
- N_d antall diskrete avhengighetsstørrelser i bildet.
- N_z antall avhengighetssoner i bildet.
- $P(i, j)$ er avhengighetsmatrisen.
- $p(i, j)$ er normalisert avhengighetsmatrise, definert som $\frac{P(i, j)}{N_z}$

Tabell B.6: Oversikt over GLDM-egenskapene brukt i oppgaven.

Egenskap	Definisjon
DependenceEntropy_d_1 Mål på tilfeldigheten i avhengigheter	$DE = -\sum_{i=j}^{N_g} \sum_{j=1}^{N_d} p(i, j) \log_2(p(i, j) +)$
DependenceNonUniformity_d_1 Måler likheten av avhengighet gjennom bildet	$DN = \frac{\sum_{j=1}^{N_d} (\sum_{i=1}^{N_g} P(i, j))^2}{N_z}$
DependenceNonUniformity Normalized_d_1 Normalisert versjon av DN	$DN = \frac{\sum_{j=1}^{N_d} (\sum_{i=1}^{N_g} P(i, j))^2}{N_z^2}$
DependenceVariance_d_1 Måler variansen i avhengighetsstørrelse i bildet	$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) (j - \mu)^2,$ $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} j p(i, j)$
GrayLevelNonUniformity_d_1 Måler likheten av gråtoneintensiteter i bildet	$GLN = \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_d} P(i, j))^2}{N_z}$
GrayLevelVariance_d_1 Mål på variansen i gråtoner i bildet	$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) (i - \mu)^2,$ $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} i p(i, j)$
HighGrayLevelEmphasis_d_1	$HGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j) j^2}{N_z}$

Mål på distribusjonen av høyere gråtoneverdier

LargeDependenceEmphasis_d_1	$LDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i,j)j^2}{N_z}$
------------------------------------	---

Mål på distribusjonen av store avhengigheter

LargeDependenceHighGrayLevelEmphasis_d_1	$LDHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i,j)i^2j^2}{N_z}$
---	---

Mål på felles distribusjon av stor avhengighet med høyere gråtoneverdier

LargeDependenceLowGrayLevelEmphasis_d_1	$LDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)j^2}{i^2}}{N_z}$
--	--

Mål på felles distribusjon av stor avhengighet med lavere gråtoneverdier

LowGrayLevelEmphasis	$SDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)}{i^2}}{N_z}$
-----------------------------	--

Måler distribusjonen av lave gråtoneverdier

SmallDependenceEmphasis_d_1	$SDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)}{i^2}}{N_z}$
------------------------------------	--

Mål på distribusjonen av små avhengigheter

SmallDependenceHighGrayLevelEmphasis_d_1	$SDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)i^2}{j^2}}{N_z}$
---	--

Mål på felles distribusjon av liten avhengighet med høyere gråtoneverdier

SmallDependenceLowGrayLevelEmphasis_d_1	$SDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)}{i^2j^2}}{N_z}$
--	--

Mål på felles distribusjon av liten avhengighet med lavere gråtoneverdier

Gray Level Co-occurrence Matrix

For beregning av GLCM-egenskapene i tabell B.7 gjelder følgende [81]:

- : et vilkårlig lite og positivt tall. Her satt til $2.2 \cdot 10^{-16}$

- $P(i, j)$: co-forekomst-matrise for vilkårlig δ og Θ
- $p(i, j)$: Normalisert form av co-forekomst-matrise, lik $\frac{P(i,j)}{\Sigma P(i,j)}$
- N_g : antall diskrete intensitetsverdier i bildet
- $p_x(i) = \Sigma_{j=1}^{N_g} p(i, j)$: marginale radsannsynligheter
- $p_y(j) = \Sigma_{i=1}^{N_g} p(i, j)$: marginale kolonnesannsynligheter
- μ_x : Gjennomsnittlig gråtonesannsynlighet for p_x
- μ_y : Gjennomsnittlig gråtonesannsynlighet for p_y
- σ_x : standardavvik til p_x
- σ_y : standardavvik til p_y
- $p_{x+y} = \Sigma_{i=1}^{N_g} \Sigma_{j=1}^{N_g} p(i, j)$ der $i + j = k$, og $k = 2, 3, \dots, 2N_g$
- $p_{x-y} = \Sigma_{i=1}^{N_g} \Sigma_{j=1}^{N_g} p(i, j)$ der $|i - j| = k$, og $k = 0, 1, \dots, N_g - 1$
- $HX = - \Sigma_{i=1}^{N_g} p_x(i) \log_2(p_x(i) +)$: entropi til p_x
- $HY = - \Sigma_{j=1}^{N_g} p_y(j) \log_2(p_y(j) +)$: entropi til p_y
- $HXY = - \Sigma_{i=1}^{N_g} \Sigma_{j=1}^{N_g} p(i, j) \log_2(p(i, j) +)$: entropi til $p(i, j)$
- $HXY1 = - \Sigma_{i=1}^{N_g} \Sigma_{j=1}^{N_g} p(i, j) \log_2(p_x(i)p_y(j) +)$
- $HXY2 = - \Sigma_{i=1}^{N_g} \Sigma_{j=1}^{N_g} p_x(i)p_y(j) \log_2(p_x(i)p_y(j) +)$

Tabell B.7: Oversikt over GLCM-egenskapene brukt i oppgaven.

