



**AALBORG
UNIVERSITET**

Bridging the Gap

ET FEASIBILITYSTUDIE OM EN BRUGERCENTRERET FORKLARENDE
BRUGERGRÆNSEFLADE TIL OPTIMERING AF BRUGERVENLIGHED I
PRÆDIKTIVE ALGORITMER FOR TELEMONITORERING

KLINISK VIDENSKAB OG TEKNOLOGI, KANDIDATSPECIALE, AALBORG UNIVERSITET

**Title:** Bridging the gap:

A Feasibility study on User-Centered Explainable User Interface for Enhancing Usability of Predictive Algorithms for Telemonitoring

Type: Master thesis

Aalborg University, 4. Semester

Project period: 01.02.2023 – 02.06.2023

Project group: 508

Authors:

Ingunn Ryggstein - 20211829

Jennifer Dan Thu Lam - 20211829

Supervisor:

Thomas Kronborg Larsen, Associate Professor

Co-Supervisor:

Alexander Arndt Pasgaard Xylander. PhD Fellow

Number of characters incl. spaces:

113.896

Number of pages: 57 pages

Number of appendixes: 13

Reference system: Vancouver

Due date: June 01, 2023

Abstract

Introduction: Early intervention for exacerbations is crucial for enhancing patients' quality of life, reducing mortality and impeding the progression of COPD. TeleCare Nord has pilottested a predictive algorithm which should help identify exacerbations 3 days prior. However, deploying AI outputs in healthcare without explainability is a barrier for effective use. This master's thesis aims to investigate the applicability of an explainable user interface (XUI) within telemedicine to support the remote monitoring of citizens with COPD.

Methodology: Hevner's design science research is used as a framework for the design and evaluation of the prototype. The framework involves exploring the environment and the knowledge base. The environment was explored by conducting a semi-structured interview with two telemedicine practitioners. The knowledge base was explored by conducting a systematic literature search for XUI within a healthcare context. These methods laid the foundation for development and evaluation of a prototype, which consisted of a two-part evaluation with two evaluators to assess the usability of the prototype.

Results: Four requirements were identified: a global and local explanation, increased or decreased risk, as well as an interactive graph. The most significant result indicated that the end-user preferred the global explanation over the local explanation. Additionally, the application-impact of the interactive graph did not meet the initial expectations.

**Titel:** Bridging The Gap

Et feasibility studie om en brugercentreret forklarende brugergrænseflade til optimering af brugervenlighed i prædiktive algoritmer for telemonitorering

Type: Kandidat Speciale

Aalborg Universitet, 4. semester

Projektperiode: 01.02.2022 – 30.05.2022

Projektgruppe: 508

Forfattere:

Ingunn Ryggstein - 0211829

Jennifer Dan Thu Lam - 20210690

Vejleder:

Thomas Kronborg Larsen, lektor

Bi-Vejleder:

Alexander Arndt Pasgaard Xylander. Ph.d.-stipendiat

Antal anslag inkl. mellemrum: 113.986

Sideantal: 57 sider

Bilagsantal: 13

Referencesystem: Vancouver

Afleveringsdato: 01. juni 2023

Abstrakt

Introduktion: Tidlig intervention for exacerbation er afgørende for at forbedre borgernes livskvalitet, reducere dødeligheden og sygdomsprogression af KOL. TeleCare Nord har gennemført et pilotstudie af en prædiktiv algoritme, der skal hjælpe med at identificere exacerbationer tre dage forud. Imidlertid udgør implementering af AI-outputs uden forklaring en hindring for effektivt brug. Dette kandidatspeciale har til formål at undersøge anvendeligheden af en forklarende brugergrænseflade (XUI) til at understøtte telemonitorering af borgere med KOL.

Metode: Hevners design science research er anvendt som ramme for design og evaluering af prototypen. Rammen indebærer udforskning af miljøet og vidensbasen. Miljøet blev udforsket ved at gennemføre et semistruktureret interview med to telemedicinsk sundhedsfaglige aktører. Vidensbasen blev afdækket ved en systematisk litteratursøgning med fokus på XUI inden for sundhedsområdet. Metoderne var byggesten for udvikling og evaluering af prototypen. Evalueringen bestod af en todelt evaluering med to evaluators, som vurderede brugervenligheden.

Resultat: Fire krav blev identificeret: en global og lokal forklaring, øget eller reduceret i risikoen samt en interaktiv graf. Det mest betydningsfulde resultat viste, at slutbrugeren foretrak den globale forklaring over lokale forklaring. Den interaktive graf levede ikke op til forventningerne om at være anvendelig.

Konklusion: Evalueringen bekræftede, at en brugercentreret XUI kunne understøtte slutbrugere i at vurdere risikoen for en exacerbation. Den globale forklaring blev foretrukket frem for den lokale forklaring, og den interaktive graf medvirkede ikke til et retrospektivt overblik for borgerens målinger. Baseret på disse uoverensstemmelser, er yderligere forskning nødvendigt for en succesfuld implementering.

Forord

Kandidatspecialet er udarbejdet i perioden 31. januar til 1. juni 2023, som afsluttende projekt på kandidatuddannelsen Klinisk Videnskab og Teknologi ved Det Sundhedsvidenskabelige Fakultet på Aalborg Universitet.

Specialgruppen er bestående af Ingunn Ryggstein, radiograf med ekspertise inden for radiografi, og Jennifer Dan Thu Lam, bioanalytiker med ekspertise inden for bioanalytisk diagnostik.

Specialet henvender sig til TeleCare Nord, og de personer i fælles nordjyske servicefunktion, der beskæftiger sig med udvikling og forbedring af Telma.

Formålet med dette speciale er at undersøge om en brugercentreret explainable user-interface kan have potentiale og/eller barrierer for anvendelse af en prædiktiv algoritme til monitorering af borgere med KOL.

Tak til TeleCare Nord for at stille data til rådighed, og til projektleder Ria Thaarup Høegh for at være behjælpelig med gennemgang af informationssystemer anvendt i telemonitorering samt rekruttering af slutbrugere.

Tak til vores hovedvejleder, Thomas Kronborg Larsen, og vores bi vejleder, Alexander Arndt Pasgaard Xylander, for sparring, konstruktiv vejledning og et godt samarbejde.

Sidst skal der også rettes en stor tak til de slutbrugere og UX-eksperter, der har deltaget i en af specialets anvendte metoder.

.

Læsevejledning

Denne specialerapport indeholder 63 sider, og er fordelt på 11 kapitler, og anbefales at læses sekventielt. I et separat dokument vedhæftes 13 bilag, som er præsenteret i kronologisk rækkefølge efter indtræden i specialet.

Til håndtering af bibliografiske formater, anvendes referencesystemet Vancouver og vil fremgå som tal til slut i sætningen. Præsentation af studier med en eller to forfattere refereres med efternavn. Ved tre eller flere forfattere bliver først forfatterens efternavn nævnt efterfulgt af et al.

Anvendte udtalelser som citeres præsenteres i citationstegn og kursiv skrifttype, hvor kilden fra udtalelsen forekommer direkte efter sidste citationstegn.

Tal fra et til ti vil blive skrevet ud, hvor tal efter ti bliver angivet i tal for at bidrage til læsevenligheden. Dog vil der være undtagelser ved år, beløb og procentsats, som vi præsenteres med tal og passende enhed.

Ordene "informant" og "evaluator" anvendes som synonyme for deltagende i specialets metoder. Ved henvisning til "slutbruger" refereres der til reelle slutbruger af it-applikationen.

Begrebsafklaring

Exacerbation:

En KOL-eksacerbation defineres som en akut forværring af de respiratoriske symptomer (åndenød, hoste og ekspektoration), som nødvendiggør en supplerende behandling (1).

GOLD-kriterier:

Et sæt af kriterier udarbejdet af Global Initiative for Chronic Obstructive Lung Disease (GOLD) til at klassificere sværhedsgraden af luftvejsobstruktion ved KOL. Denne klassificering består af fire stadier: mild, moderat, svær og meget svær. Denne klassificering består yderligere af en kombineret "ABE" tilgang til at vurdere patienter baseret på deres symptomer og tidligere historie med forværringer (1).

TeleKit:

Er en betegnelse for det telemedicinske udstyr borgerne får udleveret via TeleCare Nord ved opstart i telemonitorering. Den består af en tablet (elektronisk skærm), en iltmætningsmåler (til KOL-patienter) og evt. en vægt (2).

Telemedicinsk sundhedsfaglige aktører:

Er en betegnelse for forskellige sundhedsprofessionelle der varetager bruger og patientadministrationen ved telemonitorering, herunder opfølgning på og patientkontakten (3)

Prædiktive algoritme:

Når der henvises til en specifik algoritme eller prædiktionsalgoritme, refereres der til en algoritme udviklet af Aalborg Universitet (4) .

Indholdsfortegnelse

1. Introduktion	1
2. Problemanalyse	2
2.1 Kronisk Obstruktiv Lungesygdom	2
2.2 Omkostninger forbundet med KOL	3
2.3 TeleCare Nord og TeleKit	3
2.4 Udfordringer ved manuel monitorering	4
2.5 Kunstig intelligens som beslutningsstøtteværktøj	5
2.5.1 Udfordringer ved implementering af kunstig intelligens	5
2.6 State of art: Forklarlighed for en AI-model	6
2.7 Udgangspunkt for videre forskning	7
2.7.1 Slutbruger og kontekst	7
3. Problemafgrænsning	8
4. Præsentation af case	9
5. Metode	9
5.1 Pragmatisme	9
5.2 User-Centered Design (UCD)	10
5.3 Specialets design for udvikling af en XUI	10
5.3.1 Vidensbasen (Litteratursøgning)	11
5.3.2 Miljøet (Interview)	12
5.3.3 Udvikling og evaluering af prototype	12
5.4 Vidensbase (Litteratursøgning)	12
5.4.1 Valg af databaser	14
5.4.2 Initierende søgeprofil	14
5.4.3 In- og eksklusionskriterier	16
5.4.5 Kvalitetsvurdering af artikler	17
5.4.5 Ekstrahering af data fra artiklerne	17
5.5 Miljøet (Interview)	18
5.5.1 Rekruttering	19

5.5.2 Interviewguide	19
5.5.3 Udførelse af interview	19
5.5.4 Ekstrahering af data fra dyadeinterview	20
5.6 Prototypeudvikling	20
5.6.1 Design behov.....	20
5.6.2 Skitsering	21
5.6.3 Visualisering	21
5.7 Evaluering af prototype	21
5.7.1 Rekruttering.....	22
5.7.2 Udarbejdelse af evalueringen	22
5.7.3 Pilottest af evalueringen	23
5.7.4 Databehandling af evaluering.....	23
6. Projektets resultater	25
6.1 Vidensbase: Resultater fra litteratursøgningen.....	25
6.1.1 Præsentation af videnbasens inkluderede artikler.....	26
6.1.1 Fokusspørgsmål 1: Behov for anvendelse af XUI i sundhedsvæsenet.....	27
6.1.2 Fokusspørgsmål 2: Udvikling af en brugervenlig visualisering af algoritmens prædiktion til klinisk brug.....	28
6.1.3 Fokusspørgsmål 3: Standardiseret metode og værktøjer til evaluering af en XUI	30
6.2 Miljøet: Resultater fra interview	30
6.2.1 Præsentation af informanter	30
6.2.2 Vurdering af data.....	31
6.2.3 Anvendelighed af algoritme.....	32
6.2.4 Et behov for information.....	32
6.2.5 Et samlet overblik.....	33
6.3 Sammenfatning af krav til prototypen.....	34
6.4 Prototypen	35
6.4.1 Opgaveoversigt.....	35

6.4.2 Monitoreringsoversigt	37
6.5 Resultater fra evaluering af prototype	39
6.5.1 Præsentation af evalutore.....	39
6.5.2 Heuristiske overtrædelser	40
6.5.3 Prototypens krav og grad af overtrædelse	41
6.5.4 Udtalelser fra evalueringens.....	43
7. Diskussion.....	51
7.1 Vidensbase.....	51
7.2 Miljøet	52
7.3 Evalueringen.....	52
7.1.1 Imødekommelse af krav 1 og 2 - Global og lokal forklaring.....	53
7.1.3 Imødekommelse af krav 3 - Interaktiv graf	54
7.3 Videre forskning.....	55
8. Konklusion.....	56
9. Perspektivering.....	57
10. Bibliografi	58
11. Bilag.....	71

1. Introduktion

I kommende år vil antallet af ældre stige i Danmark (5). Her spiller alderdommen en rolle, både indirekte og direkte indikator for øget forekomst af kroniske sygdomme, som kronisk obstruktiv lungesygdom (KOL) (6). På nuværende tidspunkt er der ingen officielle statistikker over antallet af borgere med KOL, men det estimeres, at cirka 3-400.000 danskere har en lungefunktionsnedsættelse forenelig med KOL (7). Hvert år bliver ca. 20.000 borgere indlagt på landets lungeafdelinger, hvoraf 20% genindlægges i løbet af den første måned efter udskrivelsen (7). Data fra Sundhedsdata Styrelsen viser også, at borgere med KOL i gennemsnit fire gange så mange indlæggelser og syv gange flere genindlæggelser end den generelle befolkning (8). Baseret på dette samspil, forventes KOL at blive en af de dominerende sygdomme på landets hospitalsafdelinger (6-8). Der forventes også, at denne stigning vil have en særlig indvirkning på ressourceforbruget i sundhedsvæsenet (6-8).

Danmark har den højeste KOL-relaterede dødelighed pr. indbygger i Europa, med ca. 5.500 dødsfald om året (7,9). Dette er som konsekvens af forværringer (exacerbationer), som forårsager skader på lungevæv og bronkier (10). En exacerbation er en subjektiv vurdering og kan opstå pludseligt (akutte exacerbationer) (8,11). For at mindske risikoen for exacerbationer er det vigtigt med tidlig opsporing og igangsætte tidlig forebyggende behandling (11). I denne sammenhæng blev der iværksat et initiativ i form af et pilotprojekt med fokus på monitorering af borgere fra eget hjem (12). For at vurdere gevinster ved telemonitorering blev projektet udført i storskala i perioden 2012-2015, hvor telemedicin blev tilbudt til borgere med KOL (13). Som følge af den daværende business case for projektet, blev der indgået en aftale i forbindelse med økonomiaftalerne mellem Regeringen, KL og Danske Regioner, om at gå fra et projekt i storskala til et projekt med landsdækkende implementering inden udgang af 2019 (14).

Efterfølgende år blev der indgået en aftale om at afprøve anvendelse af kunstig intelligens (AI) inden for sundhedsområdet, velfærd og grøn omstilling (15). Dette førte til 15 signaturprojekter, herunder "*Prædiktation af forværring blandt KOL- og hjertesvigtpatienter*", som havde til formål at muliggøre hurtigere identificering af exacerbationer (15). Dog er der barrierer for anvendelse af AI i klinisk praksis (16), Hvor forklarlighed af algoritmen er en væsentlig faktor for at sikre brugerens accept og opretholdelse af tillid til AI-systemet (17)

I nedenstående problemanalyse vil der blive redegjort for behovet for en forklarlig algoritme til at understøtte personalet til monitorering af borgere med KOL for exacerbationer.

2. Problemanalyse

I det kommende afsnit belyses de kliniske, organisatoriske og tekniske udfordringer, der kan være forbundet med anvendelsen af AI. Dette leder op til specialets problemformulering.

2.1 Kronisk Obstruktiv Lungesygdom

KOL er en progressiv og irreversibel lungesygdom, som primært forekommer hos den ældre befolkning, hvor der gradvis er sket en forværring over årene (11). Den største årsag skyldes gentagen eksponering for tobaksrøg, hvor 85-90% af alle tilfælde af KOL skyldes rygning (18). Da sygdommen udvikler sig over 30-50 år, opfattes den oftest som en normal del af aldringsprocessen, og derfor tilpasser borgerne sig det lave funktionsniveau i lungerne (11). Denne accept af symptomerne resulterer i, at mange borgere har en reduceret lungefunktion på 50% sammenlignet med raske personer ved diagnosetidspunktet (18).

Nogle borgere er i større risiko for at opleve exacerbation end andre, hvor risikoen afhænger af borgerens stadie af KOL samt mængden af symptomer, som klassificeres ud fra Det Globale Initiativ for KOL (GOLD) (19). Her bliver der taget udgangspunkt i borgerens symptomer såsom antal exacerbationer over et år, antallet af hospitalsindlæggelser og resultater fra en lungefunktionsmåling (19). Dette muliggør vurdering af sværhedsgrad fra mild KOL med få exacerbationer til meget svær med over to exacerbationer årligt (20). En exacerbation varer typisk i syv til ti dage, og kan i gennemsnit tage op til 14 dage, før lungefunktionen er genvundet (11). Dog vil restitutionen nedsættes for hvert ekstra år med sygdommen, hvilket gør det sværere for patienterne at komme sig efter exacerbationer over tid (21). Ved akutte exacerbationer kræves der øjeblikkelig behandling og ofte altid indlæggelse, da det er den hyppigste dødsårsag hos lungepatienter (11). Dette skyldes, at akutte exacerbationer har en hazard ratio på 1,71, og forekomsten af fem eller flere akutte exacerbationer over tid medfører af hazard ratioen på 3.41 (21). Derfor er det vigtigt at identificere og behandle exacerbation tidligt for at minimere dens negative effekt på borgerens sundhedstilstand og forebygge exacerbationer (1).

Der kan findes utallige former for medicin til behandling af borgere med KOL (11). Behandling afhænger af borgerens sygdomsstadie og sigter mod at lindre aktuelle

symptomer samt virke forebyggende for exacerbationer (11). Udbrud for exacerbation indtræder oftest inden for 14 dage, hvor det medfører en stigning af mindst to til tre af følgende symptomer: åndenød, hoste, purulent eller slimproduktion (1). Exacerbation kan klassificeres som mild, moderat og svær (1,11). Mild kan behandles i eget hjem med bronkodilatorer, moderat kan behandles med yderligere orale kortikosteroider og antibiotika, og svær exacerbation kræver indlæggelse (1,11). På trods af bestræbelserne på at behandle og forebygge exacerbationer, kan det være en udfordring for klinikere og borgere at vurdere, om der er tegn på exacerbation ud fra symptomer (22). Det skal bemærkes, at symptomerne for exacerbationer er uspecifikke og kan skyldes andre sygdomme (22).

2.2 Omkostninger forbundet med KOL

VIVE udgav i 2004 en rapport, der sammenlignede omkostningerne for KOL-patienter og en kontrolgruppe af patienter uden KOL (23). Rapporten kunne konkludere, at behandlingen af 66.000 KOL-patienter medførte en årlig omkostning på cirka 3 milliarder kroner (23). Dette beløb udgør omkring 10% af de samlede sygehus- og sygesikringsomkostninger i Danmark (23). I sammenligning med kontrolgruppen uden KOL var de årlige omkostninger på 1,9 milliarder danske kroner, hvilket tydeligt viser, at KOL udgør en betydelig sundhedsøkonomisk byrde (23). Derfor vurderes det, at tidlig og korrekt behandling af borgere med exacerbationer kan gavne Danmarks økonomi og bidrage til at øge borgerens livskvalitet (24).

2.3 TeleCare Nord og TeleKit

For at kunne iværksætte hurtigere behandling af borgere med KOL har den danske regering, regionerne og Kommunernes Landsforening haft øget fokus på en national udbredelse af telemedicinske løsninger siden 2012 (25,26). Projektet TeleCare Nord blev derfor etableret i Region Nordjylland for at tilbyde telemedicinsk hjemmemonitorering af kronikere, herunder borgere med KOL (13). Den telemedicinske løsning kaldes for TeleKit (13). TeleCare Nord har planlagt en national udbredelse til alle relevante borgere med KOL, blandt andet borgere med to eller flere tidligere exacerbationer (25,26). Selvom løsningen viste lovende resultater ved at mindske antal indlæggelser og forbedre borgernes livskvalitet, blev der ikke observeret nogen statistisk signifikant effekt sammenlignet med standardbehandling (27). For yderligere at forbedre resultaterne fra løsningen vurderes der et behov for en bedre og mere pålidelig tilgang til identificering af enkelte borgers risiko for exacerbation (4).

Formålet med TeleKit er at overvåge samt sikre rettidig og korrekt behandling af patienter med KOL i deres eget hjem, samtidig som borgerens sundhedskompetencer skal styrkes (26). TeleKit består af en tablet med forudindstillede alarmgrænser, spørgeskemaer samt måleudstyr til puls, iltmætning, blodtryk og vægt (28). I begyndelsen vil borgeren bruge TeleKit dagligt i 14 dage, hvorefter der i samråd med almen praksis eller lungeafdelingen udformes personlige alarmgrænse på baggrund af de målte værdier (28). Baseret på alarmgrænserne er der udformet retningslinjer for, hvordan de telemedicinske sundhedsfaglige aktører (TSA), skal agere ved monitorering af borgerens sundhedstilstand (28,29). En grøn alarm betyder normale forhold, gul betyder observation og eventuel kontakt til borgeren, og rød betyder dårlig sundhedstilstand med behov for kontakt og behandling (29). De personlige alarmgrænser skal bidrage til en mere nøjagtig triagering og håndtering af borgerens sundhedstilstand (28). Når alarmgrænserne er indstillet, skal borgeren måle sine værdier og besvare spørgeskemaet to gange om ugen (28). Alle målingerne og besvarelsene bliver automatisk videresendt til TSA'erne via arbejdsplatform OpenTeleHealth (OTH) (28). En gang om ugen vurderer TSA borgernes indsendte materiale i OTH og vender tilbage til borgerne efter behov, med mulighed for at afholde videokonsultationer og booke ydelser til patienterne (28).

2.4 Udfordringer ved manuel monitorering

Introduktionen af TeleKit medførte nye arbejdsgange, som har udfordret både almen praksis og kommunernes nære sundhedsvæsen, da det er den primære sektor, der skulle varetage monitorering (13). En af de nye opgaver bestod af manuel monitorering, hvor borgerens værdier skal læses og vurderes, og efterfølgende sammenholdes med proaktive tegn på ændringer (28). Ved tegn eller tvivl om exacerbation skal TSA'en kontakte borgeren (28). I fællesskab vurderes det, om der er behov for ændring i behandlingsplanen eller opstart af en ny behandlingsplan (30). Hvis værdierne kræver lægelig vurdering, skal TSA'en henvise borgeren til sin ansvarlige behandler, som enten vil være egen praktiserende læge eller sygehuset, for eventuelle behandling (30).

Denne nye monitoreringsopgave beskrives i studiet af Christensen som udfordrende, især for de kommuner, hvor der ikke er specialiserede sygeplejersker i KOL til at varetage opgaven (31). I studiet blev der observeret forskelle i, hvordan TSA var i stand til at arbejde med og tolke de indsamlede data (31). TSA'en med viden om KOL og den tilhørende sygdomsforståelse var bedre til at observere og reagere på forværringer, sammenlignet med dem uden denne viden (31). Der er derfor behov for et kompetenceløft for at forbedre monitoreringen hos uerfarne kommunale sygeplejersker

(31). Udover erfaring og kompetencer kan der være andre faktorer, herunder den menneskelige kognitive funktion, der påvirker vurderingen af borgerens tilstand (32). Fejl kan opstå i den manuelle vurdering af komplekse opgaver, da de sjældent udføres perfekt af mennesker, selv under ideelle forhold (32). Kombinationen af både manglende kompetencer og risiko for menneskelige fejl ved den manuelle monitorering vurderes derfor at udgøre en hindring for tidlig identificering af exacerbationer.