Egenskap	Definisjon
Autocorrelation_d_1 Mål på størrelsen til finheten og grovheten til teksten	$autocorrelation = \Sigma_{i=1}^{N_g} \Sigma_{j=1}^{N_g} p(i, j)ij$

ClusterProminence_d_1	$cluster\ prominence = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^4 p(i, j)$
Mål på skjevhet og asymmetri i GLCM	
ClusterShade_d_1	$cluster\ shade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^3 p(i, j)$
Mål på skjevheten og uniformiteten til GLCM	
ClusterTendency_d_1	$cluster\ tendency = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^2 p(i, j)$
Mål på grupperinger av voksler med liknende gråtoneverdier	
Contrast_d_1	$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j)$
Mål på lokal intensitetsvariasjon som favorittiserer verdier unna diagonalen $i = j$	
Correlation_d_1	$correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) i j - \mu_x \mu_y}{\sigma_x(i) \sigma_y(j)}$
Viser lineær avhengighet til grånivåer til deres respektive voksler i GLCM	
DifferenceAverage_d_1	$difference\ average = \sum_{k=0}^{N_g-1} k p_{x-y}(k)$
Mål på forholdet mellom forekomster av par med liknende intensitetsverdier og forekomster av par med ikke-liknende intensitetsverdier	
DifferenceEntropy_d_1	$difference\ entropy = \sum_{k=0}^{N_g-1} p_{x-y}(k) \log_2(p_{x-y}(k) +)$
Mål på tilfeldigheten/variabiliteten i nabolagets ulikheter i intensitetsverdier	
DifferenceVariance_d_1	$difference\ variance = \sum_{k=0}^{N_g-1} (k - DA)^2 p_{x-y}(k)$

Mål på heterogenitet som plasserer høyere vektorer på ulike intensitetsverdi-par som avviker mer fra gjennomsnittet

Id_d_1

$$ID = \sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1+k}$$

Mål på lokal homogenitet i et bilde

Idm_d_1

$$IDM = \sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1+k^2}$$

Mål på lokal homogenitet i et bilde

Idmn_d_1

$$IDMN = \sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{i + (\frac{k^2}{N_g^2})}$$

Mål på lokal homogenitet i et bilde

Idn_d_1

$$IDN = \sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1 + \frac{k}{N_g}}$$

Mål på lokal homogenitet i et bilde

Imc1_d_1

$$IMC1 = \frac{HXY - HXY1}{\max(HX, HY)}$$

Korrelasjonen mellom sannsynlighetsdistribusjonene til i og j

Imc2_d_1 til i og j

$$IMC2 = \sqrt{1 - e^{-2(HXY2 - HXY)}}$$

Korrelasjonen mellom sannsynlighetsdistribusjonene

InverseVariance_d_1

$$inverse\ variance = \sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{k^2}$$

Invers varians

JointAverage_d_1

$$joint\ average = \mu_x = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) i$$

Returnerer gjennomsnittlig gråtoneintensitet i i -distribusjonen

JointEnergy_d_1

$$joint\ energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i, j))^2$$

Mål på homogene mønstre i bildet

JointEntropy_d_1

$$joint\ entropy = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log_2(p(i, j) +)$$

Mål på tilfeldigheter i nabolagets intensitetsverdier

MCC_d_1 $MCC = \sqrt{\text{nest største eigenverdi til } \bar{Q}},$

Mål på kompleksiteten til teksturen $Q(i, j) = \sum_{k=0}^{N_g} \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$

MaximumProbability_d_1 $\text{maximum probability} = \max(p(i, j))$

Forekomster av mest pre-dominante par av nabolagets intensitetsverdier

SumAverage_d_1 $\text{sum average} = \sum_{k=2}^{2N_g} p_{x+y}(k)$

Måler forholdet mellom forekomster av par med lavere intensitetsverdier og forekomster av par med høyere intensitetsverdier

SumEntropy_d_1 $\text{sum entropy} = \sum_{k=2}^{2N_g} p_{x+y}(k) \log_2(p_{x+y}(k) \log_2(p_{x+y}(k) + 1))$

Sum av nabolags intensitetsverdier

SumSquares_d_1 $\text{sum squares} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_x)^2 p(i, j)$

Mål på distribusjon av nabolagets intensitetsverdier om gjennomsnittlig intensitetsverdi i GLCM

Local Binary Pattern

Rotasjon av bildet fører til ulik tolkning av LBP-verdier. Når LBP-egenskaper hentes ut er det nødvendig å gjenkjenne de unike mønstrene, for å unngå å bruke for mye ressurser på å hente ut mønstre som ikke er unike [83]. Ved å gjøre dette kan antall mønstre reduseres fra 64 til 10. Hver sentervoksel har 6 naboer fordelt på 3 akser, dersom det ikke tas hensyn til naboene på diagonalen. Hver av disse aksene kan inneholde følgende informasjon:

1. Ingen av naboene på akse er høyere enn terskelverdi.

2. En av naboene på aksene er høyere enn terskelverdi.
3. Begge naboene på aksene er høyere enn terskelverdi.

LBP egenskapene er navngitt ved tre siffer, der det første sifferet forteller hvor mange av aksene som faller innenfor kategori 1, det andre sifferet forteller hvor mange akser som faller innenfor kategori 2 og det siste sifferet forteller hvor mange akser som faller innenfor kategori 3 [83]. Eksempelvis vil egenskap *LBP*_111 tilsvare mønsteret med 1 akse som faller innenfor hver kategori.

LBP-egenskaper i tabell B.8 ble hentet ut fra PET- og CT-bildene:

Tabell B.8: Oversikt over LBP-egenskapene brukt i oppgaven.

Egenskap

LBP003

LBP012

LBP021

LBP030

LBP102

LBP111

LBP120

LBP201

LBP210

LBP300

Tillegg C

Resultater fra kryssvalidering og prediksjon

Dette tillegget tar for seg resultatene hver klassifiseringsalgoritme oppnådde på hvert undersett av egenskaper uten vekselvirkninger under både kryssvalidering og testing på eksternt datasett. Resultatene er presentert i tabeller for hver klassifiseringsalgoritme som viser ytelsen til hver metrikk (gjennomsnittlig for kryssvalidering) for hvert undersett.

Resultatene er presentert i en vedlagt PDF, med følgende oppbygging:

- Kryssvalidering
 - DFS
 - OS
 - LRC
- Ekstern validering
 - DFS
 - OS
 - LRC

KRYSSVALIDERING

DFS

Logistisk regresjon

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.676	0.687	0.675	0.685	0.692	0.692	0.674	0.681
F1: 0	0.750	0.762	0.758	0.757	0.763	0.763	0.739	0.739
Nøyaktighet	0.719	0.732	0.725	0.727	0.734	0.734	0.712	0.714
ROC AUC	0.714	0.726	0.718	0.721	0.728	0.728	0.708	0.711
MCC	0.437	0.463	0.447	0.453	0.466	0.466	0.422	0.428

Random forest

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.649	0.661	0.671	0.679	0.675	0.678	0.673	0.640
F1: 0	0.727	0.737	0.742	0.746	0.743	0.748	0.737	0.736
Nøyaktighet	0.695	0.706	0.713	0.718	0.715	0.719	0.710	0.698
ROC AUC	0.689	0.700	0.708	0.714	0.710	0.714	0.706	0.690
MCC	0.386	0.409	0.424	0.436	0.428	0.436	0.419	0.392

KNN

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.601	0.632	0.620	0.647	0.645	0.633	0.655	0.628
F1: 0	0.724	0.733	0.726	0.749	0.748	0.749	0.743	0.756
Nøyaktighet	0.676	0.692	0.684	0.708	0.707	0.704	0.707	0.707
ROC AUC	0.665	0.684	0.675	0.699	0.698	0.693	0.700	0.694
MCC	0.346	0.380	0.363	0.413	0.411	0.405	0.411	0.412

SVM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.491	0.686	0.675	0.666	0.664	0.710	0.633	0.606
F1: 0	0.752	0.755	0.755	0.761	0.765	0.771	0.747	0.745
Nøyaktighet	0.668	0.727	0.722	0.723	0.725	0.745	0.702	0.692
ROC AUC	0.644	0.722	0.716	0.714	0.715	0.741	0.692	0.679
MCC	0.356	0.453	0.443	0.444	0.448	0.490	0.402	0.385

AdaBoost

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.601	0.632	0.615	0.627	0.584	0.623	0.632	0.626
F1: 0	0.669	0.695	0.679	0.693	0.663	0.705	0.709	0.703
Nøyaktighet	0.641	0.669	0.653	0.666	0.630	0.671	0.677	0.671
ROC AUC	0.638	0.666	0.650	0.662	0.626	0.665	0.672	0.666
MCC	0.279	0.336	0.304	0.330	0.256	0.337	0.351	0.340

QDA

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.666	0.656	0.682	0.680	0.685	0.682	0.670	0.674
F1: 0	0.732	0.735	0.754	0.749	0.755	0.755	0.741	0.744
Nøyaktighet	0.705	0.702	0.724	0.720	0.725	0.725	0.711	0.715
ROC AUC	0.700	0.696	0.719	0.715	0.720	0.720	0.706	0.710
MCC	0.408	0.400	0.447	0.438	0.449	0.449	0.419	0.427

Passive aggressive

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.623	0.645	0.635	0.646	0.651	0.654	0.653	0.641
F1: 0	0.685	0.712	0.707	0.722	0.724	0.729	0.724	0.720
Nøyaktighet	0.659	0.684	0.678	0.692	0.694	0.698	0.695	0.689
ROC AUC	0.657	0.680	0.674	0.687	0.690	0.693	0.691	0.683
MCC	0.318	0.368	0.356	0.382	0.388	0.395	0.394	0.377

LightGBM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.633	0.650	0.640	0.648	0.653	0.661	0.646	0.626
F1: 0	0.699	0.713	0.710	0.714	0.723	0.721	0.710	0.694
Nøyaktighet	0.672	0.686	0.681	0.687	0.694	0.696	0.683	0.666
ROC AUC	0.668	0.683	0.677	0.683	0.690	0.693	0.680	0.662
MCC	0.341	0.371	0.359	0.372	0.386	0.392	0.364	0.329

OS**Logistisk regresjon**

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.659	0.674	0.659	0.661	0.673	0.660	0.660	0.638
F1: 0	0.808	0.812	0.802	0.804	0.816	0.809	0.809	0.804
Nøyaktighet	0.756	0.762	0.751	0.753	0.766	0.757	0.757	0.747
ROC AUC	0.731	0.740	0.729	0.731	0.741	0.732	0.732	0.718
MCC	0.481	0.469	0.472	0.477	0.503	0.483	0.483	0.458

Random forest

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.596	0.608	0.629	0.630	0.630	0.633	0.633	0.633
F1: 0	0.805	0.806	0.809	0.807	0.809	0.810	0.810	0.805
Nøyaktighet	0.739	0.742	0.750	0.748	0.749	0.751	0.751	0.747
ROC AUC	0.697	0.703	0.715	0.715	0.715	0.717	0.717	0.715
MCC	0.435	0.443	0.463	0.460	0.462	0.465	0.465	0.458

KNN

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
-----------	-----	-----	-----	-----	-----	-----	-----	-----

F1: 1	0.586	0.605	0.570	0.590	0.577	0.585	0.585	0.627
F1: 0	0.811	0.818	0.800	0.816	0.788	0.801	0.801	0.806
Nøyaktighet	0.742	0.751	0.728	0.747	0.719	0.732	0.732	0.746
ROC AUC	0.695	0.706	0.683	0.699	0.680	0.689	0.689	0.712
MCC	0.442	0.464	0.410	0.456	0.392	0.420	0.420	0.454

SVM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.583	0.651	0.605	0.695	0.670	0.666	0.666	0.490
F1: 0	0.811	0.812	0.809	0.815	0.826	0.809	0.809	0.809
Nøyaktighet	0.741	0.757	0.744	0.771	0.773	0.758	0.758	0.723
ROC AUC	0.693	0.728	0.702	0.755	0.742	0.735	0.735	0.655
MCC	0.442	0.480	0.447	0.519	0.514	0.485	0.485	0.410

AdaBoost

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.616	0.611	0.595	0.604	0.653	0.649	0.649	0.632
F1: 0	0.775	0.770	0.754	0.766	0.807	0.800	0.800	0.797
Nøyaktighet	0.718	0.713	0.696	0.708	0.754	0.747	0.747	0.740
ROC AUC	0.696	0.691	0.676	0.685	0.728	0.723	0.723	0.712
MCC	0.402	0.391	0.359	0.381	0.475	0.461	0.461	0.445