2.5 Kunstig intelligens som beslutningsstøtteværktøj

For at forbedre telemedicinske løsninger har AI vist sig anvendeligt til forskellige opgaver, såsom patientovervågning, informationsteknologi, intelligent assistance, diagnose og samarbejde om informationsanalyse (33). Dette vidner om, at anvendelsen af AI som et beslutningsstøtteværktøj kan minimere fejl, der opstår under manuel monitorering, og dermed sikre en ensartet vurdering på tværs af forskellige kompetenceniveauer (34–36).

I forbindelse med signaturprojektet fra 2020 "*Prædiktion af forværring blandt KOL- og hjertesvigtpatienter*", blev der udviklet en prædiktiv algoritme, som kunne give beregne en sandsynlighed for exacerbation tre døgn forud (15,37). Implementeringen af den prædiktive algoritme skulle medvirke til hurtigere identificering af exacerbationer (15,37). Dette blev gjort ud fra en beregning af værdierne for puls, diastolisk blodtryk og saturation, som borgeren allerede måler ved brug af TeleKit (37). Denne nøjagtighed blev vurderet klinisk relevant for telemonitorering af borgere (37). Algoritmen er blevet testet i OTH, hvor der afventes svar på dens potentiale for at understøtte tidlig identifikation af exacerbation (4,38).

2.5.1 Udfordringer ved implementering af kunstig intelligens

Udvikling af AI-systemer omfatter både teknologiske aspekter (forbundet med naturvidenskab og teknik) og sociale aspekter (forbundet med samfundsvidenskab og humaniora), men de bliver ofte holdt adskilt fra hinanden (39). Ved udviklingen af algoritmer tages oftest kun forbehold for de tekniske aspekter (40). Fordi udviklingen af et AI-system indeholder både teknologiske og sociale aspekter, skal begge inddrages, idet algoritmens effektivitet er afhængig af, hvorvidt mennesket forstår algoritmens beslutningsprocessen (37,39,41). Manglende forståelse for algoritmens beslutningsproces udgør en af de største udfordringer ved anvendelse inden for sundhedsvæsenet (42,43). Dette skyldes, at sundhedspersonalet har behov for at kunne argumentere for deres valg, hvilket kræver forståelse for algoritmens resultater (43–45). For at imødekomme denne udfordring arbejder forskere med Explainable Artificial Intelligence (XAI), som omfatter algoritmer og metoder til at gøre AI-systemers adfærd

gennemsigtig og forståelig (46,47). Dette er ikke uden vanskeligheder, da forskerne ikke er enige om, hvad forklaring og forståelse er, samt hvornår det er opnået (48,49).

På trods af manglende definition af forklaring og forståelse mener forskere og udviklere, at visse algoritmer i sig selv er forståelige, mens andre kræver yderligere forklaringer for at kunne forklares (50,51). Algoritmer, som siges at være forståelige betegnes som white-box, mens algoritmer, der er svære at forstå og kræver XAI for at kunne forklare deres output, betegnes black-box (52). Den algoritme, som er udviklet til telemonitorering ved TeleCare Nord, er en logistisk regressionsmodel, som kategoriseres som en white-box algoritme, og derfor forventes den at være forståeligt af TSA'erne (53). Selvom white-box-algoritmer i teorien bør være forståelige, hersker der stadig tvivl over for dette argument (50). Lipton er en af de forskere, som mener, at ingen algoritme kan være forklarlig nok i sig selv, når menneskers begrænsede kognitivitet tages i betragtning (50). Det kan ikke garanteres, at modtageren af algoritmen kan forstå den, selvom det er en white-box-algoritme (54). Derfor bør udformningen af en forklaring, der er forståelig, altid tage udgangspunkt i brugerens eksisterende viden og mål for at modtage forklaringen (54).

2.6 State of art: Forklarlighed for en AI-model

Målet med XAI er at muliggøre, at slutbrugere bedre kan forstå, have tillid til og effektivt styre AI-systemer (55). For at nå dette mål er det vigtigt at anerkende, at XAI ikke kan stå alene, da forklarligheden eller effekten af en forklaring afhænger af brugerens opfattelse og reaktion (54). Derfor er udvikling af forklarlige modeller og forklarlig brugergrænseflade (XUI) nødvendig (47). Ved XAI er der ofte fokus på udviklingen af forklarlige modeller, men en effektiv XAI kræver også en velfungerende XUI (39,54). En XUI kan defineres som summen af output fra en XAI-proces, der præsenteres for brugerne, og som er forståelig og intuitiv (47). Det handler om at udvikle en brugergrænseflade (UI), hvor XAI-forklaringerne er præsenteret på en forståelig og intuitiv måde for brugerne (47)(figur 1). Derfor betragtes XAI som både en designudfordring og en algoritmisk udfordring (55). Selvom der er meget fokus på udviklingen af XAI-metoder, er der begrænset forskning i XUI (47).

Der findes ikke en XAI-metode, som passer alle brugere (54). De forskellige XAI-metoder kategoriseres ofte som post-hoc fortolkningsanalyse og ad-hoc modellering af fortolkningsmuligheder (56). Post-hoc fortolkning forsøger at forklare AI-forudsigelsen uden at forklare den interne mekanisme for denne model, mens ad-hoc modellering

forsøger at generere forklarlige modeller, herunder white-box modeller (57). Lipton foreslår dog, at post-hoc analyser også kan være nyttige ved white-box-algoritmer (50).

De største udfordringer ved at designe en XAI handler om at forstå, hvordan det enkelte menneske forventer, at en forklaring skal være (58). Derfor bør der først undersøges, hvilket behov brugerne har, hvorefter der vælges en metode med dette udgangspunkt (54). Inden for sundhedsområdet findes forskellige professioner med forskellige behov afhængig af hvilken sundhedsprofessionel der anvender algoritmen og til hvilket formål den bliver anvendt i (59,60). Læger har brug for en fleksibel forklaring, hvor de kan fremsøge yderligere informationer om nødvendigt, mens sygeplejersker ønsker simpel information som de nemt kunne handle ud fra (60). Dermed har sundhedsprofessionelle har forskellige behov, og derfor er det vigtigt at undersøge TSA'ernes behov for anvendelse af en algoritme til telemonitorering. Dette understreger vigtigheden af at tage hensyn til slutbrugeren og anvendelseskonteksten i udvikling af en XUI, hvor en brugercentreret tilgang vil bidrage til at øge brugernes accept og anvendelse af algoritmen (54,61).

2.7 Udgangspunkt for videre forskning

Inden pilottest af den prædiktive algoritme udførte Bender et al. en evaluering på patientrelaterede perspektiver (38). Studiet kunne konkludere, at borgerne generelt var positive over for inkluderingen af en prædiktiv algoritme, dog blev der rapporteret en mindre teknisk udfordring af en TSA'er, der bør adresseres ved videre forskning. Udfordringen omhandlede designet af platformen, hvilket skabte forvirring for TSA'en, som forventede at se et flag ved exacerbation og ikke en alarmklokke (38). Denne forventning medførte forvirring, for hvad alarmklokken skulle signalere (38). Der blev ikke yderligere udført udvikling og test af algoritmen, for at undersøge, hvor potentielle kliniske udfordringer kunne opstå.

2.7.1 Slutbruger og kontekst

Baseret på resultater fra Bender et al., og signaturprojektet *"Kunstig intelligens til prædiktation af forværringer ved multisygdom med letforståelig forklaring af komplekse resultater"* blev der gennemført et semesterprojekt på kandidatuddannelsen Klinisk Videnskab og Teknologi, som undersøgte visualisering af komplekse resultater til monitorering af borgere med KOL (62).

Sander & Madsen havde et centralt fokus på, hvordan den prædiktive algoritmes outputtet kunne visualiseres på en mere forståelig måde. Der blev blandt andet undersøgt TSA'ernes holdning til algoritmen, som nu blev visualiseret med en rød klokke og

mulighed for at få tekstbaseret forklaring "*risiko for exacerbation grundet puls og saturation*" i OTH (38,62). Resultaterne indikerer, at der på trods af tekstuelle forklaringer stadig var manglede forståelse, hvordan og hvorfor algoritmen alarmerede (62). Derudover gav informanterne udtryk for, at de savnede muligheden for at sammenholde risikoen med borgerens værdier (62).

Baseret på oplysningerne fra informanterne blev der udviklet en prototype for at imødekomme personalets behov og udfordringer (62). Prototypen havde primært fokus på at præsentere algoritmens output på en mere intuitiv måde ved hjælp af visualisering (62) dette ikke understøtter tidligere udtalelser fra TSA'erne om behov for yderligere information om virkemåden (62).

Derfor vurderes det relevant at genoverveje udviklingen af prototypen med fokus på brugerbehov, som blev identificeret af Sander & Madsen. Dette bør ligeledes ske i forbindelse med den landsdækkende udbredelse af telemedicin, hvor der anvendes et nyt it-system kaldet Telemedicinsk Monitorerings Applikation (Telma) (71).

3. Problemafgrænsning

For at nedsætte progressionen af KOL og forebygge mod exacerbationer er det vigtigt at identificere og behandle exacerbationer tidligt. Telemonitorering blev iværksat for at kunne understøtte monitorering af borgerne, men der er behov for yderligere at forbedre løsningen. En prædiktiv algoritme har potentiale til at understøtte TSA'erne i at identificere optræk til på exacerbationer, men for at algoritmen kan være effektiv i klinisk praksis, er der behov for at forklare dens beslutningsproces. Anvendelsen af algoritmen vurderes vigtigt, idet den kan medvirke til reduktion af indlæggelser og forbedre livskvaliteten for borgeren. Da TSA'erne tidligere har givet udtryk for et behov for at forstå algoritmens virkemåde, for at kunne anvende den, vurderes det nødvendigt at imødekomme dette behov. For at imødekomme dette behov, vurderes det relevant at undersøge XAI forklaringer. Forklaringer skal medvirke til at brugerne bedre kan forstå, have tillid til og effektivt styre AI-systemer. Dog vil effektiviteten af XAI og anvendeligheden af outputtet betinges af, hvordan slutbrugeren forstår forklaringen. Derfor vil dette speciale fokusere på designudfordringer ved XAI, der omhandler præsentation af XAI-forklaringer på en forståelig og intuitiv måde. For at kunne gøre dette med henblik på en high Fidelity prototype, der er tilpasset brugerne, for at finde en forklaring der kan øge algoritmens anvendelighed.

I dette speciale er formålet at undersøge, hvordan XUI kan muliggøre opfyldelsen af personalets behov for at forstå algoritmens resultater. Dette fører til følgende problemformulering for specialet:

“Hvordan kan en brugercentreret explainable user interface for telemonitorering medvirke til at øge anvendeligheden af prædiktiv algoritme?”

4. Præsentation af case

I samarbejde med TeleCare Nord tager dette specialet afsæt i udfordringer ved anvendelse af en prædiktiv algoritme til forudsigelse af exacerbationer. Under test af algoritmen i OTH, blev identificerede udfordringer ved anvendelsen, herunder manglende forklaring og forståelse af algoritmen. For at øge anvendeligheden for den prædiktive algoritme og dens output, er det nødvendigt at adressere og løse disse udfordringer. I forbindelse med skalering af telemedicin til nationalt brug for monitorering af borgere med KOL, udarbejdes en prototype i form af en XUI. Dette er specialets forsøg på at imødekomme slutbrugernes udfordringer ved anvendelsen.

5. Metode

I følgende afsnit præsenteres specialets metodiske design, herunder afsæt i en pragmatisk tilgang og user experience (UX), framework for udvikling og evaluering af specialets prototype.

5.1 Pragmatisme

Specialet bygger på en pragmatisk tilgang, hvor der lægges vægt på at anvende viden og teori til at løse problemstillinger på en praktisk måde, hvor løsningen skal være til gavn for mennesket og samfundet (63). For at kunne gøre dette, er det nødvendigt at interagere med den verden som udfordringen findes i (63). Dette skyldes, at mennesket i pragmatismen spiller en central rolle i fortolkningen af, hvad der vil være en passende løsning (63). Deres perspektiv sikrer, at løsningerne er meningsfulde (63). Løsningsforslagene skal evalueres og testes for at vurdere deres faktiske effekt og konsekvens i relation til den kontekst, hvor problemet foreligger (63).

Pragmatismen anses som en passende videns teoretisk tilgang i dette speciale, da projektet har til formål at løse problemet med manglende forståelse for en algoritme på en måde, der er nyttig for brugerne. Dette opnås ved at inddrage brugerne som primære grundlag for udviklingen af løsningen og samtidig basere sig på støttende evidens fra litteraturen om udvikling af brugercentrerede XUI.

5.2 User-Centered Design (UCD)

Prototypen tager inspiration fra Telma, og evidensen, der understøtter behovet for at kunne fremsøge information ved behov, hvilket er en del af en human-computer interaction (HCI) (64,65). Dette begreb fokuserer på design, evaluering og implementering af interaktive computerbaserede systemer, som ligeledes har henblik på at forbedre brugeroplevelsen (64,65). For at kunne udvikle effektive forklaringer er det vigtigt at inddrage brugerne under udviklingen, idet de kan have forskellige behov for forklaring (54). Dette gøres ved hjælp af user centered design (UCD), som sikrer større fokus på brugernes behov (66). UCD-metoder anses værende særlig anvendelige for at støtte en positiv brugeroplevelse ved at fremhæve forståelse for brugernes behov og forventning (67). Her giver UX-evaluering mulighed for at få brugerens samlede oplevelse på et interaktivt brugercentreret design og tager højde for brugeren gennem hele udviklingsprocessen (68). Dette gøres med udgangspunkt i Frank Gous kernelementer for UX, som er bestående af (69).

- Usability: Er produktet nemt at bruge?
 - Usability handler om hvor nemt brugeren kan udføre deres arbejdsopgaver under interaktion med produktet.
- Value: Er produktet nyttigt?
 - Value handler om, hvorvidt produktet, herunder dens funktionalitet og egenskaber, er af nytte for brugeren i deres liv eller arbejde.
- Adaptability: Er produktet nemt for brugeren at begynde at anvende?
 - Handler om prototypen, imødekommer brugernes forskellige behov på trods af individuelle præferencer. Denne del kan give en indikation på, hvorvidt brugerne vil blive ved med at anvende elementet frem for alternativet.
- Desirability: Er produktet anvendeligt og bedre end alternativet?
 - Handler om at skabe oplevelser, der går ud over at opfylde behov. Her fokuseres der på følelser og tilfredshed hos brugerne.

Kerneelementerne blev anvendt ved udarbejdelsen af spørgsmålene anvendt ved de semistrukturerede interviews for evaluering af prototypen.

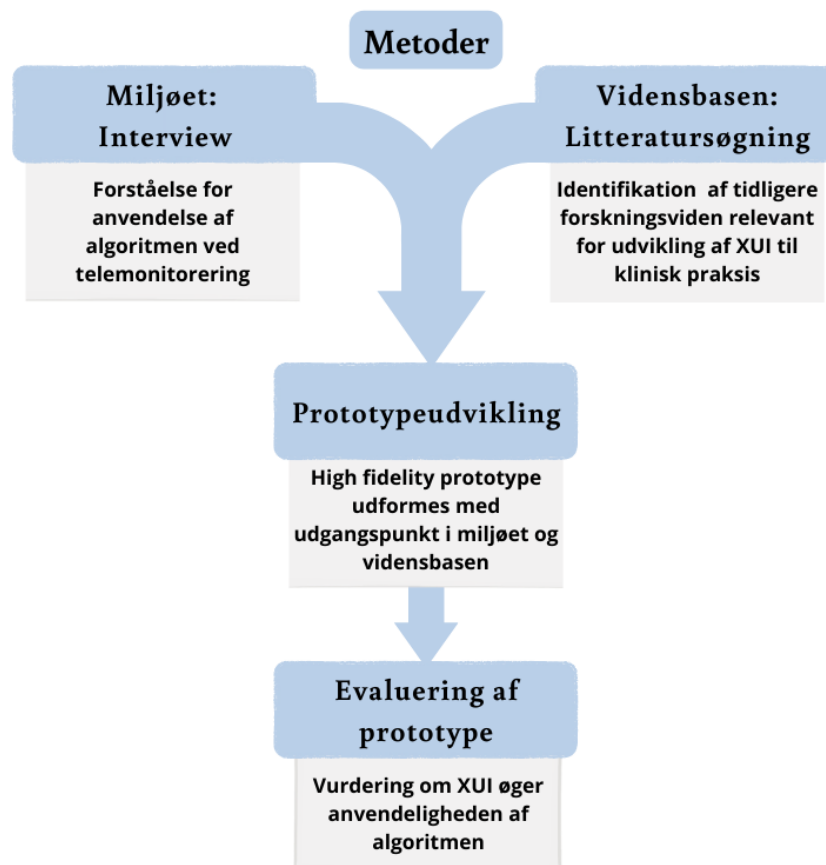
5.3 Specialets design for udvikling af en XUI

Udviklingen af en XUI til telemonitorering af borgere med KOL tog udgangspunkt i Hevners design science research (Hevners framework) som teoretisk ramme.

Denne omhandler forståelse, eksekvering og evaluering af forskning i informationssystemer, for at løse organisatoriske problemstillinger (70). Centralt for

Hevners framework er at informationssystemer agerer i et samspil mellem mennesker, organisationer og teknologier. Derfor er det vigtigt at tage hensyn til det miljø hvor produktet skal anvendes i, samt tidligere forskning relevant for produktudviklingen og evalueringen. Hevners framework består derfor af tre elementer: Miljøet, vidensbasen og informationssystemets forskning. Udviklingen og evalueringen af prototypen er det, der udgør elementet i informationssystemets forskning. I dette speciale udvikles et designforslag til en XUI, i form af en high fidelity prototype, der kan understøtte anvendelsen af output fra en prædiktiv algoritme for exacerbation.

Afdækningen af miljøet og vidensbasen blev udført parallelt. Resultaterne fra vidensbasen og miljøet blev sammenholdt for at identificere centrale krav til prototypen. Oversigt over metodernes rækkefølge og deres bidrag beskrives i nedenstående figur 1.



Figur 1: Specialets anvendte metoder og deres bidrag med inspiration fra Hevners framework. De blå bokse indikerer metoden, der blev anvendt, hvor de grå bokse beskriver deres bidrag (70)

5.3.1 Vidensbasen (Litteratursøgning)

Vidensbasen sikrer stringens ved udvikling og evaluering af prototypen. Den består af fundamenter og metoder fra tidligere forskning, der skal anvendes ved evaluering og

udvikling. Derfor blev der udført en systematisk litteratursøgning for at opnå viden om forskning i XUI.

5.3.2 Miljøet (Interview)

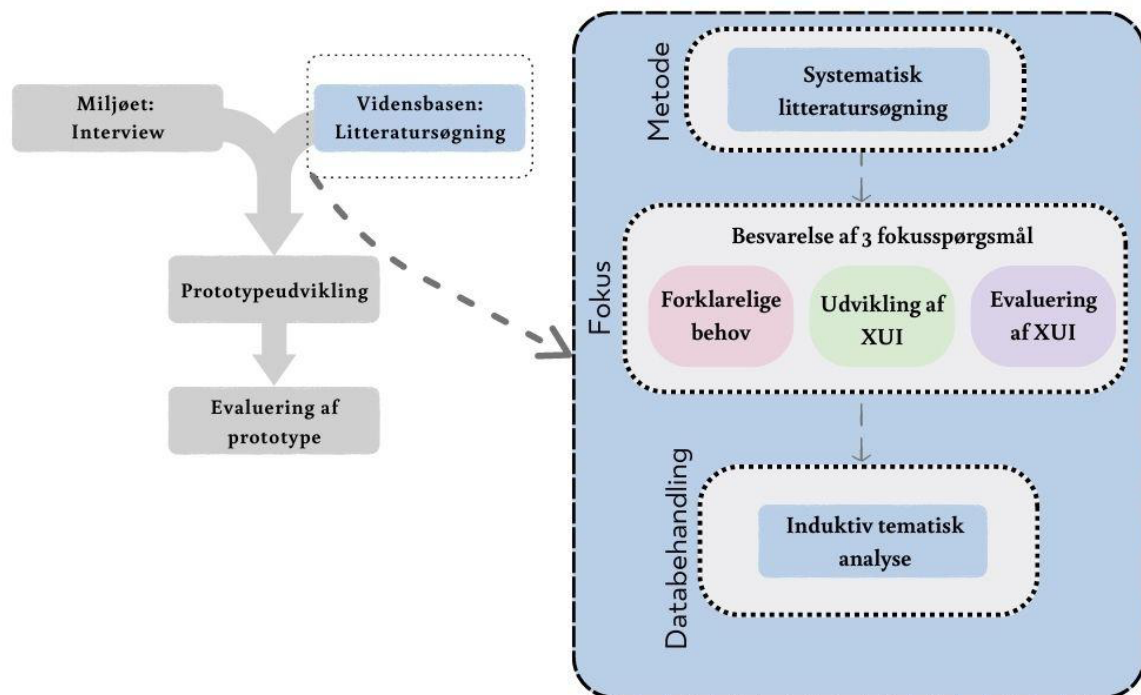
Miljøet skal sikre, at prototypen er til nytte for brugerne, herunder identificering af brugerbehov og krav, der bør tages forbehold for ved udvikling af en prototype. Derfor blev der afholdt et interview med TSA'er, som havde kendskab til algoritmen fra testperioden i OTH. Da specialets forfattere ikke tidligere har stiftet bekendtskab med hverken OTH eller Telma, blev der forud interviewet foretaget en gennemgang af platformene. Dette skulle give specialets forfattere et visuelt overblik over platformenes design, som der kunne drages nytte af under interviewet.

5.3.3 Udvikling og evaluering af prototype

Hevners framework tager udgangspunkt i en iterativ tilgang til udvikling og evaluering, hvor processen fortsætter, indtil prototypen opfylder sit formål. På baggrund af specialets korte tidsramme udføres kun en iteration af udvikling og evaluering af prototypen. Hevners framework beskriver, at det er problemet, som prototypen forsøger at løse, som er afgørende for evalueringsmetoden. Derfor vil der ved vidensbasen forsøges at identificere passende metoder til evaluering af en XUI, hvis formål er at øge brugerens forståelse for algoritmens output.

5.4 Vidensbase (Litteratursøgning)

For at afdække vidensbasen blev der udført en systematisk litteratursøgning, som tog afsæt i tre fokusspørgsmål. Derefter blev der udført en induktiv tematisk analyse af resultaterne. Denne proces visualiseres i figur 2.



Figur 2: De metodiske tiltag for afdækning af vidensbasen.

Formålet med de tre fokusspørgsmål for litteratursøgningen var at indsamle viden om relevante fundamenter og metoder til udvikling og evaluering (tabel 1). Det første fokusspørgsmål var fokuseret på at erhverve yderligere viden om sundhedsprofessionelle generelle behov, som blev sammenholdt med miljøets behov. Det andet fokusspørgsmål fokuserede sig på at tilegne viden om udviklingen af en XUI, og det sidste fokusspørgsmål havde fokus på evalueringstilgange for XUI. Præsentation samt uddybende beskrivelser for de respektive fokusspørgsmål kan ses i tabel 1.