QDA

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.686	0.703	0.698	0.685	0.688	0.679	0.679	0.690
F1: 0	0.811	0.812	0.813	0.805	0.826	0.832	0.832	0.824
Nøyaktighet	0.766	0.771	0.770	0.760	0.778	0.773	0.773	0.776
ROC AUC	0.749	0.759	0.757	0.747	0.754	0.747	0.747	0.754
MCC	0.506	0.522	0.519	0.498	0.527	0.515	0.515	0.525

Passive aggressive

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.622	0.643	0.624	0.636	0.648	0.647	0.647	0.646
F1: 0	0.776	0.786	0.785	0.790	0.800	0.800	0.800	0.797
Nøyaktighet	0.721	0.732	0.729	0.736	0.747	0.747	0.747	0.745
ROC AUC	0.701	0.711	0.705	0.714	0.724	0.724	0.724	0.722
MCC	0.411	0.433	0.424	0.442	0.464	0.465	0.465	0.461

LightGBM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.659	0.668	0.651	0.651	0.633	0.637	0.637	0.638
F1: 0	0.800	0.804	0.793	0.790	0.782	0.782	0.782	0.782
Nøyaktighet	0.750	0.755	0.742	0.740	0.729	0.729	0.729	0.730
ROC AUC	0.729	0.735	0.722	0.721	0.708	0.709	0.709	0.710
MCC	0.471	0.482	0.456	0.451	0.427	0.428	0.428	0.430

LRC

Logistisk regresjon

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.419	0.385	0.263	0.513	0.434	0.485	0.360	0.477
F1: 0	0.834	0.844	0.817	0.861	0.868	0.862	0.822	0.875
Nøyaktighet	0.744	0.752	0.709	0.785	0.787	0.783	0.724	0.800
ROC AUC	0.624	0.609	0.544	0.681	0.638	0.664	0.590	0.661
MCC	0.270	0.257	0.103	0.390	0.346	0.369	0.201	0.394

Random forest

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.146	0.021	0.027	0.138	0.008	0.134	0.149	0.000
F1: 0	0.838	0.859	0.858	0.854	0.861	0.857	0.859	0.861
Nøyaktighet	0.730	0.753	0.752	0.751	0.756	0.756	0.759	0.756
ROC AUC	0.517	0.502	0.503	0.526	0.501	0.528	0.534	0.500
MCC	0.054	0.011	0.013	0.099	0.005	0.111	0.130	-0.002

KNN

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.318	0.039	0.376	0.356	0.305	0.344	0.244	0.269
F1: 0	0.822	0.814	0.837	0.855	0.858	0.849	0.840	0.856
Nøyaktighet	0.720	0.717	0.743	0.765	0.764	0.756	0.737	0.760
ROC AUC	0.570	0.603	0.602	0.600	0.581	0.593	0.550	0.567
MCC	0.160	0.215	0.235	0.260	0.231	0.233	0.138	0.207

SVM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.380	0.420	0.321	0.555	0.257	0.000	0.426	0.461
F1: 0	0.854	0.859	0.847	0.861	0.874	0.861	0.858	0.866
Nøyaktighet	0.765	0.774	0.752	0.789	0.785	0.756	0.773	0.786
ROC AUC	0.610	0.630	0.584	0.707	0.575	0.500	0.633	0.652
MCC	0.274	0.316	0.212	0.428	0.271	0.000	0.313	0.362

AdaBoost

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.250	0.264	0.269	0.323	0.383	0.314	0.334	0.329
F1: 0	0.791	0.795	0.803	0.807	0.825	0.797	0.810	0.803
Nøyaktighet	0.675	0.682	0.692	0.702	0.729	0.689	0.706	0.698
ROC AUC	0.525	0.533	0.593	0.567	0.603	0.558	0.574	0.568
MCC	0.054	0.072	0.088	0.144	0.223	0.124	0.158	0.144

QDA

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.000	0.400	0.420	0.391	0.345	0.395	0.374	0.458
F1: 0	0.861	0.568	0.687	0.787	0.820	0.841	0.852	0.864

Nøyaktighet	0.756	0.502	0.597	0.687	0.719	0.750	0.762	0.784
ROC AUC	0.500	0.565	0.600	0.596	0.581	0.614	0.607	0.649
MCC	-0.001	0.117	0.175	0.188	0.185	0.259	0.263	0.354

Passive aggressive

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.426	0.485	0.473	0.513	0.508	0.491	0.466	0.448
F1: 0	0.800	0.810	0.816	0.833	0.841	0.845	0.839	0.851
Nøyaktighet	0.705	0.720	0.730	0.753	0.762	0.765	0.755	0.786
ROC AUC	0.620	0.642	0.653	0.680	0.679	0.668	0.653	0.647
MCC	0.236	0.279	0.302	0.357	0.364	0.352	0.322	0.331

LightGBM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.202	0.255	0.263	0.270	0.292	0.326	0.346	0.326
F1: 0	0.821	0.788	0.819	0.816	0.818	0.807	0.821	0.815
Nøyaktighet	0.710	0.688	0.711	0.708	0.712	0.705	0.724	0.713
ROC AUC	0.524	0.537	0.546	0.547	0.557	0.570	0.585	0.571
MCC	0.060	0.088	0.107	0.108	0.132	0.154	0.195	0.161

EKSTERN VALIDERING

DFS

Logistisk regresjon

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.734	0.710	0.667	0.667	0.701	0.672	0.696	0.734
F1: 0	0.373	0.333	0.333	0.333	0.375	0.313	0.300	0.373
Nøyaktighet	0.626	0.596	0.556	0.556	0.596	0.556	0.576	0.626
ROC AUC	0.574	0.544	0.513	0.513	0.552	0.510	0.524	0.574
MCC	0.189	0.111	0.031	0.031	0.120	0.023	0.059	0.189

Random forest

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.709	0.737	0.698	0.651	0.672	0.697	0.677	0.722
F1: 0	0.479	0.462	0.435	0.389	0.400	0.394	0.423	0.431
Nøyaktighet	0.626	0.646	0.606	0.556	0.576	0.596	0.586	0.626
ROC AUC	0.595	0.605	0.571	0.524	0.541	0.555	0.554	0.585
MCC	0.204	0.242	0.155	0.051	0.090	0.125	0.115	0.194

KNN

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.626	0.688	0.632	0.667	0.609	0.617	0.550	0.622
F1: 0	0.482	0.466	0.500	0.487	0.458	0.410	0.308	0.430

Nøyaktighet	0.566	0.606	0.576	0.596	0.545	0.535	0.455	0.545
ROC AUC	0.554	0.578	0.566	0.577	0.533	0.514	0.431	0.526
MCC	0.108	0.164	0.132	0.156	0.067	0.029	-0.141	0.053

SVM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.571	0.705	0.645	0.688	0.672	0.667	0.701	0.727
F1: 0	0.516	0.305	0.405	0.429	0.400	0.377	0.375	0.455
Nøyaktighet	0.545	0.586	0.556	0.596	0.576	0.566	0.596	0.636
ROC AUC	0.551	0.532	0.528	0.562	0.541	0.529	0.552	0.597
MCC	0.101	0.082	0.058	0.135	0.090	0.064	0.120	0.219

AdaBoost

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.733	0.662	0.542	0.574	0.567	0.628	0.641	0.701
F1: 0	0.478	0.353	0.462	0.410	0.333	0.416	0.343	0.328
Nøyaktighet	0.646	0.556	0.505	0.505	0.475	0.545	0.535	0.586
ROC AUC	0.609	0.517	0.506	0.492	0.452	0.523	0.500	0.536
MCC	0.244	0.038	0.012	-0.016	-0.098	0.047	0.000	0.088

QDA

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.682	0.722	0.677	0.687	0.696	0.696	0.672	0.722
F1: 0	0.406	0.431	0.382	0.344	0.349	0.300	0.313	0.431
Nøyaktighet	0.586	0.626	0.576	0.576	0.586	0.576	0.556	0.626
ROC AUC	0.550	0.585	0.538	0.531	0.539	0.524	0.510	0.585
MCC	0.109	0.194	0.084	0.072	0.093	0.059	0.023	0.194

Passive aggressive

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.614	0.485	0.651	0.614	0.700	0.701	0.595	0.735
F1: 0	0.476	0.442	0.389	0.476	0.276	0.328	0.364	0.235
Nøyaktighet	0.556	0.465	0.556	0.556	0.576	0.586	0.505	0.606
ROC AUC	0.546	0.472	0.524	0.546	0.520	0.536	0.481	0.539
MCC	0.091	-0.056	0.051	0.091	0.052	0.088	-0.039	0.126

LightGBM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.723	0.715	0.622	0.645	0.672	0.639	0.581	0.636
F1: 0	0.471	0.361	0.430	0.405	0.474	0.456	0.395	0.319
Nøyaktighet	0.636	0.606	0.545	0.556	0.596	0.566	0.505	0.525
ROC AUC	0.600	0.557	0.526	0.528	0.573	0.547	0.488	0.488
MCC	0.222	0.139	0.053	0.058	0.150	0.095	-0.024	-0.027

OS

Logistisk regresjon

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.672	0.677	0.683	0.703	0.694	0.698	0.698	0.720
F1: 0	0.438	0.459	0.444	0.457	0.486	0.472	0.472	0.521
Nøyaktighet	0.586	0.596	0.596	0.616	0.616	0.616	0.616	0.646
ROC AUC	0.570	0.581	0.580	0.598	0.601	0.600	0.600	0.631
MCC	0.157	0.179	0.180	0.229	0.224	0.226	0.226	0.294

Random forest

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.679	0.617	0.661	0.655	0.672	0.672	0.672	0.667
F1: 0	0.630	0.549	0.500	0.481	0.474	0.474	0.474	0.519
Nøyaktighet	0.657	0.586	0.596	0.586	0.596	0.596	0.596	0.606
ROC AUC	0.655	0.583	0.585	0.574	0.582	0.582	0.582	0.596
MCC	0.310	0.166	0.179	0.158	0.179	0.179	0.179	0.201

KNN

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.620	0.615	0.655	0.632	0.654	0.649	0.649	0.706
F1: 0	0.612	0.574	0.512	0.469	0.593	0.552	0.552	0.557
Nøyaktighet	0.616	0.596	0.596	0.566	0.626	0.606	0.606	0.646
ROC AUC	0.619	0.595	0.587	0.556	0.624	0.600	0.600	0.635
MCC	0.237	0.190	0.180	0.116	0.248	0.204	0.204	0.286

SVM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.696	0.661	0.661	0.683	0.672	0.688	0.688	0.661
F1: 0	0.605	0.468	0.468	0.480	0.474	0.429	0.429	0.584
Nøyaktighet	0.657	0.586	0.586	0.606	0.596	0.596	0.596	0.626
ROC AUC	0.651	0.573	0.573	0.592	0.582	0.578	0.578	0.622
MCC	0.306	0.157	0.157	0.201	0.179	0.182	0.182	0.246

AdaBoost

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.586	0.577	0.661	0.556	0.672	0.672	0.672	0.661
F1: 0	0.586	0.460	0.432	0.467	0.438	0.438	0.438	0.394
Nøyaktighet	0.586	0.525	0.576	0.515	0.586	0.586	0.586	0.566
ROC AUC	0.589	0.519	0.561	0.511	0.570	0.570	0.570	0.548
MCC	0.178	0.039	0.134	0.023	0.157	0.157	0.157	0.111

QDA

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.683	0.661	0.667	0.692	0.693	0.683	0.683	0.734
F1: 0	0.444	0.432	0.417	0.412	0.451	0.444	0.444	0.514
Nøyaktighet	0.596	0.576	0.576	0.596	0.606	0.596	0.596	0.657
ROC AUC	0.580	0.561	0.559	0.577	0.589	0.580	0.580	0.639