Litteratursøgningens fokusspørgsmål			
	Fokusspørgsmål 1	Fokusspørgsmål 2	Fokusspørgsmål 3
Spørgsmål	Hvad er brugernes behov ved anvendelse af XUI i sundhedsvæsenet?	Hvordan kan en XUI udvikles til klinisk brug af sundhedsprofessionelle?	Hvordan kan en XUI evalueres?
Uddybende fokus	Artikler indeholdende beskrivelser om sundhedsprofessionelle krav for anvendelse af XUI. Artikel indeholdende forfatterens vurdering af sundhedsprofessionelles behov ved XUI.	Artikler indeholdende beskrivelser om metoder og design til udvikling af en XUI for sundhedsvæsenet.	Artikler indeholdende metoder, retningslinjer eller rammer for evaluering af en XUI, herunder specifikke værktøjer, som kan evaluere forklaringen og forståelse for en XUI.

Tabel 1: Videnbasens tre fokusområder og deres uddybende fokus.

5.4.1 Valg af databaser

Den systematiske litteratursøgning blev udført på databaserne PubMed, Embase, Cinahl, Scopus og Web of Science.

Cinahl, PubMed og Embase blev valgt grundet deres sundhedsfaglige, biomedicinske, samfundsvidenskabelige fokus, som var relevante for udviklingen af XUI til sundhedspersonale (71–73). Scopus blev vurderet relevant grundet dens brede fokus på både sundhedsområdet, humaniora, teknologi m.m. (74). Derudover indeholder Scopus flest publikationer omhandlende XAI sammenlignet med SpringerLink, IEEE Xplore og ACM (75). Web of Science blev valgt på baggrund af databasens fokus på videnskab, humaniora og samfundsvidenskab (76). Databasen indeholder et bredt sortiment af use-case, som ligeledes er relevante for specialets fokus (77).

5.4.2 Initierende søgeprofil

Der udarbejdede en søgeprofil bestående af to blokke "XAI" og "Interface" (Tabel 2). Denne skulle sikre en systematisk tilgang til litteratursøgningen. De anvendte søgeord blev

identificeret under udformningen af problemanalysen samt gennem ustrukturerede-, og kædesøgninger. Yderligere blev der udført strukturerede søgninger på google scholar til identifikation af eksisterende systematiske review, for at afdække hyppigt anvendte synonymmer og termer.

Ved den systematiske litteratursøgning blev databasernes individuelle thesaurus gennemgået i overensstemmelse med søgeprofilen. Dette blev gjort for at sikre inddragelse af relevante emneord for de respektive databaser. Hvis et fritekst søgeord allerede eksisterede som en indekseret term, blev det anvendt i forlængelse med fritekstordene. Både trunkering og anførselstegn blev anvendt ved fritekstordene, som tillod udvidelse ved at inkludere forskellige variationer af ordet, mens anførselstegn blev inddraget til frase søgning, for at præcisere søgeordet og fjerne unødvendig støj (78). Den booleske operator "OR" brugt for at udvide søgningsfeltet, mens den booleske operator "AND" blev brugt mellem søgeblokkene for at indsnævre søgning (78).

Søgeprofil			
		Blok 1: Explainable Artificial Intelligence	Blok 2: Interface
Fritekst søgeord		AND →	
Gældende for alle databaserne	OR ↓	"Explainable Artificial Intelligence" "Explainable AI" XAI "Understandable AI" "Interpretab* AI" "Trustworthy AI" "Explainable machine learning"	"Explana* interface" XUI "Explaina* user interface" "Human-machine interface" HMI "User interaction" "User interface" "Computer interface" "User computer interface" UI
Kontrollerede emneord			
Embase		'Explainable artificial intelligence' exp	'Human-machine interface' exp 'Computer interface' exp

	OR ↓		'User computer interface' exp
Pubmed og Cinahl		Ingen	"User-Computer Interface"[Mesh]
Scopus og Web of Science		Ingen	Ingen

Tabel 2: Søgeprofil for den systematiske litteratursøgning.

5.4.3 In- og eksklusionskriterier

Inklusions- og eksklusionskriterierne for søgningen blev udarbejdet, med udgangspunkt i alle tre fokuspørgsmål (Tabel 1). Artikler blev ekskluderet, hvis de ikke: 1) havde et fokus på XUI med forbehold for sundhedsvæsenet eller sundhedsprofessionelle som slutbruger, 2) centraliserede sig om metoder for biologiske signaler eller andre former for AI-modeller end prædiktive, eller 3) ikke passede ind i et af specialets fokuspørgsmål. Ved de forskellige databaser blev søgningen begrænset til kun at indeholde publikationer på dansk eller engelsk, der var udgivet efter 2015. Dette skyldes et øget fokus på brugercentret XUI siden 2015, hvilket betyder, at der ikke foreligger meget forskning om XUI før denne tid (75,79,80). Alle hitsene fra databaserne blev importeret til Reyran for at fremskynde screeningsprocessen på titel- og abstract-niveau samt identificering af dubletter (69). Alle dubletter blev gennemgået og håndteret af specialets forfattere under screeningen. Artiklerne blev screenet af specialeforfatterne individuelt for at minimere påvirkelse af hinanden. Eventuelle konflikter mellem specialets forfattere blev løst mundtligt. Dette blev gjort for at øge reliabiliteten ved inkludering af artikler. Screeningsprocessen tog udgangspunkt i kriterierne vist i tabel 3.

Inklusionskriterier	Eksklusionskriterier
• Publikations år: 2016 - 2023 • Skrevet på dansk eller engelsk • Adgang til fuldtekst • Beskriver XUI i sundhedsvæsenet • Bidrager til besvarelse af minimum et af de tre fokusspørgsmål	• XUI tager ikke forbehold for slutbrugerne • XAI-metoder der ikke er anvendelige til prædiktionsmodeller • XAI for biologiske signaler • Vurderet ikke gyldig, pålidelig eller anvendelig under kvalitetsvurdering • Konference resume, protokoller, bøger og magasiner

Tabel 3: Krav for ekskludering eller inkludering af artikler.

5.4.5 Kvalitetsvurdering af artikler

Alle artikler, som blev inkluderet efter fuld tekstlæsning, blev kvalitetsvurderet i henhold til tjeklister fra Let Evidence Guide Every New Decision (LEGEND) (82). Kvalitetsvurderingen blev udført individuelt af specialets forfattere, og efterfølgende diskuteret i plenum. For at vurdere kvaliteten var der et krav om enighed hos begge speciale forfattere, og ved uenighed om artiklernes kvalitet blev dette løst mundtligt. Artikler, der blev evalueret ikke gyldig/pålidelig, blev ekskluderet.

5.4.5 Ekstrahering af data fra artiklerne

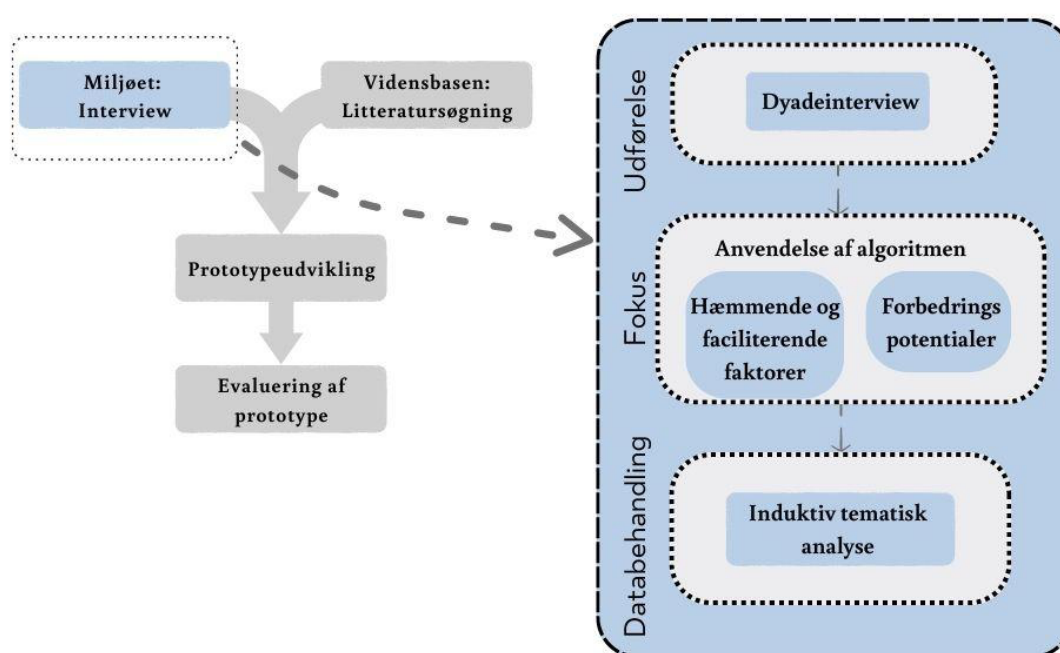
Artiklernes indhold blev først kategoriseret ift. de tre fokusspørgsmål, hvorefter der blev udført en induktiv analyse af udtræk fra artiklerne. Analysen blev udført i tre faser, som delvis overlappede hinanden.

I fase et blev der foretaget en frie linje-for-linje-kodning af artiklernes relevante indhold for besvarelse af de tre fokusspørgsmål (83). I fase to blev koder med tilnærmelsesvis identisk betydning organiseret, for at konstruere »beskrivende« temaer (83). I fase tre blev der udformet »analytiske« temaer (83). Specialets forfattere udførte fase et

uafhængigt af hinanden, hvor fase to og tre indledningsvis blev udført af specialets ene forfatter. Efterfølgende blev denne proces fremlagt for specialets anden forfatter, hvor uenigheder om de udvalgte temaer blev løst mundtligt. Den tematiske analyse kan ses i bilag 1.

5.5 Miljøet (Interview)

For at afdække miljøet blev der udført et semistruktureret dyadeinterview, med fokus på hæmmende eller faciliterende faktorer og forbedringspotentialer for anvendelse af den prædiktive algoritme. Ved databehandlingen blev der anvendt en induktiv tematisk analyse af resultaterne. Denne proces visualiseres i figur 3.



Figur 3: Den metodiske tilgang til afdækning af miljøet.

Dyadeinterview blev valgt, fordi den gav muligheden for at undersøge TSA'erne holdninger og meninger om anvendelse af en algoritme, samt evaluere ligheder eller afvigelser i deres udsagn (84). Ydermere gav metoden specialeforfatterne mulighed for at stille spørgsmål om bestemte områder, hvor informanterne selv kunne bringe relevante emner i spil. Til udarbejdelse og analysering af interviewets indhold blev der taget udgangspunkt i Kvale og Brinkmanns interviewforløb bestående af syv faser (85). Rammen bidrog til at sikre en systematisk tilgang ved udarbejdelsen af interviewet. Uddybende forklaring og proces for interviewets udformning i henhold til Kvale og Brinkmanns syv faser henvises til bilag 2.

5.5.1 Rekruttering

I samråd med projektlederen fra TeleCare Nord, blev der valgt at foretage en formålsbestemt udvælgelse af informanter. Gennem dette blev der rekrutteret to TSA'er, som havde stiftet erfaring med den prædiktive algoritme under testperioden i OTH.

5.5.2 Interviewguide

Inden gennemførelsen af det semistrukturerede interview blev der udarbejdet en interviewguide, som fungerede som en retningslinje for interviewet. Interviewguiden indeholdt praktisk information til informanterne og de specifikke spørgsmål som skulle (Bilag 3). Derudover blev der udarbejdet deltagerinformation (Bilag 3.1) og en samtykkeerklæring (Bilag 3.2), som blev inkluderet i interviewguidens briefing-del.

Ved udformning af spørgsmål til interview, blev der taget udgangspunkt i to fokusspørgsmål, som kan ses i tabel 4. Begge fokusspørgsmål bidrog til en dybere forståelse for anvendelse af den prædiktive algoritme ved telemonitorering.

Fokus	Fokusspørgsmål
Evaluerings af TSA-anvendelse af algoritmen i klinisk praksis	Hvilke barrierer og faciliterende faktorer er der for anvendelse af algoritmens prædiktions?
Information om behov til interface for brug af algoritmens prædiktions	Hvilke forbedringspotentialer i interfacen er vigtige for øget anvendelse for algoritmens prædiktions?

Tabel 4: Fokus for interviewet med tilhørende spørgsmål, som ønskes besvaret.

5.5.3 Udførelse af interview

Interviewet blev afholdt online, hvor begge informanterne blev interviewet samtidig, så informanterne havde mulighed for at supplere hinanden. Under interviewet var begge af specialets forfattere til stede, hvor den ene påtog rollen som facilitator, mens den anden påtog rollen som observatør. Formålet med observatøren var at notere uddybende spørgsmål ned undervejs, som blev stillet inden interviewets afslutning.

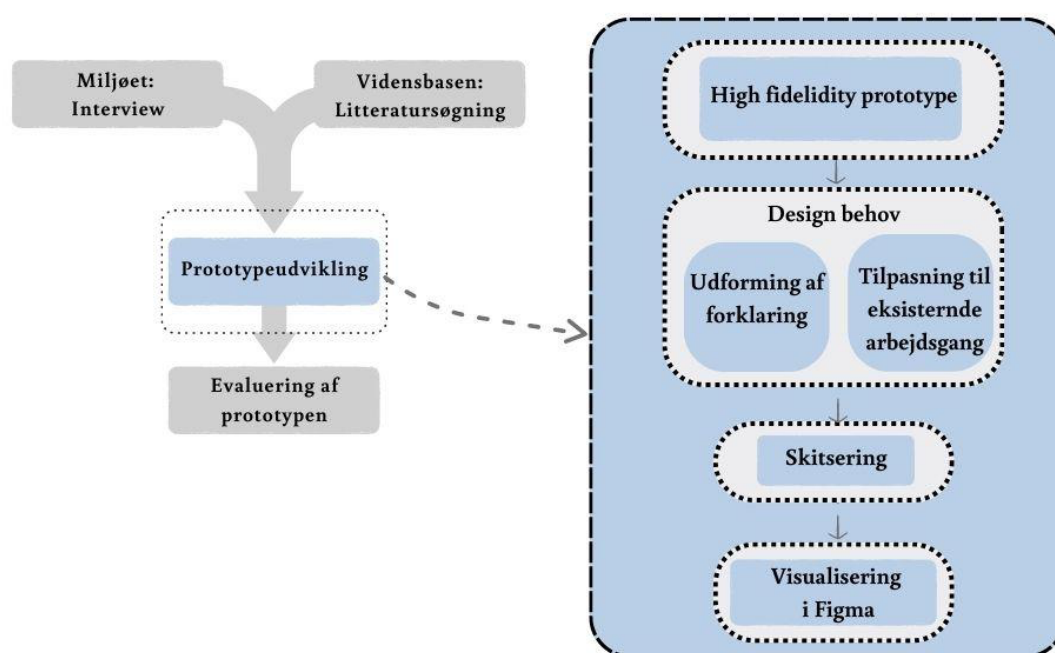
For at sikre, at spørgsmålene i interviewguiden var forståelige, blev interviewet pilottestet på to studerende fra uddannelsen Klinisk Videnskab og Teknologi. Efterfølgende blev der givet mundtlig feedback på forståelsen, rækkefølgen og forbedringsforslag. Ved revurdering af interviewets spørgsmål med den mundtlige feedback, blev der taget hensyn til de studerendes erfaring til konteksten og anvendte termer, hvor interviewguiden blev tilpasset herefter.

5.5.4 Ekstrahering af data fra dyadeinterview

Første del af databehandling for interviewet bestod af at transskribere lydoptagelsen (Bilag 4). Derefter blev der udført en induktiv tematisk analyse, som tog udgangspunkt i seks trin bestående af: 1) Familiarisering med data, 2) Generering af initiale koder, 3) Søgning efter temaer, 4) Gennemgå temaer, 5) Definering og navngivning af temaer, 6) Udarbejdelse af rapport (86)(Bilag 5). Databehandling af interview blev udført af observatøren, hvorefter facilitatoren vurderede indholdet ift. egen forståelse. Dette medvirkede til enighed mellem specialets forfattere i henhold til de valgte temaer og tilhørende udtalelser.

5.6 Prototypeudvikling

I dette afsnit beskrives tilgangen for udviklingen af specialets prototype, hvis formål var at øge TSA'ernes forståelse for algoritmen. Denne proces udformede sig i tre faser: 1) Identifikation af design behov, 2) Skitsering af designforslag og 3) Visualisering af designforslaget (figur 4).



Figur 4: Oversigt over tilgang for udvikling af prototypen

5.6.1 Design behov

Som udgangspunkt for udviklingen af prototypen, blev der udformet krav til prototypen ved sammenfatning af vidensbasen og miljøet. Dette skulle sikre, at der under udviklingen blev taget hensyn til brugernes behov og viden fra tidligere forskning. Denne sammenfatning kan ses i afsnit "6.3 Sammenfatning af krav til prototypen". For at

undersøge, hvordan kravene kunne passe ind i TSA'ernes arbejdsplatform, blev der i samråd med projektlederen for TeleCare Nord taget billeder af OTH og Telma (Bilag 6).

5.6.2 Skitsering

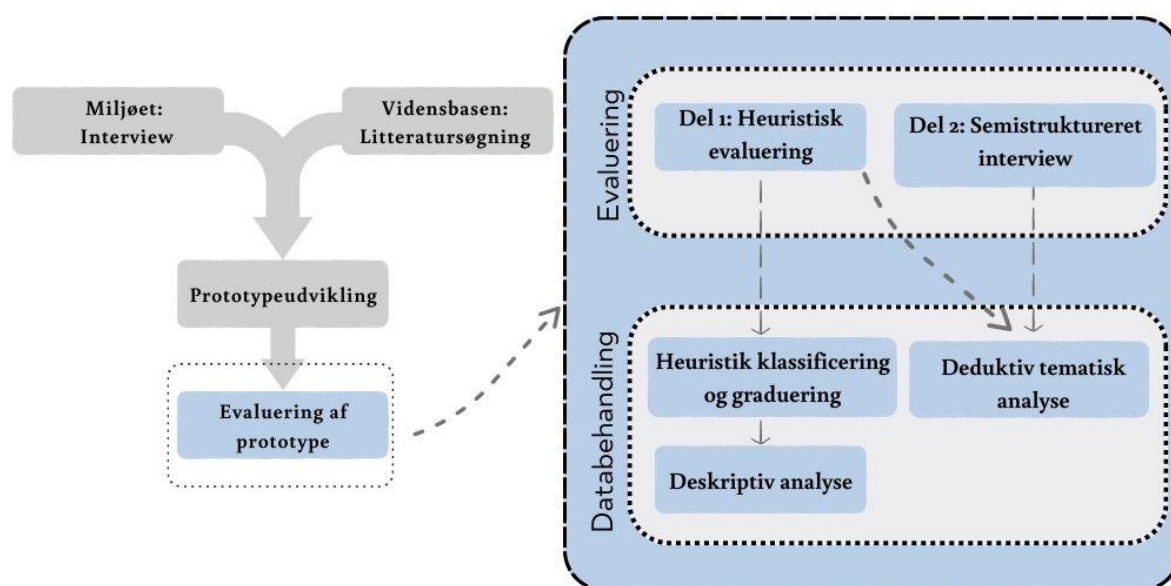
Der blev afholdt fælles brainstorming for at undersøge og diskutere, hvordan kravene til prototypen kunne imødekommes. Hovedfokus lå på at præsentere algoritmens output og tilhørende forklaringer i Telma. Ved skitseringen af hvordan kravene kunne imødekommes, blev der taget inspiration fra vidensbasen. Skitserne blev løbende evalueret og forbedret under skitseringen.

5.6.3 Visualisering

Ved udviklingen af high fidelity prototype blev Figma anvendt, som er en gratis open-sourcehjemmeside for UI-design. Dette gjorde det muligt for forfatterne at arbejde parallelt med hinanden online under visualisering af prototypen (87). Ved udformningen af prototypen interaktive komponenter blev der fokuseret på at gøre dele nogle elementer interaktive, som havde relevans for algoritmen og dens forklaringer. Ved tilføjelse eller fjerne af elementer, blev prototypen løbende evalueret i henhold til imødekommelse af kravene.

5.7 Evaluering af prototype

Ud fra specialets prototypen for en XUI, blev der udført en evaluering som skulle bidrage til at give en pejling om, hvorvidt der var opnået forståelse for exacerbationsrisikoen ud fra de identificerede krav. Evalueringen blev udført i to dele, som illustreret i figur 5



Figur 5: Præsentation af evalueringens elementer og databehandling,

5.7.1 Rekruttering

I samråd med projektlederen fra TeleCare Nord blev der udført en formålsbestemt udvælgelse af informanter, hvor der blev rekruttering af deltagere. En slutbruger med daglig arbejdsgang i Telma blev rekrutteret. Til at supplere slutbrugerens evaluering, blev det vurderet relevant at få et professionelt og objektivt syn på prototypen. Derfor blev der rekrutteret en UX-ekspert blandt speciale forfatterens netværk. Dette blev gjort, idet det vurderes, at en UX-ekspert kunne forholde sig til prototypens usability og potentielle forbedringsforslag.

5.7.2 Udarbejdelse af evalueringen

Forud evalueringen blev der udarbejdet en evalueringsprotokol, som skulle støtte facilitatoren under evalueringen (Bilag 7). I samme forbindelse blev der udformet en deltagerinformation (Bilag 7.1) og samtykkeerklæring (Bilag 7.2). Ved udformning af evalueringsprotokollen blev der fokuseret på besvarelse af problemformuleringen, hvilket medvirkede til en todelt evaluering. Evalueringens to dele bestod af en heuristisk evaluering og opfølgende semistruktureret interview.

Heuristiske evaluering

Den heuristiske evaluering skulle medvirke til opdage eventuelle udfordringer ved evaluatorernes interaktion med prototypen, som potentielt kunne udgøre en barriere for anvendelse i klinisk praksis (88). Evalueringen blev baseret på fiktive borgere med realistiske målinger, og der blev udformet en række opgaver, der skulle stimulere en potentiel arbejdsgang for anvendelse af prototypen. Alt blev der udarbejdet 16 opgaver, hvor evaluatorerne fik besked på at tænke højt mens de løste opgaverne (Bilag 7.3). Undervejs spurgte facilitatoren ind til deres tanker om placering, forståelse og anvendelse af forklaringerne, for at understøtte deres think-aloud. Opgaverne blev struktureret i henhold til de opstillede krav, som prototypen blev udviklet ud fra. Formålet var at afgøre, om specialets løsninger var nok til, at evaluatorerne mente at de kunne anvende og forstå exacerbationsrisikoen. Idet UX-eksperter ikke kendte til konteksten for prototypens anvendelse, blev der udformet en indledende beskrivelse af casen kun til UX-eksperter, herunder hvem slutbrugeren var og deres arbejdsopgave (Bilag 8).

Semistruktureret interview

Det semistruktureret interview bidrog til en dybere forståelse for evaluatorernes vurdering af prototype gennem oplevelse og holdning (89). For at kunne afdække de væsentlige aspekter, blev der taget udgangspunkt i kerneelementerne for UX: usability,

desirability, adaptability og value (90). Dette førte til 15 spørgsmål om samlede brugeroplevelse, potentielle forbedringsområder og afklarings spørgsmål fra den heuristiske evaluering (Bilag 7.4). Otte af de 15 spørgsmål var forbeholdt begge evaluatore, hvor fem var rettet mod slutbrugeren og to var specifikke for UX-eksperter. Formålet med specifikke spørgsmål var at drage nytte af evaluatorernes forskellige perspektiver i evalueringen. Spørgsmålene til slutbrugeren omhandlede anvendeligheden af prototypen i forbindelse med telemonitorering, mens spørgsmål til UX-eksperter fokuserede på brugervenligheden og potentielle barrierer for anvendelse.