MCC	0.180	0.134	0.134	0.184	0.204	0.180	0.180	0.324
-----	-------	-------	-------	-------	-------	-------	-------	-------

Passive aggressive

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.651	0.574	0.702	0.696	0.712	0.727	0.727	0.635
F1: 0	0.348	0.489	0.418	0.300	0.424	0.455	0.455	0.596
Nøyaktighet	0.545	0.535	0.606	0.576	0.616	0.636	0.636	0.616
ROC AUC	0.527	0.532	0.586	0.551	0.596	0.616	0.616	0.616
MCC	0.063	0.063	0.210	0.145	0.237	0.288	0.288	0.231

LightGBM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.582	0.554	0.672	0.627	0.667	0.677	0.677	0.693
F1: 0	0.477	0.419	0.537	0.450	0.417	0.423	0.423	0.451
Nøyaktighet	0.535	0.495	0.616	0.556	0.576	0.586	0.586	0.606
ROC AUC	0.530	0.488	0.607	0.545	0.559	0.569	0.569	0.589
MCC	0.061	-0.024	0.222	0.094	0.134	0.158	0.158	0.204

LRC

Logistisk regresjon

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.316	0.327	0.271	0.281	0.262	0.297	0.304	0.373
F1: 0	0.723	0.741	0.691	0.657	0.672	0.581	0.538	0.618
Nøyaktighet	0.606	0.626	0.566	0.535	0.545	0.475	0.444	0.525
ROC AUC	0.536	0.549	0.494	0.490	0.481	0.484	0.481	0.565
MCC	0.062	0.086	-0.011	-0.016	-0.033	-0.027	-0.033	0.108

Random forest

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.054	0.000	0.000	0.000	0.000	0.071	0.061	0.000
F1: 0	0.783	0.875	0.875	0.855	0.875	0.847	0.812	0.875
Nøyaktighet	0.646	0.778	0.778	0.747	0.778	0.737	0.687	0.778
ROC AUC	0.432	0.500	0.500	0.481	0.500	0.490	0.458	0.500
MCC	-0.158	0.000	0.000	-0.094	0.000	-0.034	-0.112	0.000

KNN

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.235	0.246	0.357	0.263	0.059	0.200	0.120	0.150
F1: 0	0.735	0.695	0.746	0.825	0.805	0.797	0.703	0.785
Nøyaktighet	0.606	0.566	0.636	0.717	0.677	0.677	0.556	0.657
ROC AUC	0.487	0.477	0.571	0.542	0.451	0.500	0.406	0.471
MCC	-0.024	-0.040	0.125	0.095	-0.124	0.000	-0.174	-0.063

SVM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.280	0.291	0.200	0.241	0.207	0.000	0.276	0.354
F1: 0	0.757	0.727	0.730	0.686	0.864	0.875	0.700	0.571
Nøyaktighet	0.636	0.606	0.596	0.556	0.768	0.778	0.576	0.485
ROC AUC	0.523	0.519	0.464	0.471	0.542	0.500	0.500	0.539
MCC	0.042	0.034	-0.066	-0.051	0.137	0.000	0.000	0.066

AdaBoost

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.348	0.190	0.125	0.222	0.340	0.303	0.214	0.355
F1: 0	0.803	0.782	0.720	0.771	0.795	0.652	0.690	0.706
Nøyaktighet	0.697	0.657	0.576	0.646	0.687	0.535	0.556	0.596
ROC AUC	0.578	0.487	0.419	0.497	0.571	0.506	0.455	0.562
MCC	0.151	-0.027	-0.153	-0.006	0.137	0.011	-0.080	0.105

QDA

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.000	0.311	0.346	0.338	0.339	0.348	0.300	0.389
F1: 0	0.875	0.426	0.547	0.677	0.719	0.651	0.525	0.651
Nøyaktighet	0.778	0.374	0.465	0.566	0.606	0.545	0.434	0.556
ROC AUC	0.500	0.468	0.526	0.542	0.552	0.545	0.474	0.584
MCC	0.000	-0.058	0.044	0.071	0.089	0.076	-0.044	0.140

Passive aggressive

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.343	0.310	0.300	0.254	0.160	0.294	0.167	0.222
F1: 0	0.641	0.614	0.696	0.583	0.716	0.631	0.638	0.708
Nøyaktighet	0.535	0.505	0.576	0.465	0.576	0.515	0.495	0.576
ROC AUC	0.539	0.503	0.516	0.445	0.435	0.494	0.399	0.468
MCC	0.065	0.005	0.028	-0.092	-0.120	-0.011	-0.172	-0.058

LightGBM

Metrikk\τ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
F1: 1	0.057	0.103	0.204	0.267	0.235	0.196	0.258	0.226
F1: 0	0.798	0.780	0.738	0.784	0.735	0.721	0.662	0.717
Nøyaktighet	0.667	0.646	0.606	0.667	0.606	0.586	0.535	0.586
ROC AUC	0.445	0.448	0.471	0.526	0.487	0.458	0.474	0.474
MCC	-0.136	-0.115	-0.055	0.051	-0.024	-0.077	-0.044	-0.047

Tillegg D

Undersett 0.2-0.4 - LRC

Egenskapene presenterte i tabell D.1 var inkluderte i undersettene basert på $\tau_1 = \tau_2$ lik 0.2, 0.3 og 0.4 for LRC.

Tabell D.1: Oversikt over egenskapene inkludert i undersettene basert på verdier av $\tau_1 = \tau_2$ lik 0.2, 0.3 og 0.4 for LRC.

Egenskap	0.2	0.3	0.4
shape_Elongation	x	x	x
shape_Sphericity	x	x	x
first_order_MinimumCT_16	x	x	x
first_order_SkewnessCT_16	x	x	x
glszm_GrayLevelVarianceCT_16	x	x	x
gllm_RunEntropyCT_16	x	x	x
gldm_DependenceNonUniformityNormalized_d_1CT_16	x	x	x
glcm_MCC_d_1CT_16	x	x	x
glszm_GrayLevelNonUniformityNormalizedPET_8	x	x	x
gllm_ShortRunLowGrayLevelEmphasisPET_8	x	x	x
LBP_201PET_8	x	x	x
hvp_related	x	x	x
first_order_KurtosisPET_8	x	x	x
LBP_021PET_8	x	x	x
glszm_SmallAreaHighGrayLevelEmphasisPET_4	x	x	x
larynx	x	x	x

glszm_LargeAreaHighGrayLevelEmphasisCT_16	x	x	x
glszm_ZoneEntropyCT_16	x	x	x
first_order_EnergyPET_8	x	x	x
glszm_SmallAreaEmphasisPET_4	x	x	x
histgrade_high	x	x	x
glszm_LowGrayLevelZoneEmphasisCT_16	x	x	x
glrlm_RunEntropyCT_16	x	x	x
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16	x	x	x
LBP_120CT_16	x	x	x
LBP_210CT_16	x	x	x
glszm_SizeZoneNonUniformityPET_8	x	x	x
glcm_ClusterTendency_d_1PET_8	x	x	x
glcm_SumSquares_d_1PET_8	x	x	x
shape_Maximum2DDiameterColumn	x	x	x
shape_Maximum2DDiameterSlice	x	x	x
ngtdm_Strength_d_1CT_16	x	x	x
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16	x	x	x
glcm_MaximumProbability_d_1CT_16	x	x	x
ngtdm_Contrast_d_1PET_8	x	x	x
glcm_Imc1_d_1PET_8	x	x	x
glszm_GrayLevelNonUniformityPET_4	x	x	x
age	x	x	x
female	x	x	x
oropharynx	x	x	x
charlson	x	x	x
shape_LeastAxisLengthCT_16	x	x	x
first_order_10PercentileCT_16	x	x	x
first_order_90PercentileCT_16_	x	x	x
first_order_MedianCT_16	x	x	x
glrlm_HighGrayLevelRunEmphasisCT_16	x	x	x
glrlm_LongRunLowGrayLevelEmphasisCT_16	x	x	x
gldm_DependenceVariance_d_1CT_16	x	x	x
gldm_GrayLevelNonUniformity_d_1CT_16	x	x	x
first_order_SkewnessPET_8	x	x	x

glszm_LargeAreaHighGrayLevelEmphasisPET_8	x	x	x
glszm_GrayLevelNonUniformityPET_8	x	x	x
glszm_SizeZoneNonUniformityNormalizedPET_8	x	x	x
glszm_ZoneEntropyPET_8	x	x	x
LBP_003PET_8	x	x	x
LBP_102PET_8	x	x	x
ngtdm_Complexity_d_1PET_4	x	x	x
gldm_DependenceEntropy_d_1PET_4	x	x	x
gldm_SmallDependenceLowGrayLevelEmphasis_d_1PET_4	x	x	x
uicc8_III-IV	x	x	x
first_order_UniformityCT_16	x	x	
glszm_SizeZoneNonUniformityCT_16	x	x	
TLG	x	x	
glcm_ClusterProminence_d_1CT_16	x	x	
LBP_111CT_16	x	x	
LBP_201CT_16	x	x	
LBP_300CT_16	x	x	
first_order_InterquartileRangePET_8	x	x	
first_order_MinimumPET_8	x	x	
ngtdm_Coarseness_d_1PET_4	x	x	
glszm_SizeZoneNonUniformityNormalizedPET_4	x	x	
glszm_LowGrayLevelZoneEmphasisPET_4	x	x	
glszm_HighGrayLevelZoneEmphasisPET_4	x	x	
glszm_GrayLevelNonUniformityNormalizedPET_4	x	x	
first_order_RobustMeanAbsoluteDeviationPET_8	x		
gldm_GrayLevelVariance_d_1PET_8	x		
glcm_Imc1_d_1PET_4	x		
shape_MinorAxisLengthCT_16	x		
first_order_InterquartileRangeCT_16	x		
first_order_KurtosisCT_16	x		
glszm_LargeAreaLowGrayLevelEmphasisCT_16	x		
glszm_SmallAreaLowGrayLevelEmphasisCT_16	x		
glrlm_ShortRunEmphasisCT_16	x		
glcm_ClusterShade_d_1CT_16	x		

glcm_ClusterTendency_d_1CT_16	x
glcm_JointAverage_d_1CT_16	x
glcm_SumAverage_d_1CT_16	x
LBP_012CT_16	x
first_order_10PercentilePET_8	x
first_order_UniformityPET_8	x
glszm_GrayLevelVariancePET_8	x
glszm_HighGrayLevelZoneEmphasisPET_8	x
glszm_LargeAreaLowGrayLevelEmphasisPET_8	x
glszm_SmallAreaLowGrayLevelEmphasisPET_8	x
ngtdm_Complexity_d_1PET_8	x
gldm_DependenceEntropy_d_1PET_8	x
gldm_LargeDependenceLowGrayLevelEmphasis_d_1PET_8	x
glcm_ClusterShade_d_1PET_8	x
glcm_Imc2_d_1PET_8	x
LBP_111PET_8	x
LBP_210PET_8	x
glszm_SmallAreaLowGrayLevelEmphasisPET_4	x
glrlm_GrayLevelVariancePET_4	x
glrlm_ShortRunLowGrayLevelEmphasisPET_4	x
ngtdm_Strength_d_1PET_4	x
glcm_ClusterProminence_d_1PET_4	x
SUVpeak	x
cavum_oris	x