5.7.3 Pilottest af evalueringen

For at undersøge evalueringsmetodens gennemførlighed, varighed og uforudsete hændelser blev der udført pilottest. Pilottesten blev udført med en medstuderende fra Klinisk Videnskab og Teknologi. Kravene for deltagerne var, at de havde sygeplejefaglig baggrund, og tidligere erfaring med UI-udvikling. Kriterierne blev valgt, for at understøtte spørgsmål og case henvendt til slutbrugeren og UX-eksperter. Efter pilottesten modtog specialets forfattere mundtligt feedback, hvor der ved revurdering af evalueringsprotokollen med den mundtlige feedback, blev der taget forbehold deltagerens erfaring til konteksten, hvor evalueringsprotokollen blev tilpasset herefter.

5.7.4 Databehandling af evaluering

For at kunne ekstrahere data blev evalueringerne transskriberet (Bilag 9). Evaluatorernes udtalelser blev nøje nedskrevet med undtagelse for stammen, gentagelser eller fyldord. Ud fra transskriberingerne blev der udtalelser om faciliterende og hæmmende elementer ved prototypen med fokus på, hvorvidt der var opnået forståelse for exacerbationsrisikoen udtrykt. For at præcisere hvilke elementer af prototypen der udgjorde en udfordring, blev der anvendt heuristikker og graduering af heuristikkerne. En heuristik refererer til metoder eller strategier, som oftest leder til problemløsning (91), hvor dette speciale har taget udgangspunkt i 15 usability principper fra Muller et al (Bilag 10). Derudover er heuristikkerne blevet vurderet i henhold til alvorligheden, hvor dette omtales som graduering af heuristiske overtrædelser (92). Specialet anvender en gradueringsskala fra Nielsen (92), hvor gradueringsniveauerne og deres betydning for usability, kan ses i tabel 5. Analysen af heuristikkerne og deres overtrædelser blev udført af en af specialeforfatterne og efterfølgende blev denne proces fremlagt for specialets anden forfatter, hvor uenigheder om de udvalgte heuristikker og gradueringer blev løst mundtligt. Analysen kan ses i bilag 11.

Graduering	Beskrivelse
0	Ikke et usability problem
1	Kosmetiske udfordringer. Behøver ikke løses, medmindre der er ekstra tid.
2	Mindre usability udfordring. Løsning af dette bør prioriteres lavt.
3	Større usability udfordring. Løsning af dette bør prioriteres højt.
4	Usability katastrofe. Nødvendigt at løse dette, før produktet kan anvendes.

Tabel 5: Oversigt over klassificeringerne af usability (92)

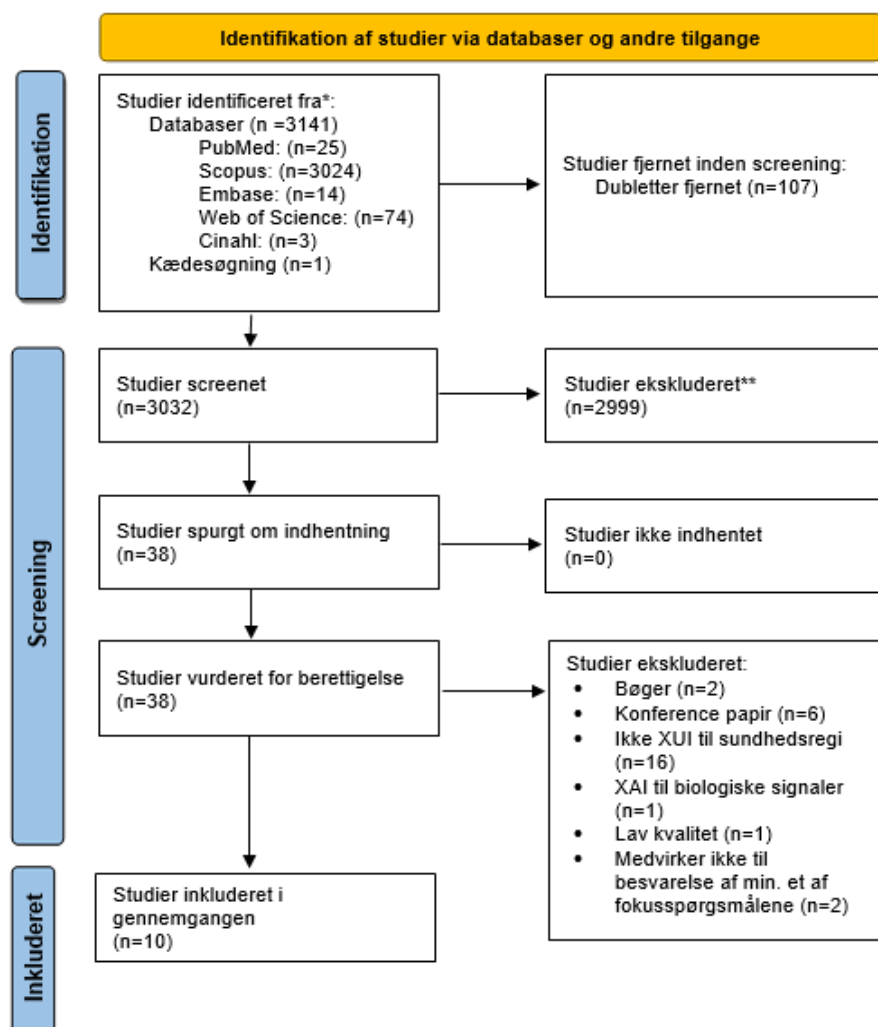
Efter vurdering af heuristik og deres overtrædelse blev der anvendt en deduktiv tilgang for at kunne tematisere evaluatorernes udtalelser (Bilag 12). Analysen bestod af en transskribering, hvor elementer fra evalueringens to dele blev udtrukket i overensstemmelse med kerneelementerne for UX: usability, value, adaptability og desirability. Under udtrækningen blev der noteret, hvilket del af evalueringen udtalelsen kommer fra og hvilken/hvilke krav udtalelsen berørte. Dette førte til lettere vurdering om evaluatorernes mening kom før eller efter opgaverne i relation til kravene, hvor dette kunne have en påvirkning på hvordan de f.eks. så på exacerbationsrisikoen. Efter databehandling blev temaerne og tilhørende udtalelser vurderet i plenum af specialets forfattere, for at sikre enighed og ekstrahering af relevante data.

6. Projektets resultater

I dette afsnit præsenteres specialets resultater, der indledningsvis starter med resultaterne fra vidensbasen efterfulgt af resultater fra miljø. Disse vil blive sammenholdt og danne grundlag for de krav, som vil være gældende for udvikling af prototypen. Derefter præsenteres prototypen og dens imødekommelse af kravene, hvorefter resultaterne fra evalueringen af prototypen.

6.1 Vidensbase: Resultater fra litteratursøgningen

Ud fra vidensbasens systematiske litteratursøgning blev i alt 3141 artikler identificeret, hvoraf 107 dubletter blev fjernet. Af de resterende 3032 artikler blev 2999 ekskluderet ved titel og abstract, da de fleste ikke var relateret til XUI og sundhedsvæsenet. Efter fuld tekstlæsning af 38 artikler blev 11 inkluderet artikler, hvoraf en af artiklerne blev ekskluderet på grund af lav kvalitet. En oversigt over screeningsprocessen kan ses i figur 6, og en oversigt over artiklernes bidrag i henhold til de tre fokusspørgsmål kan ses i tabel 6.



Figur 6: Flowdiagram over screeningsproces.

6.1.1 Præsentation af vidensbasens inkluderede artikler

Gennem den systematiske litteratursøgning blev 10 artikler inkluderet. I nedenstående tabel 6 præsenteres artiklerne og deres bidrag til et eller flere af fokusspørgsmålene.

Titel	Forfatter	Fokusspørgsmål		
		1	2	3
A Clinical Decision Support System for the Prediction of Quality of Life in ALS	Antoniadi et al.	x	x	x
A review of interpretable ML in healthcare: Taxonomy, Applications, Challenges, and Future Directions	Abdullah et al.		x	
A visual approach to explainable computerized clinical decision support	Müller et al.	x	x	x
A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare.	Barda et al.	x	x	
Explainable AI for clinical and remote health applications: a survey on tabular and time series data	Di Martino & Delmastro	x	x	x
Question-Driven Design Process for Explainable AI User Experiences	Liao et al.	x	x	x
Explaining artificial intelligence with visual analytics in healthcare	Ooge et al.	x	x	
Extending the Nested Model for User-Centric XAI: A Design Study on GNN-based Drug Repurposing.	Wang et al.	x	x	

How the different explanation classes impact trust calibration: The case of clinical decision support systems	Naiseh et al.	x	x	x
Human-centered XAI: Developing design patterns for explanations of clinical decision support systems	Schoonderwoerd et al.	x	x	x

Tabel 6: Artiklernes bidrag til besvarelse af et eller flere af fokusspørgsmålene.

6.1.1 Fokusspørgsmål 1: Behov for anvendelse af XUI i sundhedsvæsenet

Ved dataekstraktion for dette fokusspørgsmål, blev der identificeret to emner: brugernes behov for information og udvikling af en XUI til den rette kontekst. I det følgende afsnit præsenteres artiklernes bidrag i relation dette fokusspørgsmål

Brugernes behov for information

Ved anvendelse af en XUI i klinisk praksis, blev det vurderet, at tilgængeligheden af relevant information om patienten var en vigtig faktor. Ni ud af de ti inkluderede artikler havde fokus på, at brugerne kunne få adgang til handlingsrettet information for at kunne evaluere og vurdere patienten (59,60,93–99). Relevant information blev beskrevet som værende information, der både kunne understøtte og modsige prædiktionen (60). Det blev ligeledes påpeget, at det var vigtigt at vide, om der var sket en ændring i risikoen, og om stigningen var gradvis eller pludselig (60%). Desuden var der et ønske om at kende de værdier og informationer, der blev brugt i algoritmens beregning, og hvilke indflydelser de havde på forudsigelsen (60,93–95,98,99). Dog blev det påpeget af Barda et al. at behovet for forklaring foreligger forskelligt hos mennesker (60). Klinikere ønskede mere dynamisk information, som kunne hjælpe dem med at prioritere patienter, vurdere patientens status og hurtigere identificere patienter med høj risiko (60). Modsat sygeplejersker som foretrak mere simple og statiske informationer, som omhandlede forskellige anvendelser af prædiktionen (60). Dette kan yderligere understøttes af Naiseh et al., som fandt at forskellige sundhedsprofessioner ønskede at få præsenteret informationerne på en måde, der var tæt knyttet til den respektive opgave som forudsigelsen var forbundet med (98). På baggrund af ovenstående vurderes det, at forklaringen bør tilpasses til den specifikke kontekst, brugere og arbejdsopgaver som algoritmen skal understøtte.

Udvikling af XUI til rette kontekst

Udviklingen af en XUI bør tage udgangspunkt i den specifikke kliniske kontekst (60,93,94,98). En af bekymringerne blandt sundhedsprofessionelle ved brug af en XUI, var dens usability, herunder tid og opmærksomhed, de skulle bruge på at få en korrekt fortolkning og brug af AI-modellen (60,93,94,98). Antoniadi et al. understreger også, at sundhedsprofessionelle har et øget behov for at præsentere argumenterne for AI-modellens output i en form, der kan sammenlignes med deres normale beslutningsproces (94). Ooge et al., beskriver, at tilpasning af interaktioner til slutbrugerens beslutningsprocesser kan forbedre brugernes accept og tillid til systemet (97). Liao et al., fandt dog at, sundhedsprofessionelle ikke kun bruger en XUI til at verificere outputtet (59). De brugte også XUI'en til at estimere, hvor lang tid der skulle bruges til at vurdere patienterne, opnå bedre forståelse for patientinformationen, validere deres egen beslutning og lære fra modellen m.m. (100). Derfor vurderes det, at sundhedsprofessionelle finder flere måder at udnytte en XUI på udover dens oprindelige formål. På baggrund af disse fund, bør der derfor tages forbehold for brugernes forskellige målsætninger, og forsøge at imødekomme denne brede anvendelse, som litteraturen indikerer.

6.1.2 Fokusspørgsmål 2: Udvikling af en brugervenlig visualisering af algoritmens prædiktions til klinisk brug

Til besvarelse af fokusspørgsmål to blev der identificeret tre emner; XAI-metoder, visualiseringsteknikker, brugercentrering af XUI. I nedenstående afsnit præsenteres artiklernes bidrag i relation til dette fokusspørgsmål.

XAI-metoder

XAI-metoder, herunder global, lokal, kontrafaktisk og eksempel-baseret, er nogle af metoderne, der anvendes for at øge brugernes forståelse af AI-modellens output (59,95,96,101). Blandt disse fire forklaringstyper beskrives lokal og global forklaring som essentielle for at forstå en forudsigelse og vurdere dens pålidelighed (94,95). Den globale forklaring er relevant for at give brugerne en generel forståelse for hvordan en AI-model virker, mens den lokale forklaring beskriver hvordan en model udformer en forudsigelse for en specifik kontekst (59,96,101). Begge typer forklaringer anses som vigtige for, at brugere kan vurdere outputtets robusthed og sikkerhed, som medvirker til at øge brugernes tillid til systemet (95).

Både Müller et al. og Schoonderwoerd et al. beskriver begge, hvordan brugernes tillid kan øges ved brugen af lokale forklaringer (95,99). Müller et al., fandt desuden, at global forklaring var relevant, men kun i opstartsfasen (95,99). Dog fandt Naiseh et al. ud af, at sundhedsprofessionelle generelt synes, at XAI-metoderne ikke var tilstrækkeligt tilpasset til en klinisk arbejdsgang (98). Brugere i studiet også udtrykte behov for

yderligere vejledning i, hvordan XAI-forklaringen, især den globale og lokale, skulle tolkes, der antyder at metoderne kan være svære at forstå (98). På baggrund af litteraturens resultater er der en indikation for, at slutbrugeren skal have mulighed for at vurdere outputtets robusthed og sikkerhed, hvor en global forklaring og en lokal forklaring i specialets formål vil være mere anvendelig.

Visualiseringsteknikker

For at få præsenteret algoritmens prædiktion, viser artiklernes fund, at visualisering vil være et effektivt værktøj (60,93,101). Hvor visualisering i form af en graf, der f.eks. viser indflydelsen af algoritmens input for forudsigelsen, kan øge fortolkning og accept af modellen (60,101). Vidensbasens resultater præsenterer forskellige metoder for visualisering i en XUI, herunder pie-chart, donut charts, line-chart, bar-chart/reverse (60,95,97,99,101). Ud af alle visualiseringsmetoderne blev bar-charts hyppigst anvendt ved visualisering af global, og den lokale forklaring oftest visualiseres ved en reversed-barchart, som synliggør argument for og imod (59,60,94,98,99,101). Mens der ved global forklaring vises vægtningen i form af et normalt bar-chart (59,60,94,98,99,101). Disse resultater viser, at der med fordele kan anvendes både en global og lokal forklaring i en XUI.

Brugercentrering af XUI

For at kunne designe en brugercentreret XUI henviser fundene fra artiklerne til interaktive visualiseringer som en grundlæggende tilgang til AI-forklaringer (60,93). Idet visualiseringen i sig selv ikke vil indeholde analyser eller konklusioner, vil det betyde, at fortolkning og analysering vil blive udført af brugeren selv. Barda et al. anvendte i sit studie en interaktiv graf for at visualisere algoritmens input i en trend-graf (60). Dette skyldes, at flere sygeplejersker i studiet fremhævede et behov for en oversigt over risiko udvikling, som vigtig information for at kunne vurdere den kliniske relevans af forudsigelsen (60). Det blev vurderet, at inkludering af en interaktiv trend graf kunne øge brugernes opfattede nytteværdi af en prædiktionsmodel (60). En interaktiv graf vil yderligere give brugerne kontrol over hvordan de analyserer data elementerne, og mulighed for at vælge enkelte datapunkter, eller udtrække flere patientdata i en trend graf (60). Anvendelse af en interaktiv graf, eventuelt i trend format, vurderes derfor at være relevant for at understøtte brugere forskellige analysetilgange.

6.1.3 Fokusspørgsmål 3: Standardiseret metode og værktøjer til evaluering af en XUI

For at kunne besvare af fokusspørgsmål tre blev der kun identificeret et emne; Evalueringsmetoder og værktøjer. I det følgende afsnit præsenteres artiklernes bidrag i relation til dette fokusspørgsmål.

Evalueringsmetode og værktøjer

For at evaluere en forklaring henviser Di Martino & Delmastro til at dette bør gøres fra en klinisk kontekst, da denne kan være afgørende for at sikre, at XUI'en følger domæneviden (96). Gennem fuld tekstlæsning af de inkluderede artiklerne for fokusspørgsmål 3, blev der ikke præsenteret en standardiseret tilgang til at evaluere forståelsen og forklaringen i en XUI. Artiklerne anvendte forskellige metoder for at kunne evaluere på en XUI, blandt andet anvendte Liao et al. en heuristisk evaluering for at kunne identificere brugernes barriere i interaktionen (59). Her brugte Naiseh et al. et semistruktureret interview, med spørgsmål baseret på forståelighed, pålidelighed og teknisk kompetence i forhold til forklaringer i XUI (98). Müller et al. nævner, hvordan en realistisk patient case vil bidrage til at minimere fejl og misforståelser i evalueringen (95). Et fællestræk for de inkluderede artiklerne til besvarelse af dette fokusspørgsmål var, at forfatterne selv valgte og udformede en evalueringsmetode, som de mente var mest hensigtsmæssig til at støtte studiet formål.

Alle artiklernes evalueringer involverede sundhedsprofessionelle til at vurdere en XUI i praksis, hvilket gør, at artiklernes anvendte metode og værktøj kunne anvendes som inspirationskilde. Her vil metoder anvendt til evaluering være direkte overførbare til specialets formål, men hvor spørgsmålene ikke vil kunne vurderes 100% anvendelige til dette speciale.

6.2 Miljøet: Resultater fra interview

Rekrutteringen og inddragelse af to erfarne TSA'er bidrog til at identificere og verificere både kendte og ukendte udfordringer ved anvendelse af en exacerbationsrisiko. I dette afsnit vil de væsentligste temaer og tilhørende fund fra interviewet blive præsenteret.

6.2.1 Præsentation af informanter

Begge informanter har været involveret i TeleCare Nord siden 2013 og været med til at implementere telemedicin til monitorering af borgere med KOL. Informant 1 er uddannet fysioterapeut, men har arbejdet med KOL-rehabilitering i mange år og tilegnet sig kompetencer herigennem. Informant 2 er uddannet sygeplejerske og har igennem hele

sin karriere arbejdet med lungemedicin. Præsentationen af informanterne kan ses i tabel 7.

Præsentation af informanter		
Informant	Informant 1	Informant 2
Klinisk baggrund	Fysioterapeut	Sygeplejerske
Erfaring med TeleCare Nord	10 år	10 år

Tabel 7: Oversigt over informanternes kliniske baggrund.

6.2.2 Vurdering af data

Begge informanter udtalte, at de ofte anvender farvekodet triagering, hvor de først vurderer borgere med en rød farvekode, derefter borgere med en gul alarmering og slutteligt de med en grøn alarmering. Farvekoderne på triagering tildeles af en algoritme med foruddefinerede eller patientspecifikke grænseværdier, som beskriver normalområdet for borgernes vitale parametre og symptombesvarelser (29). Det blev understreget, at alle borgere vil blive vurderet, uanset hvilken farvekode de har. Det er ikke kun ved en rød eller gul alarmering, at der bliver vurderet på borgerens data.

“Altså vi er jo inde omkring alle borgere, selvom det er en alarm der lyser grøn, så er vi også inde og vurdere den. Det kan godt være, at det ikke tager så lang tid, fordi alting ser så fint ud, men det er ikke kun fordi der er blevet varslet om exacerbation, vi går ind og kigger.” - Informant 2

For at få indblik i den enkelte borgeres tilstand, læser de først borgernes besvarelser fra spørgeskemaet, som omhandler borgerens symptomer og gener, for at få et indblik i borgerens tilstand. Dernæst ser de på borgerens vitale parametre, og borgerens historik.

“Og det jeg hørte i, eller det jeg hører du siger *. Det er, at I tager udgangspunkt i spørgsmålene først, og så kigger I på værdierne. Altså de vitale værdier bagefter, er det korrekt forstået?” - Interviewer

“Det er faktisk ja, det vægter mere, hvad de svarer på spørgsmålene end de målte værdier. Saturation kan jo sagtens være lav, hvis de har kolde fingre uden at de behøves at have noget påvirket lungebetændelse i luftvejene.” - Informant 2

For at få et overblik udtalte Informant 1, at vedkommende oftest brugte grafer til at få overblik over historikken for de vitale parametre, mens informant 2 brugte tabelformat.

Selvom hver informant havde deres egne præferencer til visualisering af borgerens historik, var begge enige om, at graferne var vigtige. Dette skyldes muligheden for at ændre alarmgrænserne, og dermed farvekoderne for de vitale parametre, hvilket kan gøre det svært at gennemskue om en stigning eller fald i tabelformat.

“For eksempel med vægt, er det os selv der ændrer vægtgrænserne, og der kunne det måske være en fordel at vi bruger grafen, fordi når vi ændrer ved grænsen, så ændrer vi også farvekoderne. Og der skal man være lidt skarp på, at vi nu snakker om væggtab eller vægtstigning. Fordi de bliver udlignet.” - Informant 1

Dette indikerer at brugerne er vant til at anvende grafer til at få overblik over de vitale parametre, men foretrækker at bruge tabelformat til at få et indledende indblik i borgernes velbefindende.

6.2.3 Anvendelighed af algoritme

I dialog om den prædiktive var begge informanter enige med hinanden om, at en af årsagerne til de ikke kunne anvende den, var at den alarmerede om exacerbation hos borgere, hvor der ikke var symptomer på exacerbation. Yderligere mente de at algoritmen ville være anvendelig, så længe den kun tog udgangspunkt i de vitale parametre, uden forbehold for borgerens symptombillede.

“For at jeg skal kunne bruge det (algoritmen), så bliver det nødt til at være de rigtige besvarelser det bliver taget ud fra. Hvis vi kigger tilbage på algoritme-projektet, hvor det var faldt i saturation, at den tog ud fra, det kan jeg ikke bruge. Forstået på den måde, at borgeren kan sidde og holde vejret og så bliver saturation lav, eller man kan lige have kolde fingre” - Informant 2

Dette indikerer at informanterne ikke var tilstrækkelig informeret om det prædiktive element af algoritmen, hvor den vil alarmere forinden der foreligger symptomer.

6.2.4 Et behov for information

Informanterne udtalte begge, at det ville være irriterende for dem, hvis der kom en pop-op besked, der beskrev, hvorfor algoritmen alarmerede hver gang. De ønskede derimod, at der var en definition af algoritmen tilgængelig på deres instruksbase. Dette indikerer, at begge informanter ønskede minimal information præsenteret i deres arbejdsgang.

“Ja altså jeg tænker jo egentlig, Vi har jo en udmærket instruksbase på TeleCare Nord, og hvis definitionen der, stod hvordan at den her alarm regnede risikoen ud, så ville det være

udmærket. Det er ikke noget jeg behøver at have i mit daglige arbejde, der behøves ikke at poppe bokse op hver gang der står på baggrund af det eller det spørgsmål, eller de og de målinger. Fordi jeg tænker alle os der er i TeleCare Nord vi er vældig grundig oplært i det. Og hvis man kender hvorfor det er den alarmere, så behøves der ikke komme noget op i ens hoved hver gang, så ville det blive mere et irritationsmoment” - Informant 2

Ønske om minimal information, herunder en definition af, hvordan risikoen beregnes, kan skyldes at der allerede foreligger en algoritme i OTH, hvor fortolkningen af dens triagering beskrives i instruksbasen. Derfor vurderes det vigtigt at forsøge at undgå brugen af pop-op-beskeder i prototypens design.

6.2.5 Et samlet overblik

På OTH's platform er der mulighed for, at brugerne kan få adgang til visualisering i form af grafer. På trods af dette blev det påpeget, at de manglede et hurtigt overblik over f.eks. antallet af exacerbation samt de vitale målinger for borgeren. Informanterne udtrykte også et ønske om en hurtig oversigt.

“Ja på den facon. Jamen. Ja, jeg synes, vi scroller meget. Altså jeg synes både og det her med, at hvis vi vil have et overblik, for eksempel, hvor mange exacerbationer har borgeren haft det sidste år. Det er nogle gange vi lige går ind og kigger på det. I forhold til jamen skal vi anbefale forebyggende antibiotika og så videre. Det er faktisk ret svært at få et overblik over. Og der kunne jeg godt tænke mig at hvis der var nogle bedre søgemuligheder eller hvis vi kunne have alle målingerne på et view. Jeg synes vi scroller meget op og ned.” - informant 1

Derudover blev det indikeret, at grafisk visualisering kunne bidrage til en hurtigere vurdering af borgerens risiko. Dette vil kunne fortælle TSA'erne, om værdierne var stigende eller faldende.

“Men jeg tænker jeg. For eksempel med vægt, altså der er der jo også selv der ændrer vægtgrænserne, og der kunne det måske være en fordel at vi bruger grafen, fordi når vi ændrer ved grænsen, så ændrer vi også farvekoderne. Og der skal man være lidt skarp på, at det snakker vi nu om vægttab eller vægtstigning. Fordi de bliver udlignet”. - Informant 1

“Ja, det har du ret i. Og det kan vi sådan set kun se til støtte i vores omsorgssystem. Ikke altså, som vi dokumenterer i når vi laver vægtændringer, fordi det er ikke altid gennemskueligt i opentele.” - Informant 2

6.3 Sammenfatning af krav til prototypen

For at kunne inddrage vidensbasens og miljøets resultater til specialets prototype, blev der identificeret specifikke krav. I alt blev der fundet fire krav til prototypen, som består af følgende: 1: *Global forklaring*, 2: *Lokal forklaring*, 3: *Interaktiv graf* og 4: *Stigning/fald i exacerbationsrisikoen*. Grundlaget for kravene baseret på specialets miljø og Vidensbase præsenteres i tabel 8.

Krav nr.	Krav	Miljøet	Vidensbase
1	Global forklaring	Informanterne er ikke blevet tilstrækkeligt informeret om algoritmens virkemåde Der vurderes, at algoritmen potentielt kan understøtte dem ved gule alarmer, som de ønsker For at støtte dem ved gule alarmer, vurderes der behov for at give mere information, end de selv italesætter, at der er behov for	En global og lokal forklaring på algoritmen vurderes essentiel for at brugere kan få en forståelse for AI-systemet Ved anvendelse af XAI-metoder bør man tage forbehold for at de kan være svære at forstå, hvorfor de skal simplificeres Visualisering er en effektiv måde at præsentere en forklaring på, hvor bar-charts hyppigst anvendes global og lokal forklaring
2	Lokal forklaring	Der bør tages forbehold for ikke at påvirke deres manglende overblik yderligere De ønsker information om hvordan risikoen beregnes	Lokal forklaring er at foretrække blandt sygeplejersker Global er kun relevant i opstartsfasen
3	Interaktiv graf	Grafer er vigtige for at få overblik over ændringer i vitale parametre Borgerens historik er et vigtigt redskab De bør understøttes i vurderingen af, hvorvidt der har været optræk til forværring Behov for et bedre overblik over borgerens tilstand Den bør med fordel gøres genkendelig for brugerne i forhold til de grafer, de er vant til	Interaktive grafer er effektive til at understøtte brugeres forskellige beslutningsprocesser

4	Stigning/fald i exacerbationsrisikoen	Informanterne ønskede at se om der var optræk til forværring	Info om stigning i risiko blev vurderet mere relevant end en enkelt risikovurdering
---	---------------------------------------	--	---

Tabel 8: Miljøet og vidensbasens udtræk for de respektive krav.

6.4 Prototypen

Med udgangspunkt i de identificerede krav blev der udarbejdet en low-fidelity af prototypen bestående af en række skitser der kan ses i bilag 13.1. Baseret på low-fidelity prototype blev der udarbejdet en high-fidelity prototype, som præsenteres i dette afsnit. Yderligere billeder af prototypen kan ses i bilag 13.2. Prototypens design fokuserede på to sider: Opgaveoversigt og monitoreringsoversigt, samt de fire krav fra sammenfatningen af miljøet og vidensbasen.

- Opgaveoversigten består af et overblik over de borgere der skal monitoreres, hvor imødekommelsen af krav 1: *Globale forklaring* samt krav 4: *Stigning/fald i exacerbationsrisikoen*.
- Monitoreringsoversigten består af oversigten for en borgers målinger, hvor imødekommelsen af krav 2: *Lokale forklaring* samt en krav 3: *Interaktiv graf* præsenteres.

I nedenstående afsnit vil imødekommelsen af kravene præsenteres. Kravenes komponenter vil blive henvist til som element, hvor de visualiseres med rød indtegning samt tekst. Tabel 8 viser en oversigt over hvilke af de to sider kravene imødekommes, som hvilke elementer der udgør kravet.

Sider	Krav	UI-element
Opgaveoversigten	1: Global forklaring	Element 2 og 3
Opgaveoversigten	4: Stigning/fald i exacerbationsrisikoen	Element 1
Monitoreringsoversigten	2: Lokal forklaring	Element 7
Monitoreringsoversigten	3: Interaktiv graf	Element 6

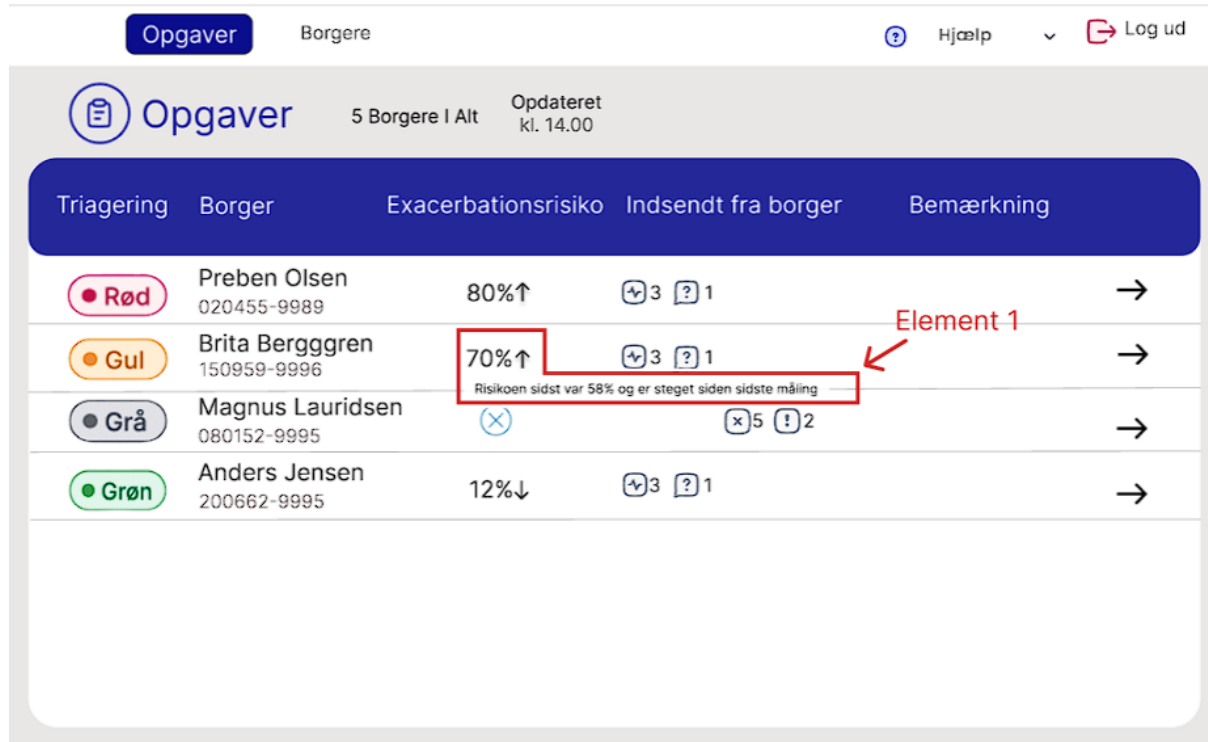
Tabel 8: Krav for den respektive side i prototypen.

6.4.1 Opgaveoversigt

Prototypens design tog udgangspunkt i, at brugerne var logget ind i systemet og allerede på siden: monitoreringsopgaverne (Billede 1). For at gøre oversigten mere realistisk og sammenlignelig med Telma, blev der udformet fire fiktive borgere med forskellige

triagering og tilhørende exacerbationsrisiko. Ud over en oversigt over borgernes risiko for exacerbation blev der tilføjet et ikon i form af pile, som indikerer, om risikoen var steget eller faldet siden sidste måling (Element 1). Dette blev gjort for at imødekomme krav 4: *Stigning/fald i exacerbationsrisikoen*.

Pile ikonerne blev gjort interaktive, så brugerne ved at holde musen over dem kun få en tilhørende tekstbeskrivelse, f.eks. at "Risikoen er steget siden sidste måling som var 65%". Dette skulle hjælpe til at brugerne bedre kunne forholde sig til alvorligheden af risikoen.



Triagering	Borger	Exacerbationsrisiko	Indsendt fra borger	Bemærkning
Rød	Preben Olsen 020455-9989	80%↑	3 1	→
Gul	Brita Bergggren 150959-9996	70%↑ Risikoen sidst var 58% og er steget siden sidste måling	3 1	→
Grå	Magnus Lauridsen 080152-9995	×	5 2	→
Grøn	Anders Jensen 200662-9995	12%↓	3 1	→

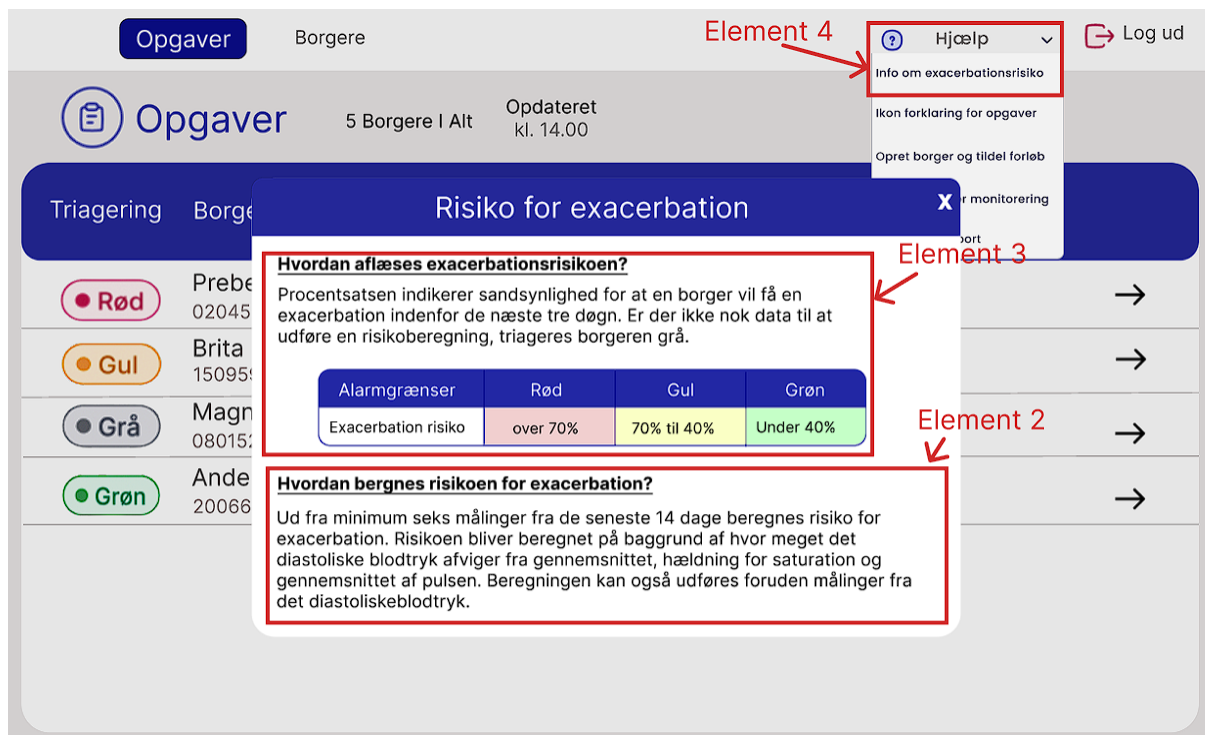
Billede 1: Opgaveoversigt med 5 borgere. Element 1 viser imødekommelse af krav 4.

For at brugerne kunne bedre forholde sig til exacerbationsrisikoen, blev der udformet et dokument med en global forklaring som udgør krav 1 (Element 2 og 3).

Forklaringen blev simplificeret, så den kun bestod af tekstbeskrivelse. Dette blev vurderet relevant grundet fund fra vidensbasen, der beskriver at global forklaring vurderes som værende svær at forstå og kun relevant ved opstartsfasen (95,98,99).

Derfor blev der fokuseret på at beskrive i tekstformat, hvilke dataelementer der anvendes, og hvordan de anvendes ved udformning af et output (Element 2). Der blev beskrevet, at algoritmen beregner risikoen på baggrund af hældningen af saturation, svingninger i det diastoliske blodtryk og gennemsnittet af pulsen de seneste 14 dage. Ydermere blev der udviklet en visuel repræsentation samt en beskrivelse af, hvordan risikoen skal aflæses (Element 3) (102). Da denne type forklaring vurderes kun værende relevant i opstartsfasen, blev der valgt at lægge den globale forklaring under hjælp funktionen, så den kunne tilgås ved behov (Element 4). For at understøtte UX-ekspertens

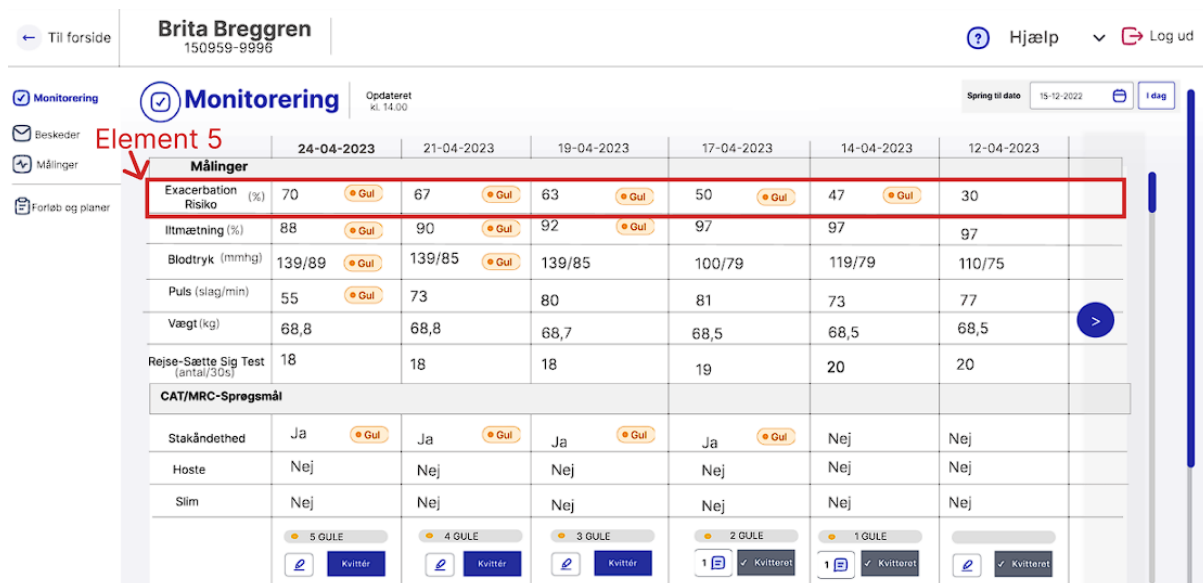
forståelse for UI'en, blev der tilføjet et dokument under hjælpefunktionen, der forklarede ikonerne anvendt i opgaveoversigten (Bilag 13.2).



Billede 2: Design af en forenklet global forklaring. Element 2 beskriver hvordan den beregnes. Element 3 beskriver, hvordan risikoen aflæses.

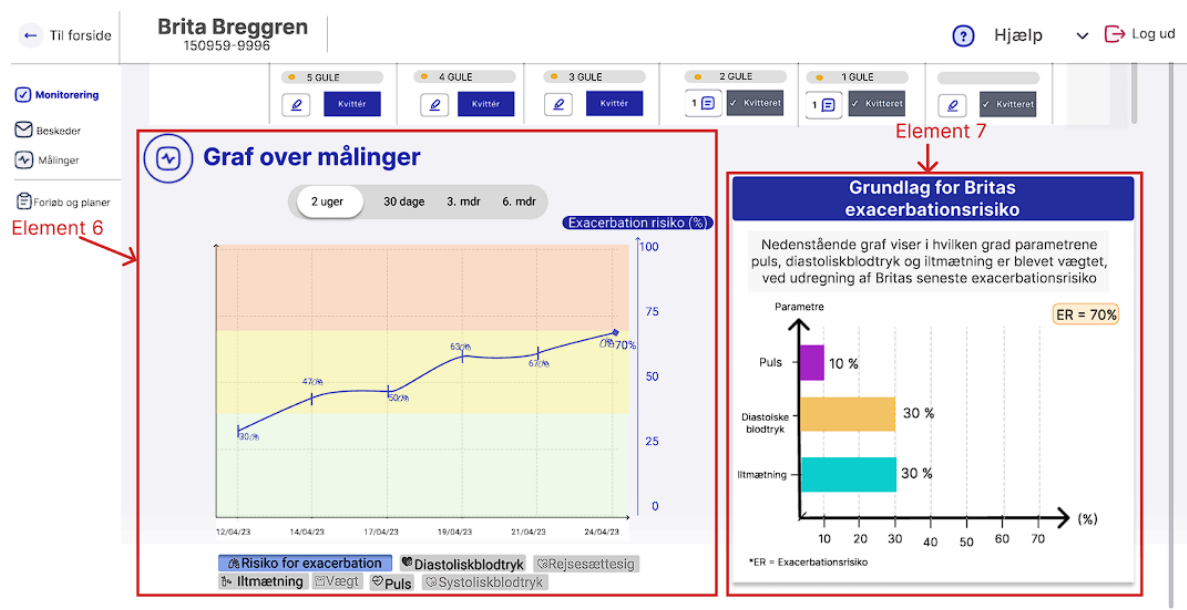
6.4.2 Monitoreringsoversigt

I miljøet blev der af informanterne italesat et behov for støtte ved gule alarmeringer. Derfor blev anvendeligheden af exacerbationsrisikoen undersøgt i henhold til et tvetydigt use-case scenarie, hvor der ikke foreligger konkrete retningslinjer for, hvordan TSA'erne skal agere. Exacerbationsrisikoen blev placeret øverst ved tabeloversigten, som del af målinger (Element 5). Dette blev gjort for at brugerne hurtigere kunne se hvordan exacerbationsrisikoen løbende er blevet ændret, baseret på borgerens målinger. Det vurderes, at dette kunne give TSA'erne et overblik over hvornår risikoen begyndte at stige, og dermed understøtte dem i analyse af om der har været optræk til forværring i længere tid.



Billede 3: Placering af exacerbationsrisiko i monitoreringsoversigten kan ses i element 5.

Ved afdækningen af miljøet blev det understreget vigtigheden af et grafisk overblik over borgerens historik og alarmering, og et behov for at reducere antallet af klik. I Telma foreligger graferne under "Målinger" der afbildes i venstre navigationspanel på figur 10, hvorfor de skal ud af monitoreringsoversigten for at tilgå dem. Baseret på dette blev der udarbejdet en interaktiv trend-graf til, at sammenholde alle vitalmålingerne for de seneste seks målinger med exacerbationsrisikoen, som blev placeret under tabeloversigten (Billede 4). Her var der mulighed for fra og tilvalg af de ønskede vitale målinger. Dette skulle medvirke til hurtigere overblik ved sammenholdelse af exacerbationsrisikoen og de vitale målinger, uden at skulle forlade siden (Element 6). Derfor ville placeringen af grafen for værdierne, medvirke til minimering af klik. Den interaktive graf blev yderligere designet, så den kunne understøtte visualisering af en enkelt parameter i henhold til dens grænseværdier, eller flere parametre i samme graf. Den lokale forklaring på borgerens seneste exacerbationsrisikoberegning, blev placeret på højre side af grafen (Element 7). Denne opsætning vurderes at have potentiale for at øge forståelse for exacerbationsrisikoen, ved at muliggøre sammenholdelse af exacerbationsrisikoen visualiseret ved krav 2: *lokal forklaring* med de vitale parametre i krav 4: *interaktive graf*.



Billede 4: Design for lokal forklaring i element 7, og den interaktive graf i element 6.

6.5 Resultater fra evaluering af prototype

I dette afsnit vil evalueringens resultater fremlægges med udgangspunkt i heuristikker og graduering af overtrædelser samt deres forbindelse med kravene for prototypen. Herefter præsenteres resultaterne fra den tematiske analyse af udtalelser fra både interaktionen og interviewet. Dette vil blive gjort med udgangspunkt i Frank Gous fire kerneelementer for UX bestående af: value, usability, adaptability, desirability.

6.5.1 Præsentation af evaluatore

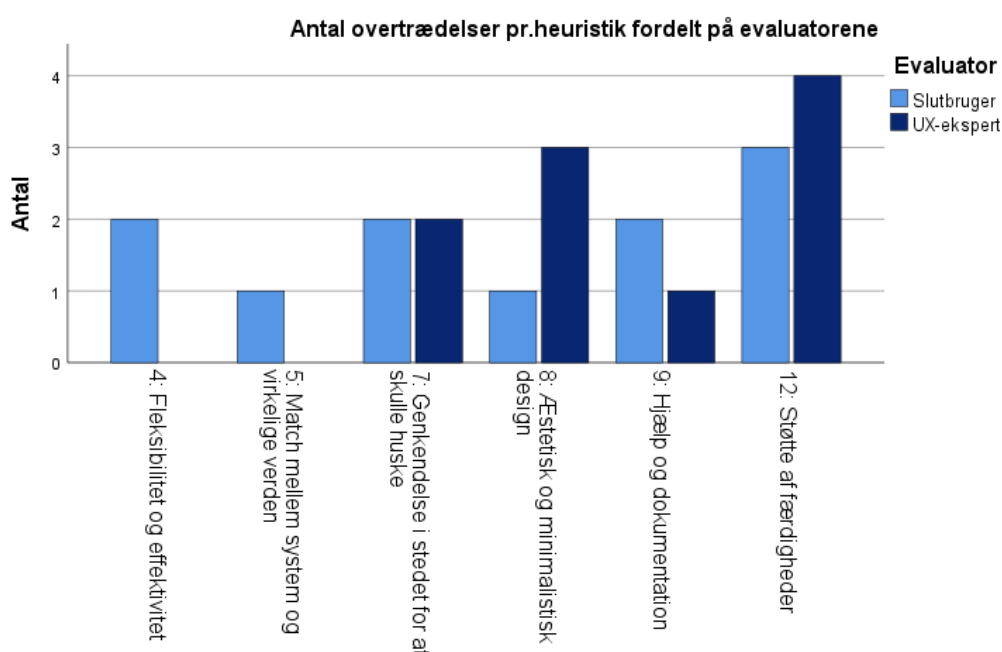
Evalueringen af prototypen blev udført med to evaluatore, bestående af en slutbruger og en UX-ekspert (Tabel 9). Slutbrugeren er uddannet sygeplejerske, og har medvirket i TeleCare Nord KOL siden opstart i 2015. UX-eksperter er uddannet produktdesigner, og har de seneste seks år i sit daglige virke arbejdet med design af hjemmesider og applikationer for at forbedre brugeroplevelsen.

Demografiske beskrivelse af informanter		
	Profession	Erfaring
Slutbruger	Sygeplejerske	8 års erfaring inden for monitorering af borgere med KOL.
UX-ekspert	Produktdesigner	6 års erfaring inden for UX-design

Tabel 9: Demografiske informationer over evaluatore i evaluering.

6.5.2 Heuristiske overtrædelser

Ved analyse af evaluatorernes interaktion under den heuristiske evaluering, blev der identificeret i alt seks heuristikker. Ud af de seks heuristikker blev der noteret 21 overtrædelser, der betegnes som en alvorlighedsgrad over nul. Fordeling af heuristiske overtrædelser blandt evaluatorerne kan ses i figur 12. Den heuristisk der forekom hyppigst var 12: *Støtte af færdigheder*, hvilket forekom syv gange, fire for UX-eksperter og tre for slutbrugerne. Slutbrugerne identificerede i alt 11 overtrædelser og UX-eksperter identificerede ti (Figur 12).

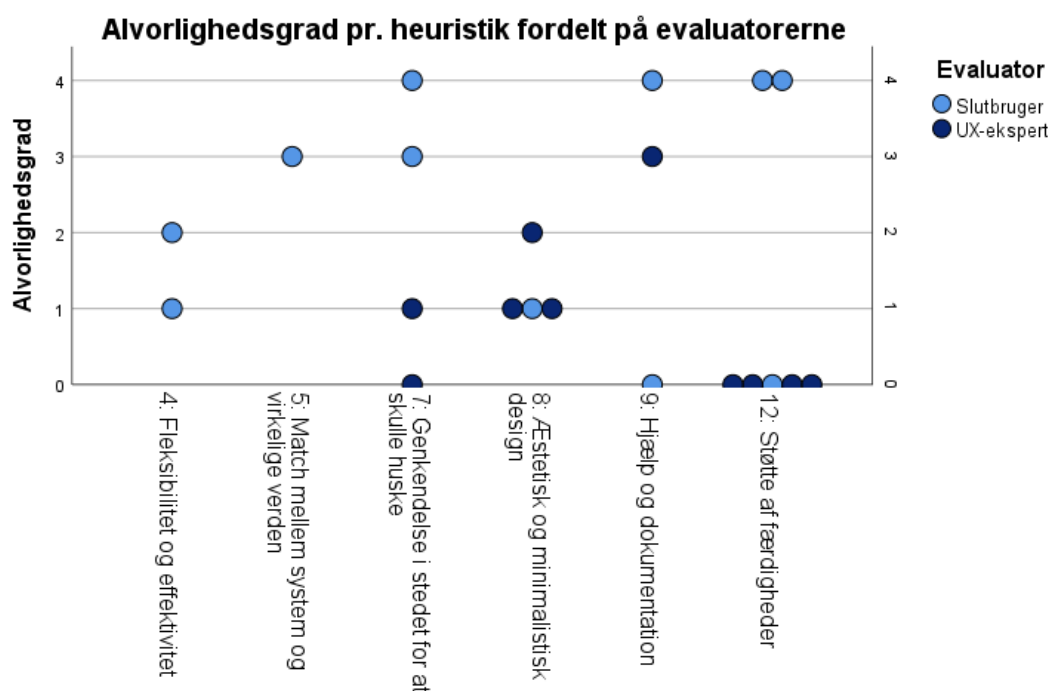


Figur 9: Antal overtrædelser pr. heuristik for hver evaluator. Slutbrugerne er repræsenteret med lyseblå farve og UX-eksperter er repræsenteret med mørkeblå farve.

Alvorlighedsgrad på fire blev hyppigt anvendt ved analyse af slutbrugerens overtrædelser. Hvor en alvorlighedsgrad nul blev hyppigst anvendt ved analyse af UX-eksperterens overtrædelser. Slutbrugerne identificerede i alt ni overtrædelser, hvor UX-eksperter identificerede fem overtrædelser. Ud af de 21 overtrædelser var den gennemsnitlige alvorlighedsgrad 1,62.

Der blev yderligere identificeret to dubletter, som defineres som samme overtrædelse og gradueringsniveau påpeget af begge evaluatorene. Den ene dubletter opstod ifm. *heuristik 8: æstetisk og minimalistisk design*, hvor begge evaluatorene fandt dele af krav 1: *Global forklaring* meget teksttung, der resulterede i en alvorlighedsgrad på to. Den anden dubletter opstod ifm. *heuristik 12: Støtte af færdigheder*, hvor hverken evaluatorene havde besvær med

krav 4: *Stigning/fald i risikoen*, der resulterede i en alvorlighedsgrad på nul. Se tabel 10 for oversigt over kravene og deres alvorlighedsgrader.



Figur 10: Grad af overtrædelse og type heuristik for hver overtrædelse hos evaluatorene. Slutbrugeren repræsenteres med lyseblå farve og UX-eksperter med mørkeblå farve.

6.5.3 Prototypens krav og grad af overtrædelse

Ved at sammenholde specialets krav og de identificerede heuristiske overtrædelser (Figur 10)). Kan det ses, at slutbrugeren generelt har væsentlige højere overtrædelsesgrader end UX-eksperter. Dog er resultaterne taget med forbehold for slutbrugers manglende erfaring i at anvende grafer, og UX-eksperter manglende kontekstforståelse. Ud af de fire krav havde krav 2: *Lokal forklaring* den højeste gennemsnitlige usability overtrædelse på 3.3 hos slutbrugeren. Hvorimod krav 1: *Global forklaring* havde den højeste gennemsnitlige usability overtrædelse på 1.3 hos UX-eksperter. Oversigt over de fire krav og deres gradueringer kan ses i nedenstående tabel 10.

Sider	Krav	Evaluering	Grad af overtrædelse	
			Slutbruger	UX-eksperter
Opgaveoversigten	1: Global forklaring	Placering	0	3
		Forståelse	1	1
		Anvendelighed	4*	0
Monitoreringsoversigten	2: Lokal forklaring	Placering	3	1
		Forståelse	3	0
		Anvendelighed	4	0
Monitoreringsoversigten	3: Interaktiv graf	Placering	4	0
		Forståelse	1	1
		Anvendelse	2	2
Opgaveoversigten	4: Stigning/fald i exacerbationsrisikoen	Aflæsning	0	0

Tabel 10: Krav sammenholdt med gradueringer. *Ved anvendelighed af den globale forklaring, blev der identificeret to overtrædelser på fire.

De fire overtrædelser med alvorlighedsgrad på fire, var i henhold til anvendelse af krav 2: *Lokal forklaring*, anvendelse af krav 1: *Global forklaring* og placeringen af krav 3: *Interaktiv graf*. Her kan det ses, at slutbrugerne for krav 1: *Globale Forklaring* havde to usability overtrædelser på fire, betydende at denne forklaring ikke er anvendelige, hvorfor problemerne skal løses før XUI'en kan anvendes. Ligeledes blev der observeret en usability overtrædelse på fire for krav 2: *Lokal forklaring*, hvor dens indhold blev vurderet ikke relevant for støtte ved identifikation af exacerbationer og skal løses før videre anvendelse. Her viser UX-eksperterne derimod en usability-overtrædelse på nul for anvendelse af både den globale og lokale forklaring, der indikerer at der ikke var en udfordring overhovedet. Ved krav 4: *Stigning/fald i exacerbationsrisikoen* var begge evaluatorene enige om overtrædelsen, og fandt ikke stigning og fald i exacerbationsrisikoen værende en udfordring. Ligeledes havde begge evaluatorene en

overtrædelse af den interaktive graf på to, grundet besvær med at aflæse grafer. Denne gradueringsgrad viser en mindre usability, men bør ikke prioriteres lige så højt som overtrædelse tre og fire.

6.5.4 Udtalelser fra evalueringens

Ud fra kerneelementerne fra Frank Gou's fire fokusområder for user experience vil resultaterne præsenteres i følgende rækkefølge: usability, value, adaptability og desirability. Diskussion vil tage udgangspunkt i kravene, der blev præsenteret i tidligere afsnit jf. *afsnit 6.3 - sammenfatning af krav til prototype tabel 8*.

Usability: Er produktet nemt at bruge?

Slutbrugeren gav udtryk for, at det var svært ved at forholde sig til procentsatsen for exacerbationsrisikoen på forsiden, idet det var noget nyt man skulle forholde sig til (Billede 1). Slutbrugeren påpeger behov for mere information, før de kan anvende exacerbationsrisikoen. Dette blev ikke italesat som et problem af UX-eksperter.

"Hm, altså jeg ved jo godt hvad det betyder. Den er nok høj." - UX-ekspert

"Den her exacerbationsrisiko med nogle procenter, og når jeg ikke sådan lige kender baggrunden for det, så ville jeg jo hvis jeg først bare skulle kigge, så tænker jeg hvad er det for noget ... Jeg har ikke set det der før ... Det er noget nyt, så det skal jeg jo have noget viden om før jeg ligesom kunne bruge det til noget tænker jeg." - Slutbruger

Selve pilen, der blev designet for at imødekomme krav 4: *Indikation for stigning/fald i risikoen* (Billede 1), blev læst og forstået uden problemer. Både slutbrugeren og UX-eksperter kunne aflæse, at der havde været en stigning i exacerbationsrisikoen.

Slutbrugeren fandt det meget intuitivt at lede efter en forklaring under hjælpe-funktionen på grund af designets similaritet med Telma. Dette var dog ikke gældende for UX-eksperter, som forventede at kunne fremsøge forklaringen ved at trykke på eller holde musen over exacerbationsrisikoen. Trods dette bemærkede UX-eksperter selv, at forklaringen nok lå under hjælp funktionen, men at dette ikke var det man ville forvente.

Ved gennemlæsning af krav 1: *Globale forklaring* (Billede 2) udtalte begge evaluatoren, at den gav et godt overblik over, hvad der menes med exacerbationsrisikoen. Dog påpegede begge evaluatoren, at element 2 af den globale forklaring var meget teksttung. De fandt dog begge, at tabellen med triagefarverne for exacerbationsrisikoen var særdeles anvendelig med hensyn til forståelse for exacerbationsrisikoen (Element 3).

“I går ligesom i detaljer med hvordan det beregnes, og hvordan det skal aflæses... Ja, jeg ved ikke hvordan jeg skal forklare det, men den her graf (tabel) giver et rigtig godt overblik over hvordan det skal aflæses, og den alene giver mig allerede en bedre forståelse af hvad jeg skal gøre, eller hvad det betyder.” - UX-ekspert

Ved evaluering af krav 3: *Lokal forklaring* (Billede 4), havde slutbrugeren svært ved at finde den, hvor der var tvivl om, hvorvidt den lå under “Forløb og planer” i navigationspanelet. UX-eksperten gættede sig til dens placering. Når det kom til selve indholdet i forklaringen, var det nemt for UX-eksperten at aflæses, men slutbrugeren havde svært ved at forholde sig til den, fordi den var for forskningsbaseret. Slutbrugerne udtalte derudover, at den ikke ville bruges ved monitorering af borgerne.

“Jeg synes den er meget forskningsbeskrevet er det sådan eller? Jeg tænker ikke det ville være noget... Jeg sådan lige umiddelbart tænker, at det ville være noget jeg vil bruge. Men det er sikkert altså megarelevant. Det er slet ikke det, men jeg tænker ikke, at det er noget i min daglige bruger, hvad hedder det?” - slutbruger

“Ikke i selve monitoreringsdelen, der tror jeg ikke, at jeg bruger den” - Slutbruger

Det blev italesat af slutbrugeren, at både krav 3: *Interaktive graf* og krav 2: *Lokale forklaring* ikke var nem at finde. Slutbrugeren havde en forventning om, at den lå under “målinger” ved navigationspanelet. For at finde forklaringen havde slutbrugeren brug for hints, mens UX-eksperten ikke havde besvær med at finde den. Både slutbrugeren og UX-eksperten kunne interagere med grafen, og få fremvist alle parametrene. Derudover italesatte slutbrugeren at de ikke ville have tid i praksis til at skulle sammenholde den lokale forklaring med den interaktive graf, ved hver borger.

“Jamen så trykker jeg måske på den hedder iltmætning dernede, og så trykker jeg, hvis jeg skulle se pulsen sammen, så trykke på pulsen og så kom de frem. Ja, nu kom der jo noget ind i det. Så skal jeg jo lige være helt sikker på hvad det er” - Slutbrugeren

“Men jeg vil ikke kunne altså, hvis jeg skulle sidde og monitorere, så vil jeg ikke tror jeg ikke bruge den 100% hver gang” - Slutbrugeren.

I forhold til selve interaktionen med prototypen, blev der af slutbrugeren påpeget, at de lige skulle få et overblik over dens design og lige finde forklaringerne. Hvorefter det ikke var så svært at anvende prototypen. UX-eksperten havde heller ikke yderligere kommentarer til interaktionen med prototypen, og vurderede at interaktionen var “*helt fint*”, samt at forklaringerne var placeret de rigtige steder.

Value: Er produktet nyttigt?

Ved evaluering af prototypen blev krav 1: *Global forklaring* (Billede 2) vurderet som en nyttig forklaring, der imødekom slutbrugerens behov for forklaring. Dette blev også bekræftet af UX-eksperter, der vurderede, at den globale forklaring gav mest forståelse. På trods af dette påpeger slutbrugerens behov for yderligere informationer og forklaringer ved den globale forklaring. Dette inkluderede behov for information om evidensgrundlaget for exacerbationsrisikoen og dens nytteværdi. Derudover udtrykte slutbrugerens undren over anvendelsen af exacerbationsrisikoen, uden en redegørelse for algoritmens input.

“Men igen tror jeg der mangler den der baggrund for hvorfor den er kommet. Hvis den kommer, så er der noget understøttende evidens.” - Slutbruger

“Men jeg undrede mig, jeg skulle lige tænke ‘nå hvor er det systoliske henne i det her’, og jeg tænker jeg bliver lidt usikker. Det var lige mine tanker omkring det.” - Slutbruger

Ved evaluering af krav 2: *Lokale forklaring* (Billede 4) beskrev slutbrugerens, at den kunne være af nytte ved at give en bedre forståelse, men at den var svær at aflæse. Derudover blev det italesat, at den lokale forklaring ikke ville anvendes ofte ved monitorering, men kunne anvende som et sparringsredskab. I forlængelse heraf mente slutbrugerens, at den måske var overflødig, end den ville være gavnlig. Derfor blev der givet et forslag til placeringen et andet sted, hvor den kunne tilgås ved behov.

“Ikke i selve monitoreringsdelen, der tror jeg ikke, at jeg bruger den. Men det kunne godt være, som jeg fik sagt tidligere, at man kunne bruge den sammen med sine kollegaer eller en ekstern kollega. Man kunne godt få dem til at kigge på den. Det kan også være, at de inde på sygehuset verden kigger meget mere på sådan en graf, så får de jo understøttelse.” - Slutbruger

“Der tænker jeg også, at man hele tiden skal finde ud af hvor meget er fyldt. Måske var det bedre, at det ligesom var et andet sted hvor jeg sådan ligesom kunne klikke ind på, så det ikke fyldte mit øje. Sådan lige min umiddelbare tanke ville være, at når jeg ikke kender den så godt. Det er så nyt det her Telma, og så skal der ikke ligge for alt for meget ind i det her overblik her. Selvfølgelig fordi så vil jeg sidde og rode rundt i det og frem og tilbage.” - Slutbruger

Derudover blev det påpeget af slutbrugerens, at placeringen af den interaktive graf, medvirkede til at minimere klik, som var dejligt, men at dette medførte andre konsekvenser.

I forbindelse med krav 3: *Interaktiv graf* sagde slutbrugeren, at det var bekvemmeligt med placering som medvirkede til at minimere klik. Dette ville medføre at der skulle scrolles mere i tabellen med de indsendte målinger.

"Jeg tænker på 2 ting. Jeg tænker den ene er at det giver jo hurtigere mening i at man sparer lidt for nogle klik man skal over i et andet skærmbillede. Det der bare er min erfaring. Nummer to ting. Det er, at jo mere der ligger på en side, jo mindre bliver det. Jo mere, skal man måske også scrolle op i selve hvad hedder det monterings overblikket." - Slutbruger

Overordnet blev prototypen vurderet at kunne være anvendelig til at understøtte monitorering af borgere med KOL, både af UX-eksperten og slutbrugeren. Hvor slutbrugeren understregede, at exacerbationsrisikoen ikke vil kunne anvendes alene og at der var behov for yderligere forklaring. Det blev i samme flæng vurderet, at prototypen kunne være anvendelig til både erfarne og uerfarne slutbrugere.

"Den kan jeg jo ikke slå fra nej, men det er klart, selvom man har erfaring, så kan man jo også godt tænke, at det er lidt spændende alligevel at have den med. Altså man kan måske bruge den med kollegerne, man kan spare med hinanden omkring den. Og sige, 'der er altså lige noget her'. Vi har faktisk set en stigning. Der er udregnet en risikovurdering, så det kunne godt være, at man kunne bruge den." - Slutbruger

UX-eksperten mente, at det kunne være relevant med yderligere information, der beskrev, hvordan der skulle ageres ud fra exacerbationsrisikoen.

"Det synes jeg også er supervigtigt at vide, for dem der har med det at gøre. Sådan en som mig, måske ikke så meget. Men den der er inden for feltet og forstår de her ting. Der synes jeg det giver mening. Eller det kan jeg tænke mig frem til, at det vil være en stor hjælp." - UX-ekspert

Adaptability: Er produktet nemt for brugeren at begynde at anvende?

Begge informanter påpeger anvendeligheden af den globale forklaring, hvor specielt den første del af beskrivelsen bidrog til en forståelse. Dog havde begge evaluators enkelte udfordringer, hvilket førte til en samlede vurdering af den globale forklaring. For UX-eksperten skyldes dette manglende kendskab til fagbegreberne, som bliver anvendt i sundhedsvæsenet.

“Det var min første tanke, men selvfølgelig hvis man er hjemmepleje og vant til at kigge på det her, og forstår alle de begreber der bliver brugt, så er det ikke noget problem.”- UX-ekspert

For slutbrugeren handlede det især om forklaringen, der beskrev, hvordan exacerbationsrisikoen blev beregnet. Her blev der udtalt en udfordring i forståelsen af denne tekst, hvor slutbrugeren italesatte, at dette var en lang tekst og havde brug for mere tid til at tænke over forklaringen.

“Men der er noget ved den her forklaring, som jeg synes, der var sådan lidt svært at forstå. Jeg vil sige der står meget tekst, men jeg tror også der er også nødt til at stå noget tekst, så det skal jeg vænne mig til. ... Jeg tror den øverste del af hvordan den læses er fin nok. Så den næste der ‘hvordan beregnes risikoen’, der skal jeg sådan lige ind og tygge lidt mere på den.” - Slutbruger

I forbindelse med aflæsningen af den globale forklaring var der tvivl hos UX-eksperten på trods af italesættelse af, at den var forståelig. Dette skyldes, at der på trods af visualisering med farver og tal i element 2, var det ikke intuitivt, hvorvidt der var tale om høj eller lav risiko for den gule alarmering.

“Grund til at jeg ligesom, lige før var lidt i tvivl hvad jeg skulle svare, var fordi "høj", hvad menes der med høj? Fordi gul kunne både være ikke så høj, men det kunne også være højt. Så det var mere, der lå og det er også noget man kunne overveje, om man måske skulle i stedet for at skrive farverne skrive om den er høj eller lav eller neutral eller sådan noget, for at en bedre kontekst ja i forhold til status”- UX-ekspert

I den lokale forklaring var der delte meninger om, hvor nem forklaring var at anvende. Dette handlede både om overblik, teksten, valg af graf og brugen af den. Begge evaluatore kom med deres forslag til, hvordan den lokale forklaring bliver nemme at anvende

“Det er fuldstændig åndssvagt, men jeg synes faktisk nogle gange det der med, at hvis man bare lige vender den om, altså hvis du bare lige vender hvad hedder de værdierne.” - Slutbruger

“I stedet for at bruge farve, men også bruge noget tekst der forklarer det. I forhold til. Fordi det jeg umiddelbart tænker, det er, at jeg ikke altid ved hvad farve betyder. Det med kun at bruge farve, det kan nogle gange, det kan bare være svært at huske.”- UX-ekspert

For slutbrugeren blev der ligeledes italesat manglende kendskab og erfaring til at anvende en graf, hvor det ikke menes, at den lokale forklaring ville anvendes som et monitoreringsværktøj. Derudover mente slutbrugeren at den lokale forklaring var svær at læse, hvilket gjorde slutbrugeren heller ville bruge den interaktive graf.

“Jeg tror det er mere min forståelse. Det gælder lidt mit syn. Det er ikke fordi at grafen er forklert. Jeg synes det er fint at have begge muligheder, altså at man har både en graf og nogle tal. Fordi den der kan sidde en anden, der godt kan lide en graf. Det kan også være, jeg skal bare vænne mig til det.” – Slutbruger

“Jeg synes den er svær lige at kigge på. Det er nok bare lige mig. Jeg tror bare, jeg ved det ikke. Så tror jeg faktisk, at jeg hellere ville bruge grafen (interaktiv graf)” - Slutbruger

For UX-eksperthen gav den lokale forklaring mening, og havde ikke yderligere besvær ved at skulle anvende grafen. Det blev dog sagt, at visualiseringen af den lokale forklaring bidrog til et bedre overblik over, hvilke parametre der vejede mest for den respektive borgers risiko.

“Ja, det giver meget klart overblik over hvilke ting der bliver taget med ind i den beregning og hvad der er resultatet. Og så også farverne hjælper mig med at forstå, at det er tre forskellige ting.” - UX-ekspert

Flere gange under vurdering af nemhed i brug af den interaktive graf, blev der italesat manglende erfaring i brug af grafer. Hvor der ikke var en indikation for, at slutbrugeren ville kunne tjene det formål med et hurtigere overblik for exacerbationer.

“Jeg har ikke brugt den så meget endnu, så jeg skal virkelig have fingeren i det, før jeg ligesom synes at den giver mening. Men der skal også en del målinger til, for at det giver mening, altså der skal jo være kommet nogle målinger.” - Slutbruger

“Det ved jeg ikke lige hvad jeg tænker. Det er måske godt nok, at man kan det, men som sagt så synes jeg det her ... jeg synes simpelthen at grafer, det er virkelig svært at aflæse. Jeg har ikke brugt det så meget.” – Slutbruger

For UX-ekspert opstod der forvirring i anvendelse af den interaktive graf ved første møde, hvor der var en forventning om, at graferne var noget der skulle klikkes af og på. Dette indikerer, at den kan være svær at begynde at anvende, især for farveblinde, idet farverne af graferne ikke var brugervenlige. Dette kunne potentielt udgøre en udfordring ved

anvendelse, hvor det ikke kunne være svært at skille farverne og dermed målingerne fra hinanden.

“Og det troede jeg var, fordi de ikke var blevet valgt. Så det jeg ligesom prøvede på lige før, var at trykke på dem her 2, det var ligesom at sige dem vil jeg ikke have. Fordi det lignede, at de var valgt, men nu da jeg så klikker på dem, kunne jeg se at det ikke er casen. Så nu gør jeg det rigtigt.” -UX-ekspert

“Ja, det er svært at se, også når man tænker på farvekombinationen af de to. Så er der jo nogle der er farveblinde. Det er i hvert fald noget vi går op i firmaet, accessibility. Og der er de to farver, er der mange mennesker der ikke kan se forskel på, fordi de indeholder lidt rød og lidt grøn. Så det er bare lige sådan en hurtig ting.” - UX-ekspert

Desirability: Er produktet anvendeligt og bedre end alternativet?

Slutbrugeren mente, at der på trods af en exacerbationsrisiko med tilhørende forklaring ikke var nok til at lægge tidligere erfaring væk og stole på algoritmen alene. Der blev også nævnt, at exacerbationsrisiko blev set på som et nyt værktøj, og potentielt kunne anvendes efter tilvænning.

“Jeg er nok sådan en lidt langsom starter, så jeg skal nogle gange lige have tingene og kigget på flere gange, for at jeg ligesom tænker, at kan jeg godt adoptere og bruge den til noget, men det kan at det sagtens kan være at man kan bruge den.” - Slutbruger

“Men nu her altså, så jeg tænker lidt at jeg kender min borger så godt til, at det kan godt være. Det er bare fordi jeg tager det lidt for givet, men jeg tænker lidt at ud fra de målinger jeg ellers ser, så ved jeg ikke lige helt om jeg behøver da en knap.” – Slutbruger

Ved den lokale forklaring var slutbrugeren ikke overbevist om, hvordan den skulle anvendes. Her foretrak slutbrugeren tal i stedet for grafer.

“Det er også fordi at jeg ikke bruger den graf på samme måde, som måske andre gør. Men jeg ser meget mere tallene, og jeg kunne bedre lide at se tallene.” - Slutbruger

For den interaktive graf mente slutbrugeren, at anvendelsen af denne ville tage for lang tid, hvor der blev refereret til de mange borgere der skulle monitoreres og vurderes værdier på. Det blev nævnt, at dette kunne være en udfordring, hvortil slutbrugeren foreslog, at grafen kunne være mere velegnet til administrativt arbejde

“Jeg tænker lidt at, jeg har heller ikke tid til at gå ind på hver enkelt og se på hver grad og se hvordan det hele er sammenholdt. Det kan være jeg kan gøre det ved et administrativt arbejde, hvor man så kan gå indtil man kan gå tilbage og kigge, men i her og nu billedet, hvor jeg ser at borgeren er gul og jeg for eksempel skal ind og forholde mig til nogle målinger, der er jeg nødt til ringe borgerne op, så det er det jeg ville gøre.”- Slutbruger

7. Diskussion

I følgende afsnit vil projektets væsentligste resultater og metodiske begrænsninger diskuteres og sammenholdes med eksisterende evidens. Dette sammenholdes med specialets formål, om en brugercentreret XUI kunne øge anvendeligheden af exacerbationsrisikoen ved telemonitorering.

7.1 Vidensbase

For at afdække vidensbasen blev der udført en litteratursøgning med to søgeblokke: 'XAI' og 'UI'. Derefter blev søgningen udført i fem databaser.

Databaserne blev valgt baseret på forfatterens tidligere erfaring med dem, eller gennem en struktureret søgning der indikerede deres relevans. Da hver af de valgte databaser dækker forskellige fagområder, der berører XUI, blev det antaget, at de ville kunne dække store dele af evidens inden for XAI og UI for sundhedsvæsenet. Idet anvendelsen og forskningen inden for XAI i sammenhæng med UI er et nyere og voksende forskningsområde, blev der under udformningen af specialet løbende lavet ustruktureret søgninger. I forbindelse med disse søgninger blev der identificeret databaser, med yderligere publikationer om XUI og sundhedsvæsenet, som kunne være relevante. Dette omfatter særdeles nyeste forskning indenfor XUI, som endnu ikke er blevet peer-reviewed. Det er dog vigtigt at bemærke, at der skal tages forbehold for kvaliteten af disse, da de muligvis ikke opfylder tidsskriftstandarden (103). Dette kunne have medvirket til relevant viden if. udviklingen og evaluering af prototypen, og dermed en værdifuld kilde til indsigt.

Speciale forfatterens ustrukturerede søgning bidrog desuden til øget forståelse for terminologianvendelsen af XUI og XAI. Det bemærkes, at en del af forskningen også anvender udtrykket XAI til at beskrive en XUI. Dette er på trods af at en XUI henviser til muligheden for at interagere på UI med den samlede output for en XAI, for at give brugeren en bedre forståelse for beslutningsprocessen (47). På baggrund af denne nye forståelse vurderes det, at den anvendte søgeprofil i dette speciale, muligvis har ekskluderet relevante publiceringer, da der anvendes søgeordene for blok to, der omhandler UI. Derfor vurderes det, at søgeprofilen kunne have været forbedret, ved at ændre søgeblokke for UI til at omhandle 'Health Care' i stedet. Denne kombination ville potentielt kunne medvirke til at præcisere søgningen og sikre inddragelsen af relevant litteratur for udvikling og evaluering af prototypen anvendt i dette speciale.

7.2 Miljøet

Som tidligere påpeget, kan en hurtigere identificering og tidligere forebyggende behandling af exacerbationer bidrage til at øge borgeren livskvalitet og reducere omkostninger forbundet hermed. Af denne grund blev der testet en prædiktionsalgoritme, der havde til formål at forudsige exacerbationer tre døgn forud, og som skulle understøtte TSA'ernes vurdering af borgerens sundhedstilstand. Tidligere forskning af denne algoritme, var der ingen indikationer på, at TSA'erne fandt den nødvendighed for dem i arbejdsgangen. Dette medførte, at specialet valgte at afdække miljøet ved at interviewe TSA'er, som skulle give en indikation på, om TSA'erne selv mente de havde brug for en exacerbationsrisiko og hvilke udfordringer der gjorde sig gældende i forbindelse med den daglige praksis. Resultaterne fra miljøet viste, at TSA'erne blev givet en grundig oplæring i aflæsningen af algoritmen. Hertil blev det italesat, at denne oplæring havde givet TSA'erne fuld forståelse for algoritmen, hvor behov for yderligere information om dens udgangspunkt ikke var nødvendig. Dog bliver der udtalt et behov for støtte ved de gule alarmer, i form af information om, hvorvidt der var optræk til exacerbationer. Denne udtalelse indikerer alligevel en manglende forståelse for algoritmen og dens formål. Denne udtalelse understøtter specialets beslutning om at udvikle en XUI, til at øge deres forståelse og anvendelse af algoritmen.

7.3 Evalueringen

På baggrund af evalueringens få evaluatore kan der ikke vurderes med sikkerhed, hvorvidt prototypen med vil øge anvendeligheden af algoritmen.

Her ses en tydelig begrænsning, da resultaterne fra evalueringen kun kan drages på den enkelte slutbruger, som betyder at resultaterne ikke kan generaliseres. Dette skyldes, at forskellige slutbrugere har forskellige behov, præferencer og ekspertise niveauer i forhold til at forstå AI-systemer, som vil påvirke, hvordan de ønsker at blive støttet i deres arbejde (59,104). Denne forskel i præferencer ses ved resultaterne fra evalueringen og litteraturen, hvor evaluatorene foretrak krav 1: *Global forklaring* (Billede 2) frem for krav 2: *Lokal forklaring* (Billede 4), som står i kontrast til litteraturen (96,105,106). Derfor har den lave samplestørrelse medvirket til, at resultaterne ikke med sikkerhed kan generaliseres til andre TSA'ernes behov. På trods af dette har evalueringen medvirket til at identificere reelle brugerbehov ved inddragelse af en slutbruger og potentielle behov ved inddragelse af UX-ekspert (107). Ligeledes beskriver Nielsen, at ved at evaluere en UI med kun en bruger har man allerede lært en tredjedel om brugervenligheden af designet (108). Derfor kan det vurderes, at denne prototype kunne øge anvendeligheden, idet evaluatorene fandt det nemmere at forholde sig til exacerbationsrisikoen efter forklaringerne.

Ud fra specialets metoder for evalueringen foreligger der en række elementer, der kunne have indflydelse evaluatorernes interaktion og vurdering af prototypen. En af de identificerede udfordringselementer er formulering af de foruddefinerede opgaver.

Under udformningen af opgaverne, var fokus på at stille spørgsmål, som kunne bidrage til at besvare, hvorvidt specialets prototypen imødekom de fire krav. Derfor blev opbygningen og formuleringen af opgaverne nedprioriteret. Dette kunne blandt andet ses under interaktion, hvor evaluatorerne ikke fandt det intuitivt, hvad der var hensigten med opgaven, der medvirkede til tøven. Her blev der stillet ledende spørgsmål for at kunne sikre evaluering af alle fire krav inden for evaluerings tidsramme. Et eksempel på et af de ledende spørgsmål indeholdt titlen på den lokale forklaring og den interaktive graf, som udgør en bias for evaluering (92). Dette skyldes at opgaverne bliver for nemme, så de medvirker til at forvrænge evaluatorernes syn på prototypen (92).

Et andet udfordrende element ved de foruddefinerede opgaver var deres centrale fokus på de fire krav. Dette kan have resulteret i manglende identificering for andre udfordringer. Dette vurderes dog ikke at have haft væsentlig indflydelse, grundet inddragelsen af en UX-eksperter i evalueringen. Her blev UX-eksperterens erfaringer bragt i spil, idet interaktionen blev udført som en think-aloud med mulighed for at afvige fra spørgsmålene under interviewet. Med andre ord blev der alligevel identificerede elementer, som kunne bidrage til en højne usability for prototypen. De udfordringer, der blev identificeret af UX-eksperter, var ikke noget specialet tidligere havde taget højde for ved udviklingen. Dette indikerer at der stadig foreligger behov for videre udvikling af prototypen, og en indikation på at der under vidensbasen stadig ikke er taget nok forbehold for forskellighed hos mennesker.

7.1.1 Imødekommelse af krav 1 og 2 - Global og lokal forklaring

For at imødekomme krav 1: *Global forklaring* (Billede 2) og krav 2: *Lokal forklaring* (Billede 4) valgte specialets forfattere at afvige fra visualisering forslagene fra vidensbasen, som anvendte bar-chart til både global og lokal forklaring. Dette skyldes vidensbasens beskrivelse af global forklaring som værende for svær at forstå, samt kun relevant i opstartsfasen. Ligeledes vurderer specialeforfatterne, at en visualisering af hvordan algoritmens input bliver vægtet generelt og ved en risikovurdering, kunne skabe forvirring. Derfor blev der valgt at udforme en forenklet global forklaring, indeholdende tekstbeskrivelse og en enkelt tabel over, hvordan risikoen skulle aflæses.

Denne afvigelse fra litteraturens beskrivelse af en krav 1: *Global forklaring*, medvirker til at besværliggøre en sammenligning af specialets resultater med andre studier. Den konklusion, at slutbrugeren fandt den globale forklaring mere anvendelig og forståelig sammenlignet med lokal forklaring, kan skyldes personlig præference for tal i stedet for

graf. Dog kan der ikke afvises, at det potentielt skyldes forklaringens design. Dette står i kontrast til litteraturen, som har vist, at sundhedsprofessionelle, særdeles sygeplejersker, foretrak den lokale forklaring (96,105,109). Dog er det ikke kun sygeplejerskerne, der foretrak lokal forklaring. I studiet af Diprose et al. blev rapporteret, at 62% af 149 læger ligeledes foretrak den lokale forklaring frem for den globale, da den var mere patientspecifik, krævede mindre kognitiv indsats og var i overensstemmelse med deres kliniske ræsonnement processer (106). Generelt anses en lokal forklaring også for at være mere pålidelig end global forklaring med hensyn til forklaringens sikkerhed, hvorfor den oftest vil være at foretrække ved validering af en prædiktion (109,110). Det er dog blevet påvist, at en global forklaring tidligere har vist sig være mere effektiv i identifikationen af falske positive og negative forudsigelser (106). Hermed kan det vurderes, at der potentiel vil være en fordel ved anvendelse af begge forklaringer. Denne vurdering understøttes af Radensky et al., der anbefaler, at både global og lokal forklaring kan være anvendelig i en højrisikos kontekst, hvor fejl og mangler kan have konsekvens for borgerens helbred (111). På nuværende tidspunkt foreligger der ikke klar indikation om, der ved telemonitoreringen er tale om en højrisiko kontekst, fordi det afhænger af hvordan der ageres på exacerbationsrisikoen i praksis (112).

7.1.3 Imødekommelse af krav 3 - Interaktiv graf

Krav 3: *Interaktiv graf* (Billede 4) blev udarbejdet på baggrund af informanterne fra miljøet, som gav udtryk for at anvende grafer ved vurdering af vitale parametre. Dette skyldes, at de havde svært ved at aflæse, om borgerens vitale parametre var steget eller faldet ved tabeloversigten. Interaktive visualisering er en af de grundlæggende tilgange til AI-forklaringer og et effektivt værktøj (60,93,101). Inkorporering af interaktiv trend graf ved siden af den lokale forklaring skulle understøtte TSA'erne i at vurdere grundlaget for exacerbationsrisikoen. Den interaktive graf kunne give TSA'erne kontrol over hvordan de analyserede data elementerne, og mulighed for at vælge enkelte datapunkter, eller udtrække flere patientdata (60). Trods af intentionerne, blev den interaktive graf beskrevet som svær at anvende med en alvorlighedsgrad på to. Her italesatte slutbrugeren at årsagen til besværet skyldes manglende erfaring med aflæsning af grafer. Slutbrugeren vurderede desuden, at de ikke ville have tid til at anvende grafen under hver monitorering, og foretrak derfor en tabeloversigt. I et studie af Brewer et al. blev det dog beskrevet, at sundhedspersonalet brugte mindre tid på at vurdere patientdata på en graf ift. en tabel (113). Selv påpegede de deltagende i studiet, at de foretrak grafen (113). Der bør også tages forbehold for, at algoritmen baserer sig på data fra virkeligheden, som kan være unøjagtige eller ufuldstændige. Derfor er det vigtigt med "sund fornuft", eller "klinisk intuition" (114). Baseret på disse faktorer bør den lokale forklaring og interaktive graf ikke tilsidesættes, trods slutbrugerens besvær.

7.3 Videre forskning

Ved evalueringen af prototypen blev der identificeret en række elementer, der bør undersøges nærmere gennem yderligere forskning. Ud fra den ovenstående diskussion, kan der identificeres to udfordringer, som er væsentlige hindringer for anvendelse af prototypen.

Første udfordring vedrører diskrepansen mellem evaluaterne og andre sundhedsprofessionelle. Denne diskrepans skaber usikkerhed og kan føre til uklarhed i udviklingsprocessen af XUI'en. Det er en udfordring at designe en XAI og XUI, da det er vigtigt at forstå, hvordan individuelle slutbrugere forventer, at en forklaring skal være (58). Derfor er der behov for inddragelse af flere slutbrugere forskellige brugere, for at få et mere tydeligt billede af deres behov for forklarlighed (59,104).

En anden udfordring vedrører usability-overtrædelserne, som udgør en hindring for XUI'ens effekt. Dette omfatter manglende forklaringer i krav 1: *Global forklaring*, hvor der skal forklares valg af parametre og evidens for algoritmens kliniske effekt. Ydermere vurderede slutbrugeren ikke krav 2: *Lokal forklaring* anvendeligt ved monitoreringen. Hvor litteraturen beskriver den lokale forklaring som et væsentligt redskab for at kunne evaluere AI-outputtet. Denne manglende nytte af den lokale forklaring og en ikke intuitiv placering af krav 3: *Interaktiv graf*, er væsentlige hindringer i prototypens design som skal løses, for at prototypen kan anvendes.

Ifølge Nielsen bør yderligere udviklinger af UI fokusere på at udføre flere mindre evalueringer med højst fem brugere per iteration (108). Der foreligger dog ikke tydeligt, om dette ligeledes vil være tilfældet ved en XUI. På baggrund af XUI'ens kompleksitet med forklarlighed kan det være nødvendigt med inddragelse af flere slutbrugere på grund af deres forskellige behov (59,104). Yderligere forskning med flere TSA'er som evaluater kan øge reliabiliteten ved den heuristiske evaluering, ved at minimere indflydelsen af evaluators personlige holdninger og præferencer på vurderingen (92). Der blev observeret, at slutbrugeren havde besvær ved anvendelse af den interaktive graf og den lokale forklaring, bør det undersøges nærmere om dette skyldes personlig præference eller grafernes design. Der kan ligeledes være behov for at give slutbrugerne vejledning i brugen af disse elementer for, for at kunne understøtte deres evaluering af exacerbationsrisikoen (60,109,110,115).

8. Konklusion

Formålet med dette speciale var at besvare følgende problemformulering:

“Hvordan kan en brugercentreret explainable user interface for telemonitorering medvirke til at øge anvendeligheden af prædiktiv algoritme?”

Specialets indledende analyse af kontekstuelle og evidensbaserede viden om behov for forståelse af algoritmer, viste at en brugercentreret explainable user-interface ville kunne øge anvendeligheden af en exacerbationsrisiko. Derfor blev der i specialet udviklet og evalueret en XUI bestående af fire krav, herunder en global forklaring, lokal forklaring, interaktiv graf samt en indikation for stigning/fald i exacerbationsrisikoen.

Evalueringsresultaterne kunne det bekræftes, at en brugercentreret explainable user interface kunne understøtte slutbrugerne i vurdering af risiko for exacerbationer. Her viste evalueringen at særdeles en forenklet global forklaring, der beskriver, hvordan algoritmen beregnes og aflæses, bidrog mest forståelse og blev derfor foretrukket. Hvorimod evidensen hævder, at en lokal forklaring, der forklarer årsagen til en specifik forudsigelse, bør være den forklaring, der bidrager mest til forståelse blandt sundhedsprofessionelle.

Under evalueringen blev der identificeret barriere for anvendelse af prototypen. Dette angår anvendelsen af den lokale forklaring og den interaktive graf, der blev vurderet ikke anvendelige ved monitorering af borgerne. Her skulle den interaktive graf bidrage til et retrospektivt overblik på borgernes målinger med exacerbationsrisikoen.

Den lokale forklaring og interaktiv graf blev dog adresseret af evidensen som væsentlige redskaber ved verificering af en algoritmes output, og bør derfor ikke undlades.

Baseret på diskrepansen mellem slutbrugerens udtalelser, og den eksisterende litteratur, som indikerer behovet for at inddrage flere slutbrugere, samt den manglende anvendelighed af den lokale forklaring og den interaktive graf, kan konklusionen drages, at prototypen ikke er tilstrækkeligt moden til at blive implementeret. Derfor er det nødvendigt at videreudvikle prototypen for at opnå en mere modenhed, der muliggør en succesfuld implementering.

9. Perspektivering

Før der kan overvejes om algoritmen og tilhørende XUI bør implementeres, er det af afgørende betydning at foretage en omfattende undersøgelse af XUI'en og algoritmens reelle indvirkning på arbejdsprocesserne og kliniske resultater. Dette vil kunne bidrage til en vurdering af den økonomiske bæredygtighed ved implementering af en prædiktiv algoritme og tilhørende XUI. Ligeledes bør der tages forbehold for, at XUI er et nyt koncept, der stadig er under udvikling, hvorfor der endnu ikke foreligger klare standarder eller retningslinjer for design og implementering. Udarbejdelsen af klare standarder og retningslinjer vurderes afgørende for at sikre, at XUI'en og algoritmen opfylder de nødvendige krav forinden en implementering. Dette vurderes at kunne øge reliabiliteten, validiteten og kvaliteten af en XUI, som vil danne grundlag for en vellykket implementering. Dette vil åbne muligheden for overførsel af XUI-designet som grundsten ved andre telemonitoreringskontekster såsom hjertesvigt.

10. Bibliografi

1. Global Initiative for Chronic Obstructive Lung Disease. 2023 GOLD Report [Internet]. 2023. Tilgængelig hos: https://goldcopd.org/wp-content/uploads/2023/03/GOLD-2023-ver-1.3-17Feb2023_WMV.pdf
2. Hvad er telemedicin [Internet]. Region Nordjylland. [henvist 31. maj 2023]. Tilgængelig hos: <https://rn.dk/sundhed/til-sundhedsfaglige-og-samarbejdspartnere/telecare-nord/hvad-er-telemedicin>
3. Digitaliseringsstyrelsen. Uddannelsesindsats i TeleCare Nord [Internet]. Tilgængelig hos: <https://digst.dk/media/12352/bilag-5-erfaring-med-uddannelse-og-kompetence.pdf>
4. Secher PH, Hangaard S, Kronborg T, Hæsum LKE, Udsen FW, Hejlesen O, m.fl. Clinical implementation of an algorithm for predicting exacerbations in patients with COPD in telemonitoring: a study protocol for a single-blinded randomized controlled trial. *Trials*. december 2022;23(1):356.
5. Sundheds- og ældreministeriet. Redegørelse om ældreområdet [Internet]. Tilgængelig hos: https://sum.dk/Media/5/A/8503_SUM_Aeldreomraadet_TILG.pdf
6. Jensen H, Thygesen L, Davidsen M. Sygdomsudviklingen i Danmark fremskrevet til 2030. KOL og type 2-diabetes. Bd. 2017. Statens Institut for Folkesundhed;
7. Baes-Jørgensen J, Kommunernes Landsforening. Lungesygdommen KOL rammer geografisk skævt [Internet]. 2020. Tilgængelig hos: <https://www.kl.dk/nyheder/momentum/2020/nr-4/lungesygdommen-kol-rammer-geografisk-skaevt/>
8. Afdeling for Sundhedsanalyser, redaktør. Prædiktionsmodeller for indlæggelse af borgere med KOL [Internet]. Sundhedsdatastyrelsen; 2015. Tilgængelig hos: <https://sundhedsdatastyrelsen.dk/-/media/sds/filer/find-tal-og-analyser/sygdomme-og-behandlinger/kol/praediktionsmodeller-kol.pdf>
9. Lungeforeningen. Danmark har kedelig EU-rekord i KOL-dødsfald – og har haft det i mange år [Internet]. Lungeforeningen; 2022. Tilgængelig hos: <https://www.lunge.dk/sites/default/files/Danmark%20har%20kedelig%20EU->

rekord%20i%20KOL-
d%C3%B8dsfald%20%E2%80%93%20og%20har%20haft%20det%20i%20ma-
nge%20%C3%A5r.pdf

10. Dansk Selskab for Almen Medicin. Luftvejsinfektioner - DSAM Vejledninger [Internet]. 2014. Tilgængelig hos: <https://vejledninger.dsam.dk/luftvejsinfektioner/?mode=visKapitel&cid=746>
11. Elbrønd J, Jakobsen M, Nielsen LM, Wilcke T. KOL - DSAM Vejledninger [Internet]. 1. udg. Komiteen for Sundhedsoplysning; 2017. Tilgængelig hos: https://vejledninger.dsam.dk/kol/?mode=visKapitel&cid=1034&gotoChapter=1034#chapter_1035
12. Jensen PM. Status på landsdækkende udbredelse af telemedicin [Internet]. Sundhedsdatastyrelsen; 2016. Tilgængelig hos: https://dsk.dk/wp-content/uploads/2020/07/Status-paa-landsdaekkende-udbredelse-af-telemedicin_30092020.pdf
13. Region Nordjylland. TeleCare Nord Afslutningsrapport - telemedicinsk storskalaforsøg i Nordjylland [Internet]. 2015. Tilgængelig hos: https://rn.dk/-/media/Rn_dk/Sundhed/Til-sundhedsfaglige-og-samarbejdspartnere/TelecareNord/Telemedicin-til-borgere-med-KOL/KOL-projekt/Evaluering-og-forskning/Fra-tro-til-viden---afrapportering/TeleCare-Nord-afslutningsrapport-18-11-2015-final.ashx?la=da
14. PA Consulting Group, redaktør. Business case for landsdækkende udbredelse af telemedicinsk hjemmemonitorering til borgere med KOL. [Internet]. 2017. Tilgængelig hos: https://sundhedsdatastyrelsen.dk/-/media/sds/filer/strategi-og-projekter/telemedicin/forudsætningsprojekter_kol/genberegning_telemedicin_business_case.pdf?la=da
15. Digitaliseringsstyrelsen. Signaturprojekter med kunstig intelligens i kommuner og regioner [Internet]. Tilgængelig hos: <https://digst.dk/digital-transformation/signaturprojekter/>

16. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 29. oktober 2019;17(1):195.
17. Directorate-General for Communications Networks C and T (European C, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji. Ethics guidelines for trustworthy AI [Internet]. LU: Publications Office of the European Union; 2019. Tilgængelig hos: <https://data.europa.eu/doi/10.2759/346720>
18. Sundhedsstyrelsen, redaktør. Patientvejledning om KOL [Internet]. 2006. Tilgængelig hos: <https://www.sst.dk/~media/CE16657A4CF04EEB8C6DFF3CD6AA46B6.ashx>
19. BMJ Best Practice. Chronic obstructive pulmonary disease (COPD) - Criteria | BMJ Best Practice US [Internet]. Tilgængelig hos: <https://bestpractice.bmj.com/topics/en-us/7/criteria>
20. Seemungal T a. R, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time Course and Recovery of Exacerbations in Patients with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. maj 2000;161(5):1608–13.
21. Rothnie KJ, Müllerová H, Smeeth L, Quint JK. Natural History of Chronic Obstructive Pulmonary Disease Exacerbations in a General Practice–based Population with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 15. august 2018;198(4):464–71.
22. Kim V, Aaron SD. What is a COPD exacerbation? Current definitions, pitfalls, challenges and opportunities for improvement. Eur Respir J [Internet]. 1. november 2018;52(5). Tilgængelig hos: <https://erj.ersjournals.com/content/52/5/1801261>
23. Bilde L, Svenning AR. Omkostninger ved behandling af patienter med kronisk obstruktiv lungesygdom (KOL) [Internet]. DSI Institut for Sundhedsvæsen; 2004. Tilgængelig hos: <https://www.vive.dk/media/pure/9581/2057590?fbclid=IwAR1CbUxOVxEAI07aCAh1xMc4ZOYjDY-MM4RNjyTNYInUmEp6qNzzzkBAhAY>

24. Larsson K, Janson C, Ställberg B, Lisspers K, Olsson P, Kostikas K, m.fl. Impact of COPD diagnosis timing on clinical and economic outcomes: the ARCTIC observational cohort study. Int J Chron Obstruct Pulmon Dis. 13. maj 2019;14:995–1008.
25. Digitaliseringsstyrelsen. Telemedicine - A key to health services of the future [Internet]. 2014. Tilgængelig hos: https://en.digst.dk/media/14145/telemedicine_uk_pdfa_03_11_12.pdf
26. Danske Regioner, Kommunernes Landsforening, Ministeriet for Sundhed og Forebyggelse, Social- og Integrationsministeriet, Erhvervs- og Vækstministeriet, Økonomi- og Indenrigsministeriet, m.fl. National Handlingsplan for udbredelse af Telemedicin [Internet]. 2012. Tilgængelig hos: https://medcom.dk/media/4852/handlingsplan_telemedicin.pdf
27. Lilholt PH, Udsen FW, Ehlers L, Hejlesen OK. Telehealthcare for patients suffering from chronic obstructive pulmonary disease: effects on health-related quality of life: results from the Danish 'TeleCare North' cluster-randomised trial. BMJ Open. 1. maj 2017;7(5):e014587.
28. Region Nordjylland. Opfølgning på data i kommuner [Internet]. 2022. Tilgængelig hos: https://rn.dk/-/media/Rn_dk/Sundhed/Til-sundhedsfaglige-og-samarbejdspartnere/TelecareNord/Telemedicin-til-borgere-med-KOL/Instrukser-KOL/Kommune/KOL-Kommune-Opfoelgning-paa-data-2023-01-30.ashx?la=da
29. Almen praksis, Nord-KAP, Lungemedicinsk Afdeling på Aalborg Universitetshospital, KOL TeleCare Nord. KOL Foruddefinerede alarmgrænser [Internet]. KOL TeleCare Nord; 2022. Tilgængelig hos: https://rn.dk/-/media/Rn_dk/Sundhed/Til-sundhedsfaglige-og-samarbejdspartnere/TelecareNord/Telemedicin-til-borgere-med-KOL/Instrukser-KOL/F%C3%A6lles-instrukser/Almen-praksis-kommune-og-hospital/KOL-Foruddefinerede-alarmgraenser-2022-08-26.ashx?la=da
30. Sundhedsstyrelsen. Telemedicinsk understøttelse af behandlingstilbud til mennesker med KOL [Internet]. 2017. Tilgængelig hos: <https://www.sst.dk/-/media/Udgivelser/2017/Telemedicinsk-underst%C3%B8ttelse-af->

behandlingstilbud-til-mennesker-med-
KOL.ashx?sc_lang=da&hash=C078B97D04E3C71D12C4C0C487E3BE98

31. Christensen JKB. Does Telecare Improve Interorganisational Collaboration? *Int J Integr Care*. 2016;16(4):14.
32. Walsh T, Beatty PCW. Human factors error and patient monitoring. *Physiol Meas*. 1. august 2002;23(3):R111–32.
33. Fernandes JG. Artificial Intelligence in Telemedicine. I: Lidströmer N, Ashrafian H, redaktører. *Artificial Intelligence in Medicine [Internet]*. Cham: Springer International Publishing; 2020. s. 1–10. Tilgængelig hos: https://doi.org/10.1007/978-3-030-58080-3_93-1
34. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, m.fl. Role of Artificial Intelligence within the Telehealth Domain. *Yearb Med Inform*. august 2019;28(1):35–40.
35. Bivard A, Churilov L, Parsons M. Artificial intelligence for decision support in acute stroke — current roles and potential. *Nat Rev Neurol*. oktober 2020;16(10):575–85.
36. Exarchos KP, Kostikas K. Artificial intelligence in COPD: Possible applications and future prospects. *Respirology*. 2021;26(7):641–2.
37. Kronborg T, Hangaard S, Cichosz SL, Hejlesen O. A two-layer probabilistic model to predict COPD exacerbations for patients in telehealth. *Comput Biol Med*. 1. januar 2021;128:104108.
38. Bender C, Hangaard S, Kronborg T, Hejlesen O, Secher PH. Preliminary Qualitative Evaluation of Patient-Related Perspectives Related to the Implementation of a Predictive Algorithm in a Telehealth System for COPD. 2021. Tilgængelig hos: https://vbn.aau.dk/ws/portalfiles/portal/413365952/SHTI_281_SHTI210230.pdf
39. Dahlin E. Mind the gap! On the future of AI research. *Humanit Soc Sci Commun*. 15. marts 2021;8(1):1–4.

40. Carusi A, Winter PD, Armstrong I, Ciravegna F, Kiely DG, Lawrie A, m.fl. Medical artificial intelligence is as much social as it is technological. *Nat Mach Intell.* februar 2023;5(2):98–100.
41. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artif Intell.* 1. februar 2019;267:1–38.
42. Sujan M, Pool R, Salmon P. Eight human factors and ergonomics principles for healthcare artificial intelligence. *BMJ Health Care Inform.* 4. februar 2022;29(1):e100516.
43. Zhang AY, Lam SSW, Ong MEH, Tang PH, Chan LL. Explainable AI: Classification of MRI Brain Scans Orders for Quality Improvement. I: *Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies [Internet]*. New York, NY, USA: Association for Computing Machinery; 2019. s. 95–102. (BDCAT '19). Tilgængelig hos: <https://dl.acm.org/doi/10.1145/3365109.3368791>
44. Ribeiro M, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. I 2016. Tilgængelig hos: <https://dl-acm-org.zorac.aub.aau.dk/doi/10.1145/2939672.2939778>
45. Chaddad A, Peng J, Xu J, Bouridane A. Survey of Explainable AI Techniques in Healthcare. *Sensors.* januar 2023;23(2):634.
46. Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning [Internet]. *arXiv*; 2017. Tilgængelig hos: <http://arxiv.org/abs/1702.08608>
47. Chromik M, Butz A. Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. I: *Ardito C, Lanzilotti R, Malizia A, Petrie H, Piccinno A, Desolda G, m.fl., redaktører. Human-Computer Interaction – INTERACT 2021.* Cham: Springer International Publishing; 2021. s. 619–40.
48. Krishnan M. Against Interpretability: a Critical Examination of the Interpretability Problem in Machine Learning. *Philos Technol.* 1. september 2020;33(3):487–502.

49. Páez A. The Pragmatic Turn in Explainable Artificial Intelligence (XAI). *Minds Mach.* september 2019;29(3):441–59.
50. Lipton ZC. The Mythos of Model Interpretability [Internet]. arXiv; 2017. Tilgængelig hos: <http://arxiv.org/abs/1606.03490>
51. Rai A. Explainable AI: from black box to glass box. *J Acad Mark Sci.* 1. januar 2020;48(1):137–41.
52. Loyola-González O. Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. *IEEE Access.* 2019;7:154096–113.
53. Raschka S. What is the relation between Logistic Regression and Neural Networks and when to use which? [Internet]. Sebastian Raschka, PhD. Tilgængelig hos: <https://sebastianraschka.com/faq/docs/logisticregr-neuralnet.html>
54. Liao QV, Varshney KR. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences [Internet]. arXiv; 2022. Tilgængelig hos: <http://arxiv.org/abs/2110.10790>
55. Gunning D, Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Mag.* 24. juni 2019;40(2):44–58.
56. Manresa-Yee C, Roig-Maimó MF, Ramis S, Mas-Sansó R. Advances in XAI: Explanation Interfaces in Healthcare. I: Lim CP, Chen YW, Vaidya A, Mahorkar C, Jain LC, redaktører. *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects* [Internet]. Cham: Springer International Publishing; 2022. s. 357–69. (Intelligent Systems Reference Library). Tilgængelig hos: https://doi.org/10.1007/978-3-030-83620-7_15
57. Fan FL, Xiong J, Li M, Wang G. On Interpretability of Artificial Neural Networks: A Survey. *IEEE Trans Radiat Plasma Med Sci.* november 2021;5(6):741–60.
58. Dey S, Chakraborty P, Kwon BC, Dhurandhar A, Ghalwash M, Suarez Saiz FJ, m.fl. Human-centered explainability for life sciences, healthcare, and medical

informatics. Patterns [Internet]. 13. maj 2022;3(5). Tilgængelig hos:
<https://www.sciencedirect.com/science/article/pii/S2666389922000782>

59. Liao V, Pribić M, Han J, Miller S, Sow D. Question-Driven Design Process for Explainable AI User Experiences. 2021.
60. Barda AJ, Horvat CM, Hochheiser H. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. BMC Med Inform Decis Mak. 8. oktober 2020;20(1):257.
61. Haque AB, Islam AKMN, Mikalef P. Explainable Artificial Intelligence (XAI) from a user perspective- A synthesis of prior literature and problematizing avenues for future research. Technol Forecast Soc Change. januar 2023;186:122120.
62. Sander MD, Madsen R. Visualisering af komplekse resultater til monitorering af patienter med KOL. Aalborg Universitet; 2022. 56 s.
63. Egholm L. Videnskabsteori : perspektiver på organisationer og samfund. 1. udgave. Havemann T, redaktør. Kbh.: Hans Reitzel; 2014. 237 s.
64. Mithun AhamedM, Mithun AhamedM, Yafooz WaelMS. Extended User Centered Design (UCD) Process in the Aspect of Human Computer Interaction. I: 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE). 2018. s. 1–6.
65. Li X, Xu Y. Role of Human-Computer Interaction Healthcare System in the Teaching of Physiology and Medicine. Comput Intell Neurosci. 13. april 2022;2022:e5849736.
66. Doroftei D, De Cubber G, Wagemans R, Matos A, Silva E, Lobo V, m.fl. User-Centered Design. I 2017.
67. Rogers Y. HCI Theory: Classical, Modern, and Contemporary. Synth Lect Hum-Centered Inform. 23. maj 2012;5(2):1–129.
68. Hartson R. The Ux Book : Process and Guidelines for Ensuring a Quality User Experience. Bd. 2012. Morgan Kaufmann Pub; 2012. xxxiv, 937 p. s.

69. Guo FY. Not Just Usability – The Four Elements of User Experience [Internet]. UX Strategized; 2012. Tilgængelig hos:
http://uxstrategized.com/White_Paper_Four_Elements_of_User_Experience.pdf?fbclid=IwAR0IwXULQ8ZpF5ceiUNd1wMikdo5R8nOLI6Tj7PkjmT61t_dbrJDP6c6oLY
70. Hevner A, R A, March S, T S, Park, Park J, m.fl. Design Science in Information Systems Research. *Manag Inf Syst Q.* 1. marts 2004;28:75 s.
71. CINAHL Database | EBSCO [Internet]. Tilgængelig hos:
<https://www.ebsco.com/products/research-databases/cinahl-database>
72. Embase Coverage and Content [Internet]. Tilgængelig hos:
<https://www.elsevier.com/solutions/embase-biomedical-research/coverage-and-content>
73. About Pubmed [Internet]. PubMed. 2022. Tilgængelig hos:
<https://pubmed.ncbi.nlm.nih.gov/about/>
74. What is the complete list of Scopus Subject Areas and All Science Journal Classification Codes (ASJC)? - Scopus: Access and use Support Center [Internet]. Elsevier. 2023. Tilgængelig hos:
https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/
75. Islam MR, Ahmed MU, Barua S, Begum S. A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks. *Appl Sci.* januar 2022;12(3):1353.
76. Web of Science Research Areas [Internet]. Clarivate InCites Help. 2021. Tilgængelig hos: <http://incites.help.clarivate.com/Content/Research-Areas/wos-research-areas.htm>
77. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. 2007. Tilgængelig hos: <https://faseb.onlinelibrary.wiley.com/doi/10.1096/fj.07-9492LSF>

78. Lund H, Juhl CB, Andreasen J, Møller AM. Håndbog i litteratursøgning og kritisk læsning: redskaber til evidensbaseret praksis [Internet]. 1. udg. Munksgaard; 2014. Tilgængelig hos: <http://munksgaard.dk/Medicin/H%C3%A5ndbog-i-litteraturs%C3%B8gning-og-kritisk-l%C3%A6sning/9788762811966>
79. Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. *Artif Intell Rev*. 1. juni 2022;55(5):3503–68.
80. Clement T, Kemmerzell N, Abdelaal M, Amberg M. A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development Process. 2023. Tilgængelig hos: <https://www.mdpi.com/2504-4990/5/1/6/htm#sec2dot1-make-05-00006>
81. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 5. december 2016;5:210.
82. Jones E. Systematic Reviews: Step 6: Assess Quality of Included Studies [Internet]. UNC University libraries. 2023. Tilgængelig hos: <https://guides.lib.unc.edu/systematic-reviews/assess-quality>
83. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol*. 10. juli 2008;8(1):45.
84. Morgan DL, Ataie J, Carder P, Hoffman K. Introducing Dyadic Interviews as a Method for Collecting Qualitative Data. *Qual Health Res*. 1. september 2013;23(9):1276–84.
85. Kvale S, Brinkmann S. Interview; det kvalitative forskningsinterview som håndværk. 3. udgave. Bd. 2015. Hans Reitzel;
86. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 1. januar 2006;3:77–101.
87. About Figma, the collaborative interface design tool. [Internet]. Figma. Tilgængelig hos: <https://www.figma.com/about/>
88. Nielsen J, Molich R. Heuristic evaluation of user interfaces. I: Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people

- CHI '90 [Internet]. Seattle, Washington, United States: ACM Press; 1990. s. 249–56. Tilgængelig hos: <http://portal.acm.org/citation.cfm?doid=97243.97281>
89. Melissa DeJonckheere, Lisa M Vaughn. Semistructured interviewing in primary care research: a balance of relationship and rigour. *Fam Med Community Health*. 1. marts 2019;7(2):e000057.
90. Gou F. More Than Usability: The Four Elements of User Experience, Part I [Internet]. Uxmatters. 2012. Tilgængelig hos: <https://www.uxmatters.com/mt/archives/2012/04/more-than-usability-the-four-elements-of-user-experience-part-i.php>
91. Frimodig B. Heuristics: Definition, Examples, and How They Work [Internet]. 2023 [henvist 29. maj 2023]. Tilgængelig hos: <https://www.simplypsychology.org/what-is-a-heuristic.html>
92. Nielsen J. Usability Engineering. Bd. 1993. Morgan Kaufmann Publishers Inc;
93. Wang Q, Huang K, Chandak P, Zitnik M, Gehlenborg N. Extending the Nested Model for User-Centric XAI: A Design Study on GNN-based Drug Repurposing. *IEEE Trans Vis Comput Graph*. januar 2023;29(1):1266–76.
94. Antoniadi AM, Galvin M, Heverin M, Wei L, Hardiman O, Mooney C. A Clinical Decision Support System for the Prediction of Quality of Life in ALS. *J Pers Med*. 2022;12(3).
95. Müller J, Stoeck M, Oeser A, Gaebel J, Streit M, Dietz A, m.fl. A visual approach to explainable computerized clinical decision support. *Comput Graph*. 1. oktober 2020;91:1–11.
96. Di Martino F, Delmastro F. Explainable AI for clinical and remote health applications: a survey on tabular and time series data. *Artif Intell Rev*. 1. juni 2023;56(6):5261–315.
97. Ooge J, Stiglic G, Verbert K. Explaining artificial intelligence with visual analytics in healthcare. *WIREs Data Min Knowl Discov*. 1. januar 2022;12(1):e1427.

98. Naiseh M, Al-Thani D, Jiang N, Ali R. How the different explanation classes impact trust calibration: The case of clinical decision support systems. *Int J Hum-Comput Stud.* 169:102941.
99. Schoonderwoerd TAJ, Jorritsma W, Neerincx MA, van den Bosch K. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *Int J Hum-Comput Stud.* 1. oktober 2021;154:102684.
100. Liao QV, Gruen D, Miller S. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. I: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* [Internet]. 2020. s. 1–15. Tilgængelig hos: <http://arxiv.org/abs/2001.02478>
101. Abdullah TAA, Zahid MS, Ali W. A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions. *Symmetry.* 2021;13(12).
102. David DB, Resheff YS, Tron T. Explainable AI and Adoption of Financial Algorithmic Advisors: an Experimental Study [Internet]. *arXiv*; 2021. Tilgængelig hos: <http://arxiv.org/abs/2101.02555>
103. Kelly J, Sadeghieh T, Adeli K. Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. *EJIFCC.* 24. oktober 2014;25(3):227–43.
104. Chazette L, Klünder J, Schneider K, Balci M. How Can We Develop Explainable Systems? Insights from a Literature Review and an Interview Study. I *Pittsburgh PA USA*; s. 1–12. Tilgængelig hos: <https://dl.acm.org/doi/epdf/10.1145/3529320.3529321>
105. Pawar U, O'Shea D, Rea S, O'Reilly R. Incorporating Explainable Artificial Intelligence (XAI) to aid Understanding of Machine Learning in the Healthcare Domain. 2020.
106. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *J Am Med Inform Assoc JAMIA.* 27. februar 2020;27(4):592–600.

107. Harley A. UX Expert Reviews [Internet]. Nielsen Norman Group. 2018. Tilgængelig hos: <https://www.nngroup.com/articles/ux-expert-reviews/>
108. Nielsen J. Why You Only Need to Test with 5 Users [Internet]. Nielsen Norman Group. 2000. Tilgængelig hos: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
109. Guidotti R, Monreale A, Ruggieri S, Pedreschi D, Turini F, Giannotti F. Local Rule-Based Explanations of Black Box Decision Systems [Internet]. arXiv; 2018. Tilgængelig hos: <http://arxiv.org/abs/1805.10820>
110. Machlev R, Heistrene L, Perl M, Levy KY, Belikov J, Mannor S, m.fl. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy AI. 1. august 2022;9:100169.
111. Radensky M, Downey D, Lo K, Popović Z, Weld DS. Exploring The Role of Local and Global Explanations in Recommender Systems [Internet]. arXiv; 2021. Tilgængelig hos: <http://arxiv.org/abs/2109.13301>
112. Europa-kommissionen. Forslag til Europa-Parlamentets og Rådets forordning om harmoniserede regler for kunstig intelligens (Retsakten om kunstig intelligens) og om ændring af visse af unionens lovgivningsmæssige retsakter [Internet]. 2021. Tilgængelig hos: <https://eur-lex.europa.eu/legal-content/DA/TXT/?uri=CELEX%3A52021PC0206>
113. Brewer NT, Gilkey MB, Lillie SE, Hesse BW, Sheridan SL. Tables or Bar Graphs? Presenting Test Results in Electronic Medical Records. Med Decis Mak Int J Soc Med Decis Mak. 2012;32(4):545–53.
114. Bharati S, Mondal MRH, Podder P. A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When? IEEE Trans Artif Intell. 2023;1–15.
115. Agency for Healthcare Research and Quality,. Displaying the Data in a Health Care Quality Report [Internet]. 2019. Tilgængelig hos: <https://www.ahrq.gov/talkingquality/translate/display/index.html>

11. Bilag

Bilag 1 - Tematisering af vidensbasens inkluderede artikler	2
Bilag 2 - Brinkmann og Kvaales syv faser	33
Bilag 3 - Interviewguide for miljøet	36
Bilag 3.1 - Deltagerinformation	40
Bilag 3.2 - Samtykkeerklæring	42
Bilag 4 - Transskribering af miljøet (interview)	43
Bilag 5 - Tematisk analyse af miljøet (Interview)	59
Bilag 6 - Eksplorativ gennemgang af OTH og Telma	70
Bilag 7 - Evalueringsprotokol	77
Bilag 7.1 - Deltagerinformation	82
Bilag 7.2 - Samtykkeerklæring	85
Bilag 7.3 - Evaluering af prototype del 1: Opgaver og think-aloud	86
Bilag 7.4 - Evaluering af prototype Del 1: Opfølgende spørgsmål	91
Bilag 8 - Case for kontekstforståelse til UX-ekspert i evaluering	94
Bilag 9 - Transskribering af evaluering	95
Bilag 10 - Gennemgang af heuristikker til analysering	138
Bilag 11 - Dataekstrahering af evaluering	142
Bilag 12 - Tematisk analyse af udtalelser fra evaluering	145
Bilag 13 - Prototype udvikling	161
Bilag 13.1 - Low-fidelity prototype	161
Bilag 13.2 - High-fidelity prototype	166