Tillegg E

Undersett 0.5v for LRC

Undersettet av egenskaper med vekselvirkninger som ga det beste resultatet under kryssvalidering for LRC (0.5v) besto av følgende egenskaper:

shape_Sphericity
glrlm_RunEntropyCT_16
glrlm_RunEntropyCT_16
shape_Elongation*shape_Sphericity
shape_Elongation*gldm_DependenceNonUniformityNormalized_d_1CT_16
shape_ElongationCT_16*glszm_GrayLevelNonUniformityNormalizedPET_8
shape_Sphericity*glszm_GrayLevelVarianceCT_16
shape_Sphericity*gldm_DependenceNonUniformityNormalized_d_1CT_16
shape_Sphericity*glszm_GrayLevelNonUniformityNormalizedPET_8
shape_Sphericity*glcm_Imc1_d_1PET_8
first_order_MinimumCT_16*LBP_021PET_8
first_order_MinimumCT_16*glszm_LargeAreaHighGrayLevelEmphasisCT_16
first_order_SkewnessCT_16*LBP_201PET_8
first_order_SkewnessCT_16*hpv_related
first_order_SkewnessCT_16*histgrade_high
first_order_SkewnessCT_16*glszm_SizeZoneNonUniformityPET_8
first_order_SkewnessCT_16*shape_Maximum2DDiameterSlice
glszm_GrayLevelVarianceCT_16*glszm_GrayLevelNonUniformityNormalizedPET_8
glszm_GrayLevelVarianceCT_16*hpv_related
glrlm_RunEntropyCT_16²
glrlm_RunEntropyCT_16*LBP_120CT_16
glrlm_RunEntropyCT_16*LBP_210CT_16
gldm_DependenceNonUniformityNormalized_d_1CT_16*glszm_GrayLevelNonUniformityNormalizedPET_8
gldm_DependenceNonUniformityNormalized_d_1CT_16*hpv_related
gldm_DependenceNonUniformityNormalized_d_1CT_16*glcm_ClusterTendency_d_1PET_8
gldm_DependenceNonUniformityNormalized_d_1CT_16*glcm_Imc1_d_1PET_8
glcm_MCC_d_1CT_16*glszm_LowGrayLevelZoneEmphasisCT_16

glcm_MCC_d.1CT_16*LBP_120CT_16
 glcm_MCC_d.1CT_16*LBP_210CT_16
 glszm_GrayLevelNonUniformityNormalizedPET_8*glszm_SmallAreaHighGrayLevelEmphasisPET_4
 glszm_GrayLevelNonUniformityNormalizedPET_8*glcm_ClusterTendency_d.1PET_8
 glszm_GrayLevelNonUniformityNormalizedPET_8*shape_Maximum2DDiameterSliceCT_16
 glszm_GrayLevelNonUniformityNormalizedPET_8*ngtdm_Contrast_d.1PET_8
 glszm_GrayLevelNonUniformityNormalizedPET_8*glcm_Imc1_d.1PET_8
 glszm_GrayLevelNonUniformityNormalizedPET_8*female
 glrlm_ShortRunLowGrayLevelEmphasisPET_8*first_order_KurtosisPET_8
 glrlm_ShortRunLowGrayLevelEmphasisPET_8*larynx
 glrlm_ShortRunLowGrayLevelEmphasisPET_8*ngtdm_Strength_d.1CT_16
 LBP_201PET_8²
 LBP_201PET_8*LBP_021PET_8
 LBP_201PET_8*larynx
 LBP_201PET_8*glszm_LowGrayLevelZoneEmphasisCT_16
 hpv_related²
 hpv_related*first_order_KurtosisPET_8
 hpv_related*glszm_SmallAreaHighGrayLevelEmphasisPET_4
 hpv_related*LBP_120CT_16
 hpv_related*shape_Maximum2DDiameterSliceCT_16
 hpv_related*ngtdm_Strength_d.1CT_16
 first_order_KurtosisPET_8*larynx
 first_order_KurtosisPET_8*glszm_LowGrayLevelZoneEmphasisCT_16
 first_order_KurtosisPET_8*glszm_SizeZoneNonUniformityPET_8
 first_order_KurtosisPET_8*ngtdm_Contrast_d.1PET_8
 first_order_KurtosisPET_8*female
 LBP_021PET_8*larynx
 LBP_021PET_8*gldm_LargeDependenceHighGrayLevelEmphasis_d.1CT_16
 LBP_021PET_8*ngtdm_Strength_d.1CT_16
 LBP_021PET_8*gldm_LargeDependenceHighGrayLevelEmphasis_d.1CT_16
 LBP_021PET_8*oropharynx
 glszm_SmallAreaHighGrayLevelEmphasisPET_4²
 glszm_SmallAreaHighGrayLevelEmphasisPET_4*larynx
 glszm_SmallAreaHighGrayLevelEmphasisPET_4*histgrade_high
 glszm_SmallAreaHighGrayLevelEmphasisPET_4*ngtdm_Strength_d.1CT_16
 glszm_SmallAreaHighGrayLevelEmphasisPET_4*female
 larynx*glszm_LargeAreaHighGrayLevelEmphasisCT_16
 larynx*first_order_EnergyPET_8
 glszm_LargeAreaHighGrayLevelEmphasisCT_16*glcm_SumSquares_d.1PET_8
 glszm_LargeAreaHighGrayLevelEmphasisCT_16*ngtdm_Contrast_d.1PET_8
 glszm_LargeAreaHighGrayLevelEmphasisCT_16*female
 first_order_EnergyPET_8²
 first_order_EnergyPET_8*glszm_SizeZoneNonUniformityPET_8
 first_order_EnergyPET_8*ngtdm_Strength_d.1CT_16
 first_order_EnergyPET_8*glszm_GrayLevelNonUniformityPET_4
 first_order_EnergyPET_8*female
 glszm_SmallAreaEmphasisPET_4*female
 glszm_SmallAreaEmphasisPET_4*oropharynx
 histgrade_high*charlson

glszm_LowGrayLevelZoneEmphasisCT_16*glszm_GrayLevelNonUniformityPET_4
glrlm_RunEntropyCT_16²
glrlm_RunEntropyCT_16*LBP_120CT_16
glrlm_RunEntropyCT_16*LBP_210CT_16
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16*LBP_210CT_16
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16*shape_Maximum2DDiameterColumnCT_16
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16*age
LBP_210CT_16*gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16
shape_Maximum2DDiameterColumnCT_16*gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16
shape_Maximum2DDiameterSliceCT_16*ngtdm_Strength_d_1CT_16
gldm_LargeDependenceHighGrayLevelEmphasis_d_1CT_16*age
female*charlson



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway