UDVIKLING AF EN EXPLAINABLE MACHINE LEARNING MODEL TIL DETEKTION AF PRÆDIABETES

Kandidatspeciale

Klinisk Videnskab & Teknologi, Aalborg Universitet
4. semester

Gruppe: 10508
Andrés Rose Millan Sørensen
Michelle Vittrup Sørensen

Vejledere:
Stine Hangaard
Tanja Fredensborg Holm

2. juni 2025

Titel:
Udvikling af en explainable machine learning model til detektion af prædiabetes

Uddannelse:
Klinisk Videnskab & Teknologi

Modultitel:
Kandidatspeciale

Projektperiode:
Forår 2025

Projektgruppe:
10508

Forfattere:
Andrés Rose Millan Sørensen
Michelle Vittrup Sørensen

Hovedvejleder:
Stine Hangaard

Bivejleder:
Tanja Fredensborg Holm

Antal anslag (med mellemrum):
78.874
Svarende til 32,9 standardsider

Antal bilag:
11

Referencestandard:
Vancouver

Afleveringsdato:
02.06.2025

Resumé

Baggrund:
Prædiabetes er en asymptomatisk, men klinisk relevant tilstand, som ofte forbliver udiagnosticeret. Selvom flere screeningsværktøjer baseret på selvrapporterede data eksisterer, er de mest anvendte og validerede tests i klinisk praksis afhængige af laboratoriemålinger som HbA1c, FPG og OGTT.

Formål:
Dette speciale havde til formål at udvikle og evaluere explainable machine learning (ML) modeller til detektion af prædiabetes udelukkende baseret på ikke-laboratorieafhængige risikofaktorer.

Metode:
Der blev anvendt data fra NHANES 2011–2016 til at udvikle tre separate Decision Tree-modeller, hvor prædiabetes blev defineret ud fra én laboratorieparameter (HbA1c, FPG eller OGTT). Udviklingen omfattede præprocessering og modellering. Input omfattede 17 variable relateret til antropometri, demografi og livsstil.

Resultater:
Modellerne opnåede høj sensitivitet (recall: 0.670–0.918), men lav diskriminationsevne (AUC: 0.588–0.666). HbA1c- og OGTT-modellerne anvendte kun én inputvariabel (alder), mens FPG-modellen inkluderede to (talje-højde-ratio og køn).

Konklusion:
Explainable ML kan anvendes til at identificere personer med prædiabetes baseret på ikke-laboratorieafhængige data. Modeller med høj sensitivitet kan være relevante i screeningssammenhæng, men den lave AUC understreger behovet for forbedret datagrundlag og metode, før bred implementering kan anbefales.

Title:
Development of an Explainable Machine Learning Model for the Detection of Prediabetes

Course of Study:
Clinical Science & Technology

Module theme:
Master's Thesis

Project period:
Spring 2025

Project group:
10508

Authors:
Andrés Rose Millan Sørensen
Michelle Vittrup Sørensen

Supervisor:
Stine Hangaard

Co-supervisor:
Tanja Fredensborg Holm

Total number of characters (including spaces):
78.874
Equivalent to 32,9 standard pages

Number of appendices:
11

Reference standard:
Vancouver

Submission date:
02.06.2025

Abstract

Background:
Prediabetes is an asymptomatic but clinically relevant condition that often remains undiagnosed. Although several screening tools based on self-reported data exist, the most commonly used and validated tests in clinical practice rely on laboratory measurements such as HbA1c, FPG, and OGTT.

Aim:
This thesis aimed to develop and evaluate explainable machine learning (ML) models for detecting prediabetes based solely on non-laboratory-dependent risk factors.

Method:
Data from NHANES 2011–2016 were used to develop three separate Decision Tree models, where prediabetes was defined based on a single laboratory parameter (HbA1c, FPG, or OGTT). The development included preprocessing and modeling. Inputs included 17 variables related to anthropometry, demographics, and lifestyle.

Results:
The models achieved high sensitivity (recall: 0.670–0.918) but low discriminatory power (AUC: 0.588–0.666). The HbA1c and OGTT models used only one input variable (age), while the FPG model included two (waist-to-height ratio and gender).

Conclusion:
Explainable ML can be used to identify individuals with prediabetes based on non-laboratory-dependent data. Models with high sensitivity may be relevant in a screening context, but the low AUC highlights the need for improved data quality and methodology before broad implementation can be recommended.

Forord

Dette speciale er udarbejdet som afslutning på kandidatuddannelsen i Klinisk Videnskab og Teknologi ved Aalborg Universitet. Specialet er udarbejdet i samarbejde mellem to studerende og bygger på en fælles interesse for forebyggelse og tidlig opsporing af diabetes.

I forbindelse med specialet blev der gennemført en systematisk litteraturgennemgang med henblik på at identificere eksisterende machine learning modeller til detektion af prædiabetes. Resultaterne heraf danner grundlag for både problemanalysen (afsnit 1.5) og dele af diskussionerne. Den fulde søgestrategi og protokol er vedlagt som bilag 1.

Vi vil gerne takke vores hovedvejleder, Stine Hangaard, samt bivejleder, Tanja Fredensborg Holm, for faglig sparring og værdifuld feedback undervejs i specialet.

Indholdsfortegnelse

1. Problemanalyse

1.1 Type 2 diabetes

I 2021 levede cirka 537 millioner voksne i alderen 20-79 år med diabetes på globalt plan. Ifølge prognoser fra International Diabetes Federation forventes dette tal at stige til 643 millioner i 2030 og yderligere til 783 millioner i 2045, hvilket svarer til, at én ud af otte voksne vil have sygdommen, en stigning på 46% ift. 2021. Over 90% af alle tilfældene er type 2 diabetes (T2D), hvilket gør denne form til den mest udbredte diabetestype (1). I Danmark er diabetes en af de kroniske sygdomme, der rammer flest danskere. Antallet af personer med T2D i Danmark er mere end firdoblet siden 1996 (2,3) og prævalensen er fortsat stærkt stigende. Det estimeres, at der i 2030 vil være cirka 420.000 danskere med T2D (4). T2D er en progressiv sygdom, hvor både insulinsekretion samt insulinresistens forværres over tid. Dette resulterer i hyperglykæmi, som med tiden vil forårsage cellulære og molekylære forandringer, der bidrager til sygdommens progression samt udvikling af alvorlige komplikationer og følgesygdomme såsom kardiovaskulære sygdomme samt mikrovaskulære komplikationer som neuro-, nefro- og retinopati (5).

Diabetes koster årligt det danske samfund omkring 32 milliarder kroner, som anvendes til behandling, pleje og tabt arbejdsfunktion (6). Omkostninger har derfor konsekvenser for samfundet, men der er også konsekvenser for individet i form af udvikling af senkomplikationer, hvilket understreger nødvendigheden for at forebygge udvikling af T2D (7).

1.2 Prædiabetes

Prædiabetes er en intermediær tilstand, der ofte går forud for udviklingen af T2D, og som er karakteriseret ved forhøjede plasmaglukoseniveauer, der ligger over det normale, men endnu ikke opfylder de diagnostiske kriterier for diabetes (8). Prædiabetes er forbundet med samtidig tilstedeværelse af insulinresistens og betacelledysfunktion, hvilket er to patologiske processer, der opstår, før ændringer i glukoseniveauer er målbare (9). Tilstanden betegnes i klinisk praksis som enten nedsat glukosetolerance (Impaired Glucose Tolerance, IGT), forhøjet fasteglukose (Impaired Fasting Glucose, IFG) eller forhøjet langtidsglukose i form af glykosyleret hæmoglobin (HbA1c) afhængigt af hvilken diagnostisk test, der ligger til grund for påvisningen (7,10), se tabel 1 for oversigt over begreberne. En af de store udfordringer med prædiabetes er, at mange personer er asymptomatiske i lang tid og derfor ikke er klar over deres tilstand. Det betyder, at de ofte ikke modtager rettidig intervention (11).

Begreb	Beskrivelse	Diagnostisk test
Impaired Fasting Glucose (IFG)	Forhøjet plasmaglukose i fastende tilstand, men under diabetesniveau.	Fastende plasmaglukose (FPG)
Impaired Glucose Tolerance (IGT)	Dysreguleret glukosestofskifte, hvor plasmaglukoseniveauet er forhøjet efter et måltid eller glukoseindtag, men under diabetesniveau.	2-timers plasmaglukose efter en Oral Glucose Tolerance Test (OGTT)
Forhøjet HbA1c	Øget andel af glykosyleret hæmoglobin, som afspejler kronisk forhøjet plasmaglukoseniveau over tid, men under diabetesniveau.	Glykosyleret hæmoglobin (HbA1c)

Tabel 1: *Kort beskrivelse af de centrale prædiabetesbegreber og dertilhørende diagnostiske tests.*

1.2.1 Definition af prædiabetes

Prædiabetes defineres ud fra en række kriterier, der varierer mellem organisationer og er baseret på forskellige diagnostiske tests: HbA1c, fastende plasmaglukose (FPG) og oral glukosetolerancetest (OGTT) (7), hvilket kan ses i tabel 2.

Test	Enhed	ADA	WHO	Diabetes Canada
HbA1c	%	5.7-6.4	Ikke anvendt	6.0-6.4
	mmol/mol	39-46	Ikke anvendt	42-47
FPG	mg/dL	100-125	110-125	110-125
	mmol/L	5.6-6.9	6.1-6.9	6.1-6.9
OGTT	mg/dL	140-199	140-199	140-199
	mmol/L	7.8-11.0	7.8-11.0	7.8-11.0

Tabel 2: *Diagnostiske kriterier for prædiabetes anvendt af forskellige organisationer. ADA = American Diabetes Association(10), WHO = World Health Organization(12), Diabetes Canada(13).*

Disse tre tests identificerer forskellige undergrupper, da de måler forskellige aspekter af glukoseregulationen og fanger forskellige underliggende fysiologiske tilstande (9,14). Eksempelvis er IFG forbundet med insulinresistens i leveren, mens IGT er relateret til nedsat glukoseoptag i muskelvæv (15). Der er generelt lavt overlap mellem tests. Et amerikansk studie viste fx, at kun 4.1% opfyldte kriterierne for prædiabetes på alle tre tests (16), og at OGTT identificerede flest tilfælde (17). Det understreger, at valget af definition kan have betydning for, hvem der klassificeres med prædiabetes. Et systematisk review har desuden vist, at testenes evne til at identificere prædiabetes varierer betydeligt på tværs af populationer, og at der ikke er enighed om hvilke grænseværdier, der bør anvendes (14). Det medfører, at forskellige tests ofte identificerer forskellige grupper med begrænset overlap, hvilket understøtter, at prædiabetes kan være en heterogen og diagnostisk udfordrende tilstand.

1.2.2 Prævalens af prædiabetes

I 2021 blev det estimeret, at 10.6% af den voksne verdensbefolkning havde nedsat IGT, en andel der forventes at stige til 11.4% i 2045 svarende til 730 millioner mennesker. Derudover skønnes det, at 6.2% af den voksne befolkning havde forhøjet IFG i 2021 med en forventet stigning til 6.9% i 2045 (7). Den globale stigning i prædiabetes ses særlig høj i USA, hvor mange voksne er berørt. Ifølge Centers for

Disease Control and Prevention (CDC) viste data fra 2009-2012, at 37% af voksne over 20 år havde prædiabetes, mens forekomsten blandt personer over 65 år var helt oppe på 51% baseret på målinger af FPG og HbA1c (18). I Danmark har tidligere estimater for prævalensen af prædiabetes været baseret på OGTT, hvor diagnosen blev stillet ud fra nedsat IGT og forhøjet IFG. På baggrund af disse kriterier blev det anslået, at omkring 750.000 danskere havde prædiabetes. Med overgangen til HbA1c som primær diagnostisk metode, hvor prædiabetes defineres ved en HbA1c-værdi mellem 42-47 mmol/mol, er estimaterne imidlertid markant lavere. Nyere beregninger baseret på denne metode anslår, at 292.715 danskere havde prædiabetes i 2011 (19). Ifølge Diabetesforeningen blev det i 2023 skønnet, at næsten 500.000 danskere havde prædiabetes (20). Det fremgår dog ikke, hvilke diagnostiske kriterier der ligger til grund for dette estimat.

1.2.3 Progression til T2D

Tidlig intervention er afgørende, da personer med prædiabetes allerede har øget risiko for kardiovaskulære sygdomme og progression til T2D (11). Screening og tidlig opsporing kan derfor både forebygge sygdomsprogression og reducere risikoen for senkomplikationer. At identificere personer med prædiabetes er således centralt både i et folkesundhedsmæssigt og klinisk perspektiv, da tidlig indsats kan give store sundhedsmæssige (16) såvel som socioøkonomiske gevinster (21). Dette understreger behovet for effektive og tilgængelige metoder til tidlig identifikation af prædiabetes også uden laboratoriebaserede målinger.

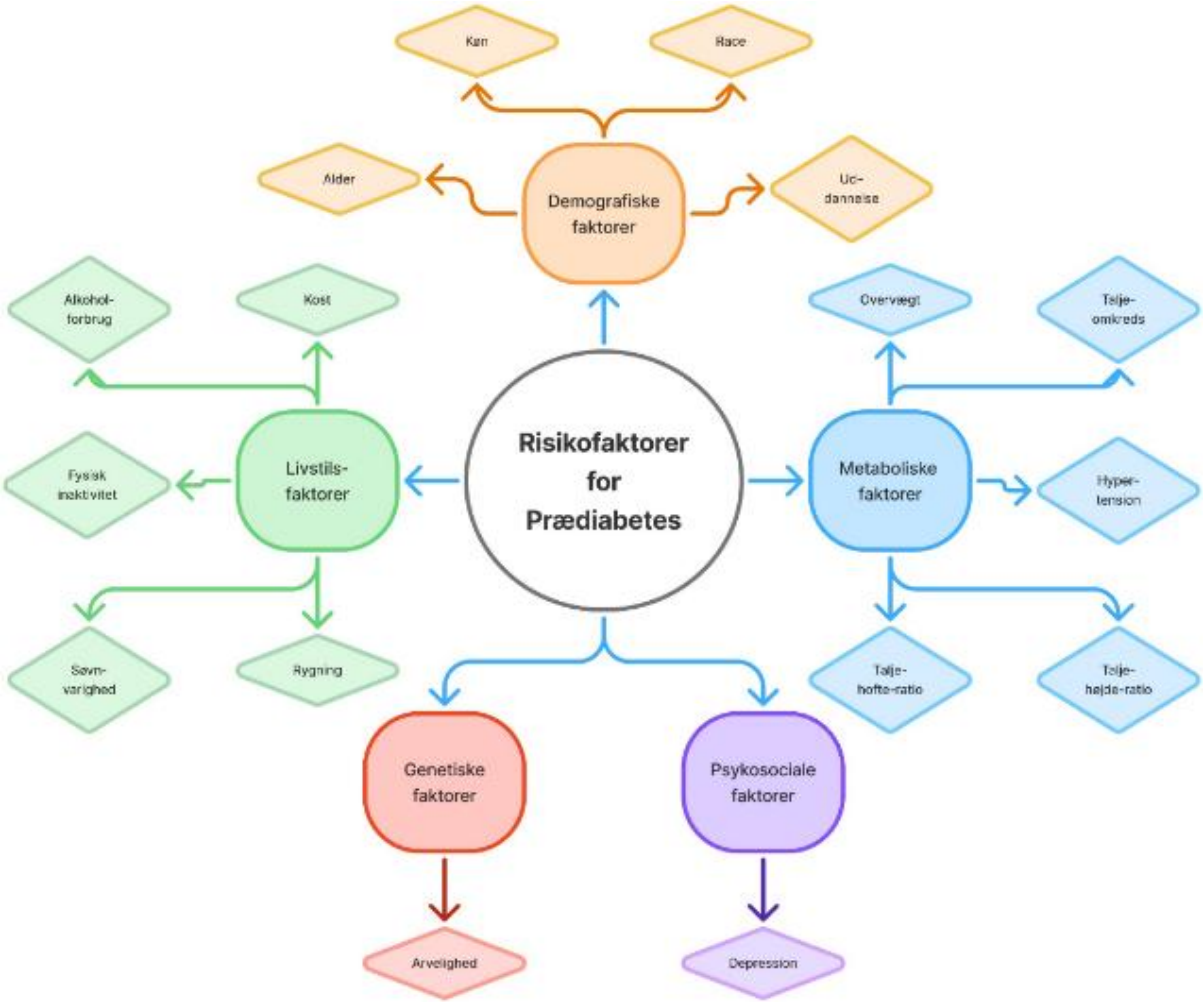
Risikoen for progression til T2D kan forebygges eller forsinkes hos en stor andel ved hjælp af tidligt målrettede livsstilsinterventioner (11,22,23). Et review konkluderede, at personer, der modtog livsstilsinterventioner, havde en 28–58% lavere risiko for at udvikle T2D sammenlignet med en kontrolgruppe uden intervention. Vægttab blev identificeret som den vigtigste faktor for at reducere risikoen (11).

Prospektive studier viste, at 5–10% af personer med prædiabetes årligt udvikler T2D (24). Inden for fem år vil én ud af fem personer med prædiabetes statistisk set udvikle T2D (25). Risikoen for progression til T2D varierer dog afhængigt af, hvilken definition af prædiabetes, der anvendes (8,11). En metaanalyse fra 2019, der inkluderede 16 studier med fem forskellige definitioner af prædiabetes, fandt, at risikoen for at udvikle T2D inden for fem år var 4–8 gange højere sammenlignet med personer uden prædiabetes(26).

1.3 Risikofaktorer for udvikling af prædiabetes

Der findes en lang række risikofaktorer for udvikling af prædiabetes, herunder metaboliske-, demografiske-, genetiske-, livsstils- og psykosociale faktorer, hvilket ses illustreret i figur 1. I dette afsnit

fokuseres der på ikke-laboratorieafhængige risikofaktorer, det vil sige faktorer, der kan identificeres uden behov for blodprøver eller andre biokemiske analyser. Biokemiske markører såsom lipidprofil, levermarkører og inflammatoriske biomarkører er også associeret med prædiabetes, men inddrages ikke i denne gennemgang, da fokus er på risikofaktorer, der kan anvendes i ikke-invasiv screening og tidlig identifikation uden laboratorieanalyser



Figur 1: *Oversigt over ikke-laboratorieafhængige risikofaktorer for udvikling af prædiabetes*

1.3.1 Metaboliske faktorer

Flere metaboliske tilstande er tæt forbundet med risikoen for prædiabetes. Det er veldokumenteret, at overvægt (BMI 25,0-29,9) og fedme (BMI >30,0) er risikofaktorer (11,27,28). Prævalensen af prædiabetes er væsentligt højere blandt overvægtige personer sammenlignet med normalvægtige voksne, og risikoen øges markant med graden af fedme (11,27). Mere end 80% af personer med selvrapporteret prædiabetes er overvægtige (11). Dette skyldes primært, at en øget mængde visceralt fedt, som ofte ses ved overvægt og fedme, bidrager til insulinresistens og nedsat glukoseregulering (11,27).

Flere antropometriske mål er ligeledes stærkt forbundet med risikoen for prædiabetes, særligt taljeomkreds, talje-hofte-ratio og talje-højde-ratio (28,29). Taljeomkreds er en central indikator for abdominal fedme og er signifikant forbundet med øget risiko for prædiabetes. Talje-højde-ratio er en stærk prædiktor for metabolisk risiko, da den tager højde for kropsstørrelse og fedtfordeling (28,29). Talje-hofte-ratio er også forbundet med øget risiko for prædiabetes, men kan have lavere prædiktiv værdi end talje-højde-ratio (28). En anden vigtig risikofaktor er hypertension. Det ses på tværs af studier, at personer med hypertension har mere end dobbelt så høj risiko for at udvikle prædiabetes, hvor særligt et diastolisk blodtryk > 100 mmHg fremhæves som en større risiko end forhøjet systolisk blodtryk (30–32).

1.3.2 Demografiske faktorer

Alder er en af de stærkeste risikofaktorer og adskillige studier, der undersøger alder og prædiabetes, indikerer alle, at risikoen for udvikling af prædiabetes stiger i takt med stigende alder (11,28,30). Årsagen til dette er blandt andet at stigende alder er forbundet med en forringelse af betacellefunktionen og en øget tendens til insulinresistens. Ligeledes forbindes en stigende alder ofte med lavere fysisk aktivitetsniveau og deraf vægtøgning, hvilket yderligere øger risikoen (30). Prædiabetes er mere udbredt hos personer over 60 år sammenlignet med personer i alderen 18–44 år (11,28). Data fra National Health And Nutrition Survey (NHANES) 2017–2020 viser, at prævalensen blandt voksne i alderen 45–64 år er 44.8% (33), hvilket indikerer en betydelig stigning i risikoen i denne aldersgruppe. Alder > 45 år anses som en væsentlig uafhængig risikofaktor for prædiabetes, selv når der kontrolleres for indflydelse fra andre livsstilsfaktorer (30).

Køn er ligeledes en risiko for at udvikle prædiabetes (27,34), hvilket understøttes af en europæisk undersøgelse, der rapporterer, at mænd har tre gange højere risiko for at udvikle prædiabetes sammenlignet med kvinder (35). En forklaring på øget prævalens hos mænd kan findes i, at mænd generelt har en højere andel af trunkalt- og visceralt fedt samt en øget leverfedtprocent og større muskelmasse i overekstremiteterne sammenlignet med kvinder med samme alder og BMI (35).

Race og etnicitet anses som væsentlige risikofaktorer for prædiabetes. Data fra NHANES i perioden 2011-2016 viser, at prævalensen er højere blandt latinamerikanske voksne (41.6 %) efterfulgt af sorte voksne (39.9 %) og hvide voksne (36,1 %) (36). Et review viser, at forekomsten af IFG og IGT varierer betydeligt mellem etniske grupper, hvor asiatiske voksne oftere udvikler IGT, mens IFG er mere almindeligt blandt hvide voksne (37). Genetiske forskelle kan spille en rolle i forhold til insulinresistens og insulinsekretion, men race og etnicitet er ikke kausale faktorer i sig selv. Forskelle i prævalens kan i stedet delvist forklares af livsstil, socioøkonomiske forhold, miljø og variationer i diagnostiske kriterier (37).

Lavt uddannelsesniveau er associeret med højere prævalens af prædiabetes. I et amerikansk studie har 49.3% af voksne uden high school-eksamen prædiabetes, mod 37.0% blandt collegeuddannede. Selvom analysen ikke er justeret, tyder resultaterne på, at uddannelsesniveau kan være en relevant risikofaktor. Lavere uddannelse er samtidig forbundet med lavere indkomst, dårligere sundhedsadgang og mindre fysisk aktivitet, hvilket kan bidrage til øget risiko (38).

1.3.3 Livsstilsfaktorer

Flere livsstilsfaktorer viser sig at påvirke udviklingen af prædiabetes. Utilstrækkelig fysisk aktivitet kan føre til nedsat insulineffektivitet samt en uhensigtsmæssig udnyttelse af glukose og fedt i kroppens celler (39). Kost spiller også en central rolle, hvor et højt fedtindtag, særligt mættet fedt, er associeret med en øget risiko for nedsat glukoseregulering (39). Den vestlige kost, som er præget af et højt indhold af raffinerede kulhydrater, sukker og fedt, er særligt forbundet med denne øgede risiko samt med stigende forekomst af overvægt og T2D (30). Rygning og et højt alkoholforbrug er ligeledes relateret til en øget risiko for prædiabetes. Alkohol synes især at øge risikoen blandt mænd, mens rygning generelt er forbundet med forringet glukosemetabolisme (39). Søvnvaner spiller også en vigtig rolle. Flere studier viser, at kort søvnvarighed (<5 timer) er associeret med en øget risiko for at udvikle prædiabetes sammenlignet med normal søvnvarighed (7–8 timer). Den generelle søvnvarighed er således en vigtig faktor i vurderingen af prædiabetesrisiko (40,41).

1.3.4 Genetiske faktorer

En europæisk metaanalyse viser, at en familiær disposition for diabetes (defineret som diabetes hos mindst én førstegradsslægtning) er en væsentlig uafhængig risikofaktor for udviklingen af prædiabetes (42). Udfordringen ved denne faktor er, at mange personer muligvis ikke kender deres families medicinske historie samt at pårørende kan have prædiabetes eller T2D uden at være diagnosticeret (43).

1.3.5 Psykosociale faktorer

Der ses positiv association mellem depression og risikoen for prædiabetes, da personer med moderat til svær depression har en øget risiko sammenlignet med ikke-deprimerede. Den præcise mekanisme er ikke fuldt klarlagt, men en dysregulering af stresshormoner samt en øget insulinresistens kan være underliggende mekanismer (44).

1.4 Traditionelle metoder til identifikation af prædiabetes

1.4.1 Laboratoriebaserede metoder

Som tidligere præsenteret diagnosticeres prædiabetes ved hjælp af blodprøverne FPG, OGTT og HbA1c. Der er dog en række udfordringer forbundet med at anvende denne metode til identifikation af prædiabetes, hvilket vil uddybes i følgende afsnit.

Udfordringer ved laboratoriebaserede metoder

Trods deres status som standard for prædiabetesdiagnostik har laboratoriebaserede metoder flere udfordringer, særligt i ressourcebegrænsede rammer. For det første er testene invasive, da de kræver blodprøvetagning, hvilket kan afholde nogle fra screening (45). For det andet er de omkostningstunge og kræver specialiseret udstyr og sundhedspersonale (45). Desuden er laboratoriefaciliteter ofte utilgængelige i landområder og udviklingslande, hvilket kan resultere i lav screeningsrate, hvor mange forbliver udiagnosticeret med prædiabetes (11,45). Endelig gør behovet for sundhedspersonale og laboratorieanalyse det logistisk udfordrende at udføre gentagne målinger over tid, hvilket reducerer muligheden for tidlig opsporing og intervention (45). Dette understreger behovet for mere tilgængelige og omkostningseffektive screeningsmetoder, der ikke kræver laboratoriebaserede målinger.

1.4.2 Screeningsmetoder

Screening for prædiabetes anses som en omkostningseffektiv strategi til at reducere presset på sundhedsvæsenet, da det muliggør tidlig intervention og forebyggelse af udvikling til T2D (46). Mange sundhedsorganisationer anbefaler rutinemæssig screening (37,47), eksempelvis anbefaler ADA universel screening hvert tredje år for alle voksne over 35 år (47). På trods af dette forbliver screeningsraterne lave (37). Kun 19% af voksne med prædiabetes i USA rapporterer at have fået stillet diagnosen af en sundhedsprofessionel, og cirka 80% af voksne med prædiabetes i USA ikke er klar over deres tilstand (33). Dette indikerer, at eksisterende screeningsmetoder har begrænset rækkevidde og effekt.

Scoringsmodeller og statistiske screeningsværktøjer

Til at understøtte screening er der udviklet en række ikke-laboratorieafhængige risikomodeller baseret på selvrapportede oplysninger. Nogle af de mest udbredte er simple vægtede scoringsmodeller, hvor risikofaktorer som alder, BMI og fysisk aktivitet manuelt tildeles point baseret på epidemiologiske studier. Et velkendt eksempel er Finnish Diabetes Risk Score (FINDRISC), som er udviklet til brug uden sundhedsfagligt personale og kan identificere personer i risiko for prædiabetes ud fra selvrapporterede oplysninger (48). Et andet eksempel er CDC's Prediabetes Risk Test, som anvender en lignende point-baseret tilgang med faktorer som alder, BMI, blodtryk og aktivitetsniveau (49). I Danmark tilbyder

Diabetesforeningen en tilsvarende digital risikotest baseret på oplysninger om vægt, taljemål, kost og fysisk aktivitet (50).

Et fælles træk ved disse modeller er, at de typisk bygger på statistiske analyser, oftest logistisk regression, hvor koefficienter for risikofaktorer er estimeret ud fra større befolkningsdata og omsat til simple pointværdier (51). Denne tilgang gør modellerne lette at anvende i praksis og mulige at implementere uden avanceret teknologi. Et eksempel på en mere dataintensiv variant er Cambridge Risk Score, som er baseret på logistisk regression og trækker på kliniske oplysninger fra patientjournaler såsom tidligere glukoseniveauer, medicinforbrug og rygestatus (52).

Udfordringer ved screeningsmetoder

En væsentlig metodisk begrænsning ved disse modeller er deres forenklede antagelser om lineære og additive sammenhænge mellem risikofaktorer og udfald. Dette kan reducere modellernes evne til at opfange komplekse og ikke-lineære interaktioner, som ofte kendetegner tidlig prædiabetes (53). Derudover er transitionszonen mellem normoglykæmi og prædiabetes præget af subtile fysiologiske forskelle, som scorerne ofte overser (54).

Yderligere udfordringer knytter sig til generaliserbarhed, da mange modeller er udviklet i specifikke befolkningsgrupper og har begrænset præcision på tværs af etnicitet, køn og socioøkonomiske forhold(55). Endelig kræver de fleste værktøjer aktiv opsøgende adfærd fra borgeren selv, hvilket kan udgøre en barriere for tidlig detektion (56).

1.5 ML til detektion af prædiabetes

Mens statistiske modeller typisk bygger på antagelser om lineære og additive sammenhænge, har machine learning (ML) potentiale til at identificere komplekse, ikke-lineære mønstre og interaktioner mellem variable (57). Dette skyldes ML's evne til at lære direkte fra data uden foruddefinerede antagelser om relationer mellem input og output. ML kan dermed opfange subtile mønstre i store datasæt og tilpasse sig individuelle variationer i befolkningen (51,56).

Samtidig rummer ML en mulighed for automatisering af screeningsprocesser, hvor prædiktive modeller kan anvendes direkte på elektroniske sundhedsdata eller selvrapporterede informationer uden behov for manuelle vurderinger (51). Dette kan mindske behovet for ressourcekrævende kliniske tests og frigøre tid og arbejdskraft i sundhedsvæsenet, hvilket gør teknologien mindre omkostningstung på sigt (51,56).

Succesfuld implementering af ML-modeller kan bidrage til at understøtte klinikere, forbedre behandlingsforløb og medvirke til omkostningsreduktion (57). Dette gør ML til et strategisk redskab, ikke kun

på individniveau, men også i en bredere sundhedspolitisk kontekst, hvor tidlig opsporing og ressourceoptimering er afgørende.

Flere ML-modeller til detektion af prædiabetes er blevet udviklet i de seneste år. Disse kan overordnet opdeles i modeller, der kræver laboratorietests eller specialudstyr, og modeller, der baserer sig på ikke-laboratorieafhængige faktorer. De to modeltyper har forskellige anvendelsesmuligheder og begrænsninger, som gennemgås i det følgende.

1.5.1 ML-modeller baseret på laboratoriedata og specialiseret udstyr

ML er blevet anvendt i en række studier med avanceret inputdata, herunder laboratorieprøver (58), bærbare sensorer og internetbaserede systemer (59), retinal billedanalyse (60), MR-scanninger (61), optisk vaskulær signalanalyse (62), parodontale målinger (63) og tungediagnostik (64). Selvom disse modeller ofte viser høj præcision, er deres anvendelighed begrænset af høje omkostninger, krav om specialudstyr og ringe tilgængelighed, især i lavressourceområder (58–63).

1.5.2 Eksisterende ML-modeller baseret på ikke-laboratorieafhængig data

Andre studier har udviklet ML-modeller, som udelukkende anvender ikke-laboratorieafhængige data. Disse modeller er særligt relevante i kontekster, hvor adgang til laboratorietests er begrænset. Tabel 3 viser fire sådanne studier fra forskellige geografiske områder (57,65–67).

Studie	Studiepopulation	ML-modeller	AUC
Abbas et al., 2021	Qatar	Random Forest, XG Boost, Gradient Boosting Machine, Deep Learning, Logistisc Regression (LR)	0.803-0.815
Birk et al., 2021	Indien	Random Forest, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net,	0.700-0.705
Choi et al., 2014	Sydkorea	Artificial Neural Network, Support Vector Machine	0.729-0.731
Dinh et al., 2019	USA	Random Forest, XG Boost, Support Vector Machine, Ensemble Model	0.731-0.737

Tabel 3: *Eksisterende ML-modeller baseret på ikke-laboratorieafhængige faktorer.*

Resultaterne viser, at ML-modeller baseret på ikke-laboratoriedata kan opnå AUC-værdier på op mod 0.815, hvilket vidner om et betydeligt potentiale. Samtidig spænder performance på tværs af studier og modeller, hvilket kan skyldes forskelle i population, featureudvælgelse og algoritmevalg. For yderligere detaljer om de fire studier med eksisterende ML-modeller henvises til Søgeprotokol i bilag 1.

1.5.3 Begrænsninger ved eksisterende ML-modeller

Flere af modellerne i tabel 3 benytter avancerede algoritmer såsom Random Forest, XGBoost, og forskellige typer af neurale netværk. Sidstnævnte, herunder Deep Learning, anvender dybe arkitekturer

med mange lag og har evnen til selv at lære relevante features fra data uden behov for manuel feature-udvælgelse. Denne fleksibilitet gør dem kraftfulde og ofte mere præcise end klassiske modeller (68,69). Et centralt problem ved især dybe neurale netværk er deres manglende gennemsigtighed. De hører under kategorien black box-modeller, hvor deres interne beslutningslogik er vanskelig at forstå eller forklare, da der ikke foreligger et klart teoretisk grundlag for, hvordan specifikke inputværdier fører til et givent output, hvilket kan hæmme klinisk accept og anvendelse. Af den grund falder de uden for kategorien explainable artificial intelligence (AI), det vil sige ML-modeller, der på en forståelig måde kan forklare, hvilke inputvariabler der har bidraget til en given beslutning (68,69).

Explainability er særlig vigtig i en klinisk kontekst, hvor sundhedsprofessionelle skal kunne forstå og stole på modellens output. Hvis en model vurderer en patient til at være i høj risiko for prædiabetes, er det afgørende, at det er tydeligt, hvilke faktorer der ligger bag denne vurdering (69). Uden explainability er det svært at bruge modellen som beslutningsstøtte i praksis, hvilket begrænser dens anvendelighed, uanset hvor præcis den måtte være (69). Af denne grund er der voksende interesse for ML-modeller, der både opretholder høj prædiktiv performance og samtidig er explainable.

1.6 Afgrænsning

Prædiabetes rammer millioner af mennesker globalt, men størstedelen forbliver udiagnosticeret. En væsentlig årsag hertil er, at de nuværende screeningsmetoder, særligt laboratoriebaserede tests, er ressourcekrævende, omkostningsfulde og utilgængelige i mange sammenhænge. Samtidig identificerer forskellige diagnostiske tests kun delvist overlappende grupper, hvilket komplicerer tidlig opsporing. Eksisterende scoringsmodeller baseret på statistiske metoder er lette at anvende, men har begrænset evne til at håndtere komplekse, ikke-lineære sammenhænge og interaktioner mellem risikofaktorer.

På denne baggrund afgrænses dette kandidatspeciale til at undersøge, om ikke-laboratorieafhængige risikofaktorer alene kan anvendes til at identificere personer med prædiabetes. Der fokuseres udelukkende på risikofaktorer inden for metaboliske, demografiske, livsstilsrelaterede, genetiske og psykosociale domæner, det vil sige faktorer, der kan indsamles via spørgeskema eller selvrapportering uden behov for blodprøver, lægefaglig vurdering eller specialudstyr.

Specialet omfatter således ikke udvikling eller evaluering af modeller, der kræver laboratoriedata eller specialudstyr som input. Outcome er baseret på tre laboratorieværdier (HbA1c, FPG og OGTT), som anvendes til at definere prædiabetes. Da disse biomarkører måler forskellige fysiologiske mekanismer

og kun delvist overlapper, udvikles der tre separate modeller, én for hver test, for at opnå mere præcise risikoprofiler og undgå heterogenitet i definitionen af outcome.

1.7 Formål

Formålet med dette speciale er at undersøge, hvordan explainable ML kan anvendes til at detektere prædiabetes baseret på ikke-laboratorieafhængige risikofaktorer. Målet er ikke at finde én samlet model, men at undersøge, hvordan prædiktionspotentialet varierer med prædiabetesdefinitionen. Med udgangspunkt i data fra NHANES udvikles tre separate modeller, som hver især anvender én laboratoriebaseret prædiabetesdefinition som outcome. Dette muliggør en sammenligning af modellernes prædiktive egenskaber afhængigt af den anvendte definition.

1.8 Problemformulering

Hvordan kan explainable machine learning anvendes til at detektere prædiabetes ud fra ikke-laboratorieafhængige risikofaktorer, og hvordan varierer modellens prædiktive performance afhængigt af den anvendte prædiabetesdefinition (HbA1c, OGTT, FPG)?

2. Metode

2.1 Design og datagrundlag

Dette speciale blev baseret på data fra NHANES, et nationalt tværsnitsobservationsstudie, der vurderer sundhedstilstanden i den amerikanske befolkning (70). NHANES gennemføres løbende blandt repræsentativt udvalgte deltagere fra den generelle befolkning i USA og anvendes bredt i epidemiologisk forskning.

NHANES data er organiseret i særskilte datasæt for hver toårige undersøgelsesperiode og består af spørgeskemadata, fysisk undersøgelsesdata og laboratoriemålinger (71–73). Der blev anvendt data fra tre undersøgelsesperioder: 2011–2012, 2013–2014 og 2015–2016, da OGTT herefter ikke længere blev indsamlet.

NHANES udgjorde et velegnet datagrundlag til udvikling af ML-modeller, da det er omfangsrigt, standardiseret indsamlet og indeholder en bred vifte af relevante risikofaktorer på tværs af køn, alder og etnicitet. Der blev udviklet tre separate ML-modeller baseret udelukkende på ikke-laboratorieafhængige inputvariable.

Specialets metode og rapportering er desuden struktureret med udgangspunkt i Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis – Artificial Intelligence extension (TRIPOD+AI) anbefalinger for udvikling og præsentation af kliniske ML-modeller (74).

2.2 Udtræk af relevante variable

Til udviklingen af de tre ML-modeller blev der udtrukket én binær outcomevariabel per model baseret på henholdsvis HbA1c, FPG eller OGTT. Samtidig blev der identificeret en fællesmængde af inputvariable, som var tilgængelige i NHANES og udelukkende baseret på ikke-laboratorieafhængige faktorer. Tabel 4 viser en oversigt over de udtrukne variable, struktureret efter NHANES' datastruktur: fysisk målte data, demografiske data, spørgeskemadata og laboratoriedata.

Inputvariable		Outcomevariable
Målt undersøgelsesdata		*Målt laboratoriedata*
BMI: Kropsmasseindeks beregnet som vægt i kg divideret med højde i meter i anden (kg/m^2).		**HbA1c:** målt uden krav om faste, reflekterer det gennemsnitlige glukoseniveau over de seneste 2-3 måneder.
Taljeomkreds: målt i cm omkring midten af kroppen.		
Højde: målt i cm stående uden sko.		
Systolisk blodtryk: målt mellem 1-4 gange.		
Diastolisk blodtryk: målt mellem 1-4 gange.		
Selvrapporterede demografisk data		**FPG:** fastende plasmaglukosemåling efter mindst 9 timers faste.
Alder: ved screeningstidspunkt, angivet i hele år. Alle over 80 år er kodet som værende 80 år.		

Køn: registreret som enten mand eller kvinde.		
Race/Etnicitet: Oplysninger om rapporteret race og etnicitet.		
Uddannelsesniveau: Højest fuldførte uddannelse blandt voksne deltagere over 20 år.		
Selvrapporterede spørgeskemadata		**OGTT:** udført ved at måle plasma-glukoseniveauet 2 timer efter indtagelse af en 75 g glukoseopløsning.
Rygning: Har røget 100 cigaretter eller mere i livet (ja/nej).		
Alkoholforbrug: Gennemsnitligt dagligt forbrug de seneste 12 måneder.		
Søvnvarighed: Gennemsnitligt antal timers søvn per nat.		

Tabel 4: *Oversigt over de udtrukne variabler til definering af input- og outcomevariable.*

2.3 Præprocessering

2.3.1 Definition af outcome

Hver af de tre ML-modeller anvendte en binær outcomevariabel baseret på én specifik laboratorietest: enten HbA1c, FPG eller OGTT. Outcome blev defineret som tilstedeværelse eller fravær af prædiabetes i overensstemmelse med de diagnostiske kriterier fra ADA, som er internationalt anerkendte og hyppigt anvendt i epidemiologiske studier (11).

En deltager blev klassificeret som havende prædiabetes i den pågældende model, hvis testværdien lå inden for de definerede grænseværdier. Deltagere med værdier uden for området blev klassificeret som normoglykæmi. Tabel 5 viser de anvendte grænseværdier.

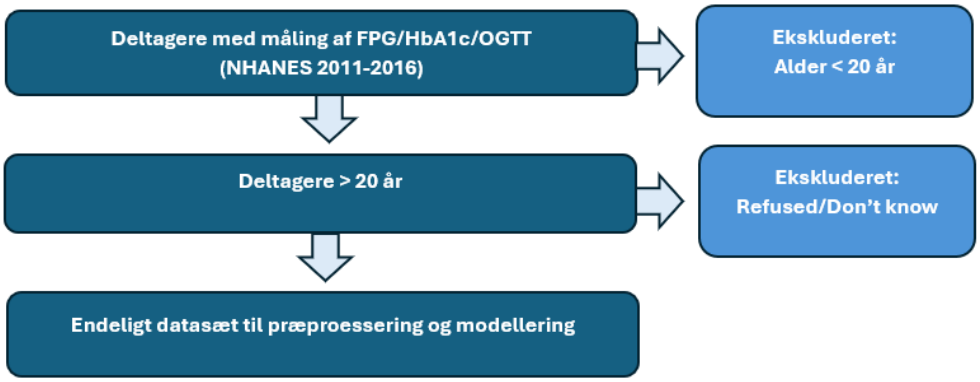
Diagnostisk test	Grænseværdi	Enhed
HbA1c	5.7-6.4	%
FPG	100-125	mg/dL
OGTT	140-199	mg/dL

Tabel 5: *Oversigt over de tre diagnostiske tests og tilhørende prædiabetesgrænser anvendt som outcome.*

2.3.2 Filtrering af data

Efter udtræk og definition af relevante variable blev datasættet filtreret med henblik på at sikre valide og komplette observationer til hver af de tre modeller. For hver model blev der først identificeret deltagere med en gyldig måling af den laboratorietest, der definerede outcomevariablen. Kun disse deltagere blev medtaget i den videre behandling. Denne filtrering blev foretaget særskilt for hver model og på tværs af alle NHANES-årgange (2011–2016).

Herefter blev inputvariable koblet til ud fra respondent-ID. Deltagere under 20 år blev ekskluderet, da variablen om uddannelsesniveau kun foreligger for voksne ≥ 20 år. Desuden blev observationer fjernet, hvor én eller flere inputvariable var kodet som “Refused” eller “Don’t know”, da disse ikke blev vurderet egnede til imputering. Filtreringen resulterede i tre separate analysepopulationer, én for hver outcomevariabel, hvilket kan ses illustreret i figur 2.



Figur 2: *Viser den trinvise eksklusion af deltagere. Filtreringen blev anvendt separat for hver outcomevariabel.*

2.3.3 Beregning af afledte inputvariable

Udvalgte inputvariable blev konstrueret ud fra eksisterende målinger i NHANES. Talje-højde-ratio, en veldokumenteret risikofaktor for prædiabetes (28,29), indgik ikke som en selvstændig variabel i datasættet og blev derfor beregnet som forholdet mellem taljeomkreds og stående højde. Højde blev udelukkende anvendt til denne beregning og indgik derfor ikke i modellen.

For blodtryk blev der beregnet én gennemsnitsværdi per deltager for henholdsvis systolisk og diastolisk tryk, da NHANES gennemførte mellem 1 og 4 målinger per person. Gennemsnitsværdierne blev anvendt som modelinput for at reducere tilfældig målevariation og sikre konsistens på tværs af deltagere.

2.3.4 Kodning af inputvariable

Kontinuerlige inputvariable

De kontinuerlige variable omfattede alder, BMI, taljeomkreds, talje-højde-ratio, systolisk- og diastolisk blodtryk. Disse variable blev anvendt i deres oprindelige skala uden yderligere transformation eller gruppering.

Kategoriske inputvariable

De kategoriske variable omfattede køn, uddannelsesniveau, søvnvarighed, alkoholforbrug, rygning og race/etnicitet. Køn og rygning blev behandlet som binære variable. Race/etnicitet blev omkodet til seks separate indikatorvariable via one-hot encoding. Uddannelsesniveau blev behandlet som en ordinal variabel med stigende kategorier. Søvnvarighed og alkoholforbrug blev omkodet til ordinale kategorier baseret på naturlig ordensstruktur, datadistribution og sundhedsfaglige anbefalinger, jf. afsnit 2.3.5.

2.3.5 Inspektion og håndtering af sjældne inputværdier

Kontinuerlige inputvariable

Der blev ikke foretaget inspektion af sjældne- eller ekstreme værdier i de kontinuerlige inputvariable, da explainable ML-modeller typisk håndterer variation i kontinuerlige variable internt gennem

modelstrukturen. I sådanne modeller tilpasses grænseværdier automatisk under træning, hvilket reducerer behovet for særskilt transformation (75).

Kategoriske inputvariable

Alle kategoriske inputvariable blev inspiceret for lavfrekvente kategorier i hver af de tre modeller med henblik på at identificere grupper med potentiel risiko for overfitting. Modellen risikerer at lære mønstre, der kun gælder for meget små og ikke-repræsentative undergrupper. Kategorier med en forekomst på under 1% af deltagerne blev derfor defineret som sjældne, hvilket er i overensstemmelse med anbefalinger i datavidenskabelig praksis (76,77).

Køn og rygning, som begge er binære, blev undtaget inspektion. For race/etnicitet og uddannelsesniveau viste inspektionen, at ingen kategorier optrådte under 1% i nogen af de tre modeller. De laveste var henholdsvis 2.8% ("Anden race") og 9.6% ("Less Than 9th Grade"). Der blev derfor ikke foretaget kodemæssige sammenlægninger og variablerne blev fastholdt i deres oprindelige inddeling. Fordelingerne af race/etnicitet og uddannelsesniveau fremgår af bilag 2.

Omkodning af alkoholforbrug og søvnvarighed

Alkoholforbrug udviste en højreskæv fordeling med mange deltagere i den lave ende af skalaen og meget få med højt forbrug. Søvnvarighed var centreret omkring et midterinterval, men med asymmetrisk fordeling og få observationer i ydergrupperne. På baggrund af disse fordelingsmønstre blev begge variable omkodet til kategoriske variable.

Alkoholforbrug blev inddelt i fire kategorier; 1–2, 3–4, 5–6 og >7 genstande per dag baseret på CDC's definition af moderat alkoholforbrug og anbefalinger fra Sundhedsstyrelsen (78,79).

Søvnvarighed blev ligeledes inddelt i fire kategorier; ≤5 timer, 6–6.5 timer, 7–9 timer og ≥9.5 timer baseret på internationale retningslinjer for optimal søvnvarighed(80–82).

Fordelingerne af råværdier for alkoholforbrug og søvnvarighed før omkodning fremgår af bilag 3. De færdigkonstruerede inputvariable efter præprocessering fremgår samlet i tabel 6. Tabellen er fælles for alle tre modeller, da de samme 17 inputvariable indgår.

Nr.	Inputvariabel	Skalaniveau	Kodning
Kontinuerlige			
1	BMI	Kontinuerlig	-
2	Taljeomkreds	Kontinuerlig	-
3	Talje-højde-ratio	Kontinuerlig	-
4	Systolisk blodtryk	Kontinuerlig	-
5	Diastolisk blodtryk	Kontinuerlig	-

6	Alder	Kontinuerlig	-
Kategoriske			
7	Køn	Binær	1/2
8	Rygning	Binær	1/2
9	Uddannelsesniveau	Ordinal	1-5
10	Alkoholforbrug	Ordinal	1-4
11	Søvnvarighed	Ordinal	1-4
12	Race: Mexicansk-amerikansk	One-hot encoded	0/1
13	Race: Andet latinamerikansk	One-hot encoded	0/1
14	Race: Ikke-latinamerikansk hvid	One-hot encoded	0/1
15	Race: Ikke-latinamerikansk sort	One-hot encoded	0/1
16	Race: Ikke-latinamerikansk asiatisk	One-hot encoded	0/1
17	Race: anden eller blandet race	One-hot encoded	0/1

Tabel 6: *Oversigt over de 17 inputvariable, deres skalaniveau og kodning.*

2.3.6 Analyse og håndtering af manglende værdier

Manglende data blev analyseret separat for de tre modeller, da hver model er baseret på en forskellig underpopulation. Andelen af manglende værdier i hver inputvariabel fremgår af bilag 4.

Den højeste andel blev fundet for alkoholforbrug (ca. 38–40%), hvor de manglende værdier forekom i alle tre årgange og blev vurderet at være Not Missing at Random (NMAR) (83) på baggrund af spørgeskemaets opbygning og svarmønstre. Den blev håndteret ved omkodning af manglende værdier til 0 svarende til antagelsen om manglende rapportering af alkoholforbrug.

For de øvrige variable var andelen af manglende data lav med maksimalt 6.3% for talje-højde-ratio. Disse manglende værdier blev antaget at være Missing at Random (MAR) (83). Blandt alle inputvariable var det kun for alkoholforbrug, at manglende data blev omkodet.

2.3.7 Imputeringsteknik

Manglende værdier i de resterende kategoriske og kontinuerlige inputvariable blev håndteret ved hjælp af IterativeImputer fra Pythonpakken Scikit-learn (84). Denne metode er baseret på principperne om multipel imputering ved chained equations (MICE) og anvendes ofte til imputering af sundhedsdata under antagelsen om MAR (85,86).

Det maksimale antal iterationer blev sat til 15, og algoritmen blev konfigureret til at stoppe tidligere, hvis konvergens blev nået. Efter imputering blev datasættets fordelinger og centrale tendenser sammenlignet med det oprindelige datasæt for at kontrollere for ændringer i datamønstret (86). Forskellene var minimale, hvilket indikerede, at imputering ikke havde introduceret systematiske skævheder.

2.4 Modellering

2.4.1 Split af datasæt

Datasættet blev opdelt i et trænings- og testdatasæt i forholdet 80/20 med henblik på at sikre tilstrækkelig datamængde til både modeltræning og evaluering (87,88). For at opretholde den relative fordeling mellem deltagere med og uden prædiabetes, se tabel 7, blev der anvendt stratificeret sampling baseret på outcomevariablen (87).

Model	Normoglykæmi (n, %)	Prædiabetes (n, %)	I alt (n)
FPG	4.950 (62,8 3%)	2.928 (37,17%)	7878
HbA1c	11.956 (73,1%)	4.399 (26,9%)	16.355
OGTT	5.989 (86,8%)	911 (13,2%)	6.900

Tabel 7: *Fordeling af normoglykæmi og prædiabetes i de tre modeller. Tabellen viser antal og procentvise andele af deltagere i hver gruppe efter den anvendte laboratoriebaserede outcomevariabel.*

Opdelingen blev udført med en fast random seed for at sikre reproducerbarhed(89). Træningsdatasættet blev anvendt til både feature selection og modeltræning, mens testdatasættet udelukkende blev benyttet til endelig evaluering af modellens performance (87,89). Fordelingerne af inputvariable blev efterfølgende sammenlignet mellem trænings- og testdatasættet for hver model med henblik på at kontrollere for eventuelle systematiske forskelle.

2.4.2 Valg af model

Decision Tree blev valgt til udvikling af modellerne. Decision Trees genererer en sekventiel struktur af "hvis-så" regler baseret på inputvariable og muliggør dermed tydelig identifikation af mønstre og prædiktioner i datasættet. Classification and Regression Tree (CART) blev anvendt som algoritme da den, på baggrund af Gini-indekset som kriterium, er i stand til at identificere det optimale split i træet (90). Denne type model er let at fortolke og visualisere, hvilket gør den egnet i sammenhænge, hvor explainability er vigtig. Tidligere studier har vist, at Decision Tree kan anvendes til klassifikation inden for diabetesrelaterede problemstillinger med tilfredsstillende performance (51,91).

2.4.3 Software og programmeringssprog

Alt databehandling- og modelleringsarbejde blev udført i programmeringssproget Python, som er udbredt i både forsknings- og udviklingssammenhænge (92,93). Python blev anvendt i kombination med open source-biblioteker som Pandas, Numpy og Scikit-learn. Disse værktøjer muliggjorde en effektiv datamanipulation, imputering, modellering og evaluering i én samlet arbejdsgang (94).
Udviklingsmiljøet bestod af Deepnote, en cloudbaseret notebook-platform (95), der tillod en interaktiv kodning og versionskontrol.

2.4.4 Feature selection

Feature selection blev gennemført separat for hver af de tre modeller med det formål at identificere de mest informative inputvariable. Der blev anvendt forward feature selection med et stopkriterium på minimum 0.01 forbedring i AUC (96).

2.4.5 Træning

Hver model blev trænet på sit eget træningsdatasæt. Til træningen blev der foretaget systematisk tuning af centrale hyperparametre, herunder træets maksimale dybde (max_depth), det minimale antal observationer per slutnode (min_samples_leaf) og vægtning af klasser (class_weight). Disse parametre blev valgt på baggrund af deres dokumenterede betydning for modelkompleksitet og risiko for overfitting i Decision Trees (97). Der blev afprøvet forskellige værdier for hver parameter for at identificere den mest optimale konfiguration.

For hver model blev træningsdatasættet opdelt i et subtrænings- og et valideringssæt i et 80/20 forhold. Modelkonfigurationer med og uden klassevægtning blev sammenlignet baseret på AUC i valideringssættet. Hvis vægtning af klasser medførte højere AUC sammenlignet med uden vægtning, blev dette valgt til den endelige model. Ved ens AUC mellem to konfigurationer, blev den med lavest modelkompleksitet (lavere max_depth eller uden class_weight) foretrukket.

De endelige modeller blev herefter trænet på hele træningsdatasættet og evalueret på et uafhængigt testdatasæt, som ikke indgik i hverken feature selection eller tuning.

Risikoen for overfitting blev vurderet ved at sammenligne AUC på valideringssættet under hyperparametertuning med AUC på det uafhængige testdatasæt (89). En forskel større end 0,1 mellem valideringsog test AUC blev fortolket som tegn på overfitting. Modelspecifikationer fremgår af tabel 8.

Model	class_weight	max_depth	min_samples_leaf
FPG	none	5	2
HbA1c	none	3	2
OGTT	balanced	3	2

Tabel 8: *Modelspecifikationer for hver model, herunder anvendt class_weight, max_depth og min_samples_leaf.*

2.4.6 Evaluering

Performance blev vurderet på testdatasættet med henblik på at vurdere evnen til at skelne mellem personer med og uden prædiabetes. Til dette formål blev AUC anvendt som primær metrik, da den giver et robust, samlet mål for diskrimination i binære klassifikationsopgaver (87). Der blev suppleret med metrikkerne recall (sensitivity), specificity, precision og accuracy for at give et mere nuanceret

billede af modellernes performance (51,98). Definitioner og beregningsformler for disse metrikker fremgår af bilag 5.

For at identificere den optimale threshold for klassifikation, blev F1-score anvendt som kriterium, da metrikken balancerer recall og precision i én samlet måling (99). For hver af de tre modeller blev der systematisk afprøvet thresholds fra 0.1-0.9 med trin på 0.1. Den threshhold, der resulterede i den højeste F1-score på testdatasættet, blev valgt som modelspecifik cut-off.

2.4.7 Supplerende analyse

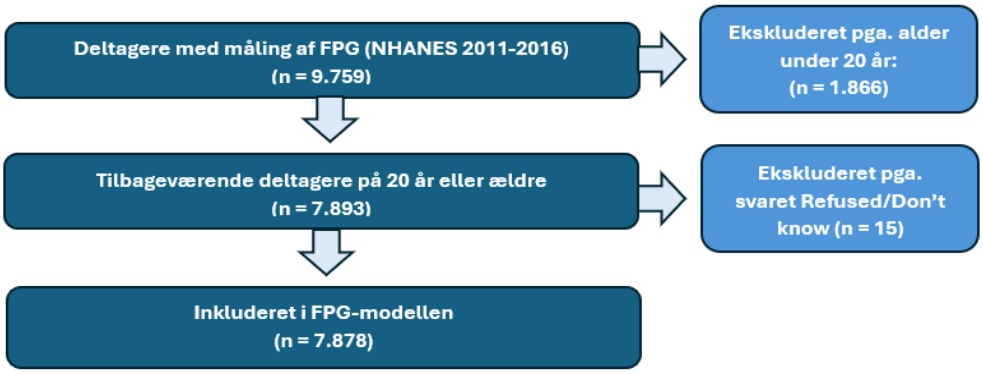
Som en supplerende analyse blev fordelingen af de endeligt inkluderede inputvariable sammenlignet mellem deltagere med prædiabetes og normoglykæmi i testdatasættet for hver af de tre modeller. Formålet var at undersøge, om de to grupper adskilte sig væsentligt på de variable, som modellerne faktisk anvendte for at vurdere, om svag modelperformance kunne skyldes lav diskriminationsevne i de anvendte variable.

Da hver model var baseret på en forskellig kombination af inputvariable efter feature selection, blev analysen udført separat for hver model og kun på de variable, der indgik i den endelige model. Resultaterne præsenteres i bilag 6.

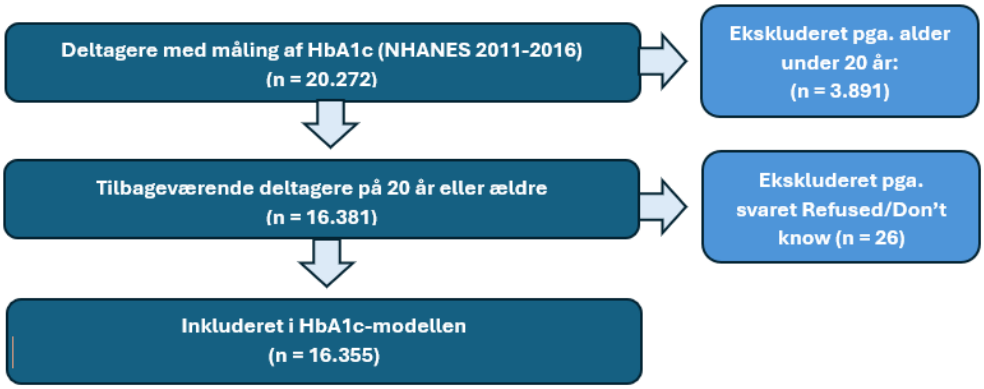
3. Resultater

3.1 Deltagerflow

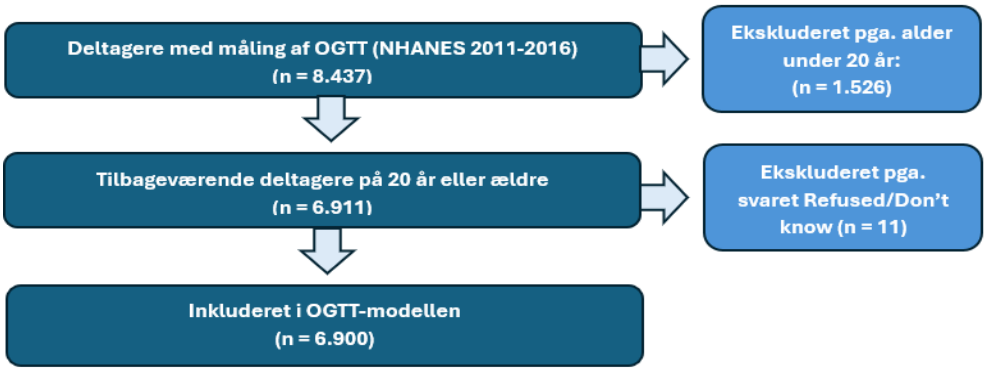
Figur 3-5 viser deltagerflowet for de tre separate modeller. I alt blev der inkluderet 16.355 deltagere i HbA1c-modellen, 7.878 i FPG-modellen og 6.900 i OGTT-modellen.



Figur 3: *Deltagerflow for FPG-modellen. Figuren viser antallet af deltagere med måling af FPG i NHANES 2011–2016 og de efterfølgende eksklusioner, herunder alder <20 år og ugyldige inputbesvarelser.*



Figur 4: *Deltagerflow for HbA1c-modellen. Figuren viser antallet af deltagere med måling af HbA1c i NHANES 2011–2016 og de efterfølgende eksklusioner, herunder alder <20 år og ugyldige inputbesvarelser.*



Figur 5: *Deltagerflow for OGTT-modellen. Figuren viser antallet af deltagere med måling af OGTT i NHANES 2011–2016 og de efterfølgende eksklusioner, herunder alder <20 år og ugyldige inputbesvarelser.*

3.2 Deltagerkarakteristika

3.2.1 Overordnede karakteristika

Deltagerkarakteristika for de tre modeller er opsummeret i tabel 9 (kontinuerlige variable) og tabel 10 (kategoriske variable). Fordelingerne afspejler datasættene efter imputering. Kontinuerlige variable er præsenteret som gennemsnit ± standardafvigelse (SD), og kategoriske variable som procentvis fordeling. I alle tre modeller var gennemsnitsalderen omkring 49 år og BMI cirka 29. Køn og uddannelsesniveau var jævnt fordelt, og størstedelen af deltagerne sov 7–9 timer og indtog 1–2 genstande dagligt.

Kontinuerlige inputvariable	FPG-model (n = 7.878)	HbA1c-model (n = 16.355)	OGTT-model (n = 6.900)
BMI	29.0 (SD=6,9)	29.2 (SD=7,0)	29,2 (SD=7,1)
Taljeomkreds (cm)	99.3 (SD=16,6)	99.5 (SD=16,7)	99,4 (SD=16,6)
Talje-højde-ratio	0.6 (SD=0,1)	0.6 (SD=0,1)	0,6 (SD=0,1)
Systolisk blodtryk (mm Hg)	123.9 (SD=18,3)	124.1 (SD=18,0)	123,7 (SD=18,0)
Diastolisk blodtryk (mm Hg)	69.3 (SD=12,5)	70.1 (SD=12,4)	69,2 (SD=12,3)
Alder (år)	49.7 (SD=17,6)	49.1 (SD=17,6)	49,6 (SD=17,4)

Tabel 9: *Gennemsnit og standardafvigelse (SD) for kontinuerlige inputvariable blandt deltagere inkluderet i hhv. FPG-, HbA1c- og OGTT-modellerne. Værdierne er opgjort efter imputering.*

Kategoriske inputvariable	FPG-model (n = 7.878)	HbA1c-model (n = 16.355)	OGTT-model (n = 6.900)
Køn			
Mænd	48.7%	48,3%	48,9%
Kvinder	51.3%	51,7%	51,1%
Rygning			
Ja	43.7%	42,8%	43,5%
Nej	56.3%	57,2%	56,5%
Uddannelsesniveau			
Less than 9th grade	9,8%	9,8%	9,6%
9-11th Grade (Includes 12th grade with no diploma)	13.6%	13,2%	13,2%
Hight School Grad/FED or Equivalent	21.5%	21,7%	21,5%
Some College or AA degree	29.5%	30,2%	29,5%
College Graduate or above	25.5%	25,1%	26,1%
Alkoholforbrug			
0 genstande	0.1%	0.1%	0.1%
1-2 genstande	61.1%	60.0%	61.5%
3-4 genstande	29.5%	30.8%	29.2%
5-6 genstande	6.0%	5.8%	5.9%
7+ genstande	3.3%	3.3%	3.2%
Søvn			
≤ 5 timer	12.6%	12.1%	12.5%
6-6,5 timer	20.1%	20.6%	20.1%

7-9 timer	60.7%	60.4%	61.1%
≤ 9,5 timer	6.7%	6.9%	6.3%
Race/Etnicitet			
Race: Mexicansk-amerikansk	13,1%	13.6%	13.4%
Race: Andet latinamerikansk	11.0%	10.7%	11.2%
Race: Ikke-latinamerikansk hvid	38.7%	37.5%	39.3%
Race: Ikke-latinamerikansk sort	21.6%	22.6%	20.5%
Race: Ikke-latinamerikansk asiatisk	12.8%	12.5%	12.7%
Race: Anden eller blandet race	2.9%	3.1%	2.8%

***Tabel 10**: Procentvis fordeling af kategoriske inputvariable blandt deltagere inkluderet i henholdsvis FPG-, HbA1c- og OGTT-modellerne. Fordelingerne er opgjort efter imputering.*

3.2.2 Sammenligning af trænings- og testdatasæt

Inputfordelingerne i trænings- og testdatasættet blev sammenlignet separat for hver model med henblik på at kontrollere for skævheder i datagrundlaget efter den stratificerede 80/20-randomisering. Tabel 11 viser fordelingen af de endeligt inkluderede inputvariable i hver model opdelt på trænings- og testdata. Enkelte mindre forskelle blev observeret i fx. uddannelsesniveau og alkoholforbrug i FPG-modellen ellers var fordelingen af inputvariable mellem trænings- og testdata generelt balanceret.

Inputvariable	FPG-model Træning / Test		HbA1c-model Træning / Test		OGTT-model Træning / Test	
BMI	29.0 (6.9)	29.4 (7.3)	29.2 (7.0)	29.1 (7.0)	29.2 (7.0)	29.1 7.(7.2)
Taljeomkreds (cm)	99.2 (16.4)	99.8 (17.3)	99.5 (16.7)	99.3 (16.7)	99.6 (16.7)	99.8 (17.2)
Talje-højde-ratio	0.6 (0.1)	0.6 (0.1)	0.6 (0.1)	0.6 (0.1)	0.6 (0.1)	0.6 (0.1)
Systolisk blodtryk (mm Hg)	123.9 (18.4)	123.7 (18.0)	124.2 (18.0)	124.0 (18.3)	123.7 (18.2)	123.7 (17.2)
Diastolisk blodtryk (mm Hg)	69.4 (12.4)	69.1 (12.8)	70.0 (12.4)	70.3 (12.6)	69.2 (12.4)	69.5 (12.2)
Alder (år)	49.7 (17.6)	49.6 (17.6)	49.2 (17.7)	48.8 (17.4)	49.6 (17.5)	49.6 (17.0)
Køn						
Mænd	48.3	50.1	48.3	48.2	48.7	49.4
Kvinder	51.7	49.9	51.7	51.8	51.3	50.6
Rygning						
Ja	43.6	44.2	42.5	44.2	43.3	44.2
Nej	56.4	55.8	57.5	55.8	56.7	55.8
Uddannelsesniveau						
< 9th grade	9.6	10.7	9.7	10.2	9.8	8.8
9-11th grade	13.5	14.0	13.2	13.0	13.3	12.8
High school / GED eller tilsvarende	21.5	21.7	21.8	21.4	21.0	23.6
Some college eller AA-degree	29.1	30.9	30.3	29.9	29.6	29.2
College graduate eller over	26.2	22.7	25.0	25.4	26.2	25.6
Alkoholforbrug						

0 genstande	0.1	0.1	0.1	0.1	0.1	0.1
1-2 genstande	61.3	60.2	60.2	59.2	61.1	63.2
3-4 genstande	29.6	29.1	30.7	31.5	29.6	27.6
5-6 genstande	5.6	7.5	5.9	5.4	5.8	6.6
7+ genstande	3.3	3.2	3.2	3.8	3.4	2.5
Søvnvarighed						
≤ 5 timer	13.1	10.5	12.0	12.4	12.6	12.2
6-6,5 timer	19.6	21.9	20.7	20.2	20.1	20.1
7-9 timer	60.6	61.0	60.5	60.0	61.1	61.0
≤ 9,5 timer	6.7	6.5	6.8	7.4	6.2	6.7
Race/Etnicitet						
Race: Mexicansk-amerikansk	12.9	13.8	13.5	13.9	13.4	13.5
Race: Andet latin-amerikansk	10.7	12.2	10.7	10.8	11.2	11.3
Race: Ikke-latiname-rikansk hvid	39.1	37.2	37.9	35.8	39.5	38.6
Race: Ikke-latiname-rikansk sort	21.3	22.8	22.4	23.6	20.1	21.9
Race: Ikke-latiname-rikansk asiatisk	13.0	11.9	12.4	12.9	12.9	12.1
Race: Anden eller blandet race	3.1	2.1	3.2	3.0	2.9	2.6

Tabel 11: *Fordeling af de endeligt inkluderede inputvariable i trænings- og testdatasættet for hver model. Værdierne er opgjort som procent for kategoriske variable og gennemsnit (standardafvigelse) for kontinuerlige variable.*

3.3 Valgte inputvariable efter feature selection

Tabel 12 præsenterer de variable, der blev inkluderet i de endelige modeller efter feature selection. Som supplement blev der gennemført en univariat AUC-analyse for hver model for at belyse den individuelle prædiktive styrke af hver variabel. Resultaterne af denne analyse fremgår af bilag 7. Derudover er en manuel forward feature selection med alle 17 variable er præsenteret i bilag 8, som viser ændringen i AUC efter hver tilføjelse.

Inputvariable	FPG-model	HbA1c-model	OGTT-model
BMI			
Taljeomkreds			
Talje-højde-ratio	✓		
Systolisk blodtryk			
Diastolisk blodtryk			
Alder		✓	✓
Køn	✓		
Rygning			
Uddannelsesniveau			
Alkoholforbrug			

Søvn			
Race: Mexicansk-amerikansk			
Race: Andet latinamerikansk			
Race: Ikke-latinamerikansk hvid			
Race: Ikke-latinamerikansk sort			
Race: Ikke-latinamerikansk asiatisk			
Race: Anden eller blandet race			

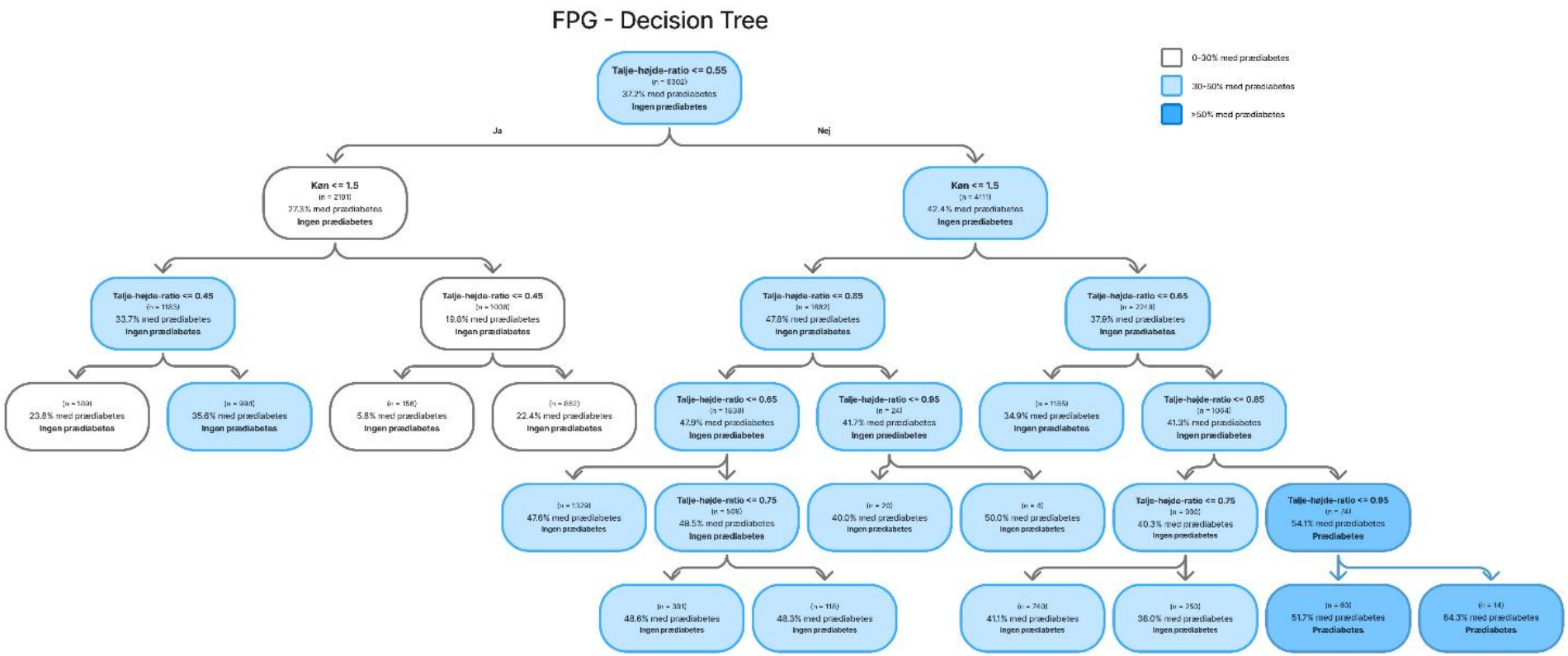
Tabel 12: *Oversigt over inkluderede inputvariable i de endelige modeller efter forward feature selection. "✓" angiver, at variablen blev inkluderet i den pågældende model. Tomme felter angiver, at variablen blev ekskludere.*

3.4 Modellernes struktur og forklaring

Figur 6-8 viser de endelige Decision Trees for henholdsvis FPG-, HbA1c- og OGTT-modellen. De illustrerer den sekventielle opdeling af observationsdata baseret på udvalgte inputvariable, der anvendes til at klassificere prædiabetesstatus.

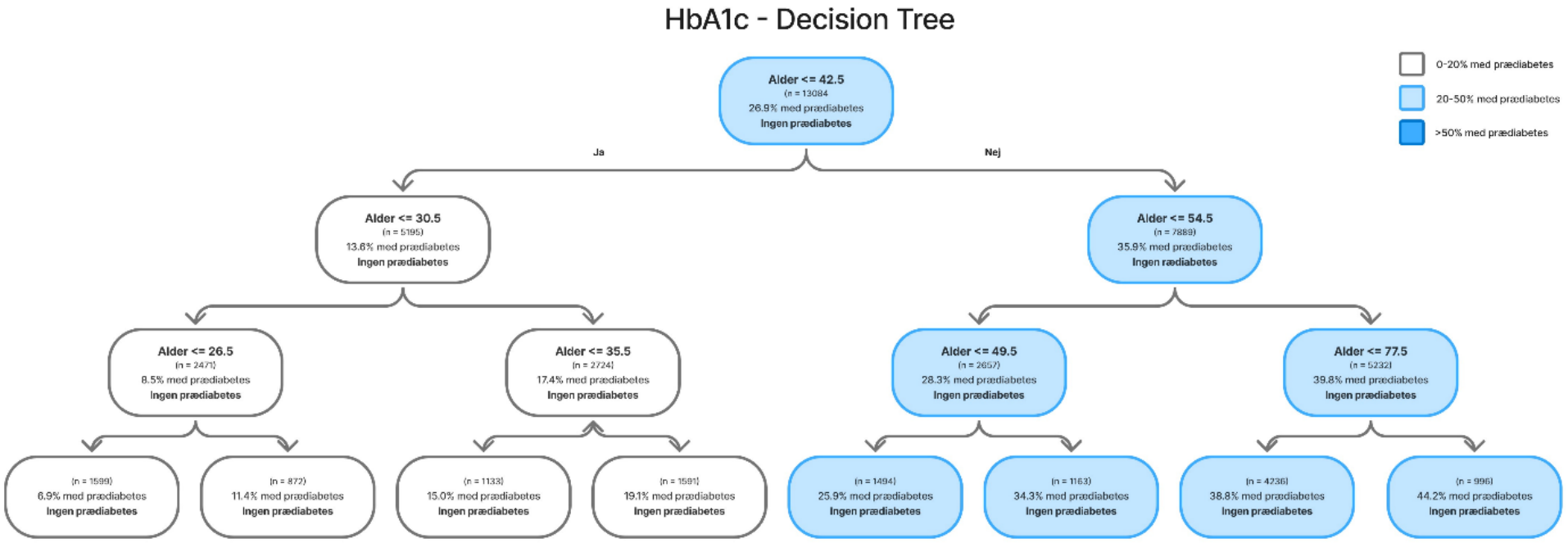
Hver node repræsenterer en binær opdeling baseret på en threshold i en given variabel. Farverne angiver andelen med prædiabetes og er manuelt tilpasset de tre modellers anvendte thresholds jf. afsnit 3.5.3 (0.20 for HbA1c, 0.30 for FPG og 0.50 for OGTT). Noder med en andel under threshold vises som hvide og klassificeres som "Ingen prædiabetes", mens noder med en andel over de valgte threshold vises i lyse- eller mørkeblå, afhængigt af om prædiabetesandelen er under eller over 50 %. Bemærk, at tekstetiketten i hver node ("Prædiabetes" eller "Ingen prædiabetes") genereres automatisk ud fra Scikit-learns standardklassifikationsgrænse på 0.5. Det betyder, at der i FPG- og HbA1c-modellen kan forekomme noder, hvor teksten angiver "Ingen prædiabetes", selvom nodens prævalens og farve ifølge specialets justerede threshold indikerer prædiabetes. For OGTT-modellen svarer threshold til standardgrænsen, hvorfor tekst og farve stemmer overens. De fulde træstrukturer kan ses i bilag 9.

FPG-modellen består af to inputvariable: talje-højde-ratio og køn. Første opdeling sker på talje-højde-ratio efterfulgt af køn.



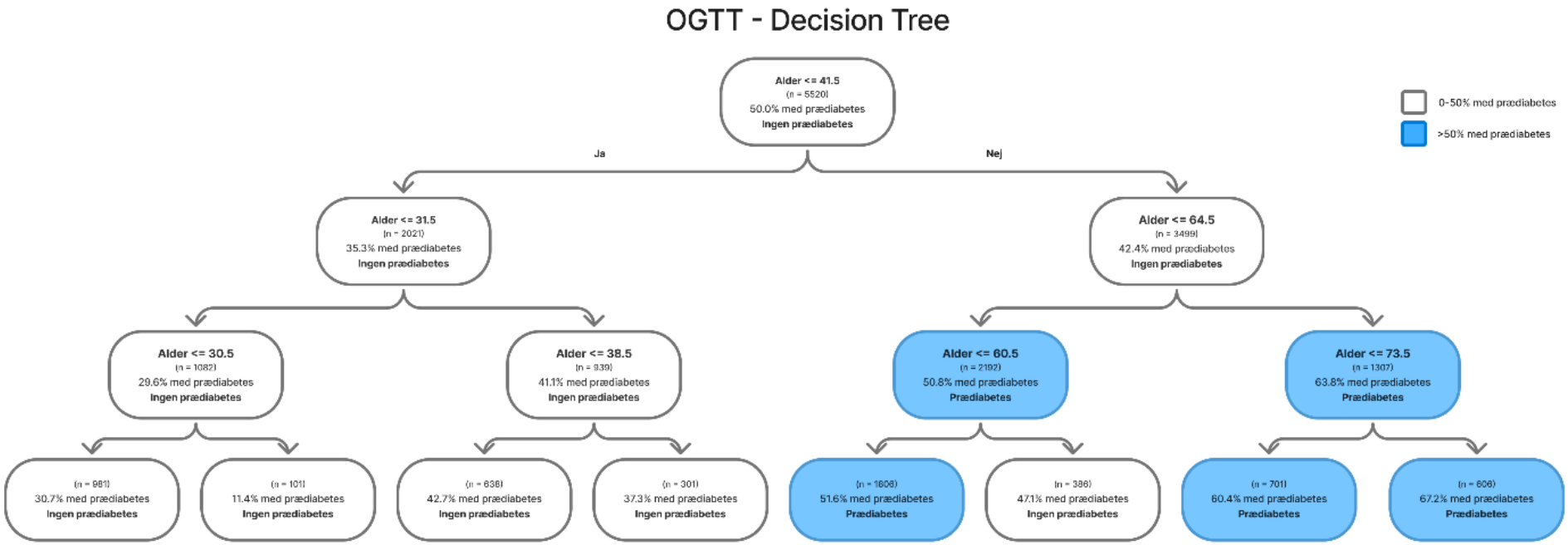
Figur 6: *Decision Tree for FPG-modellen (uden class_weight), trænet med max_depth = 5 og min_samples_leaf = 2. Modellen er baseret på talje-højde-ratio og køn. Nodefarverne er tilpasset en justeret klassifikations-grænse på 0.3 og viser andelen med prædiabetes, mens tekstetiketter følger Scikit-learns standardgrænse på 0.5.*

HbA1c-modellen er baseret på én inputvariabel: alder. Træet indeholder sekventielle split på forskellige aldersgrænser. Første opdeling sker ved 42,5 år efterfulgt af yderligere opdelinger i aldersbaserede intervaller.



Figur 7: *Decision Tree for HbA1c-modellen (uden class_weight), trænet med max_depth = 3 og min_samples_leaf = 2. Modellen er baseret på alder som eneste inputvariabel. Nodefarverne er tilpasset en justeret klassifikationsgrænse på 0.2 og viser andelen med prædiabetes, mens tekstetiketter følger Scikit-learns standardgrænse på 0.5.*

OGTT-modellen er ligeledes baseret på én inputvariabel: alder. Første opdeling sker ved 41,5 år, og træet indeholder flere aldersbaserede split, herunder omkring 60–65 år. Flere slutnoder er domineret af prædiabetesklassifikation.



***Figur 8**: Decision Tree for OGTT-modellen (class_weight="balanced"), trænet med max_depth = 3 og min_samples_leaf = 2. Modellen er baseret på alder som eneste inputvariabel. Nodefarverne afspejler en klassifikationsgrænse på 0.5 og viser andelen med prædiabetes, mens tekstetiketter svarer til Scikit-learns standardgræns*

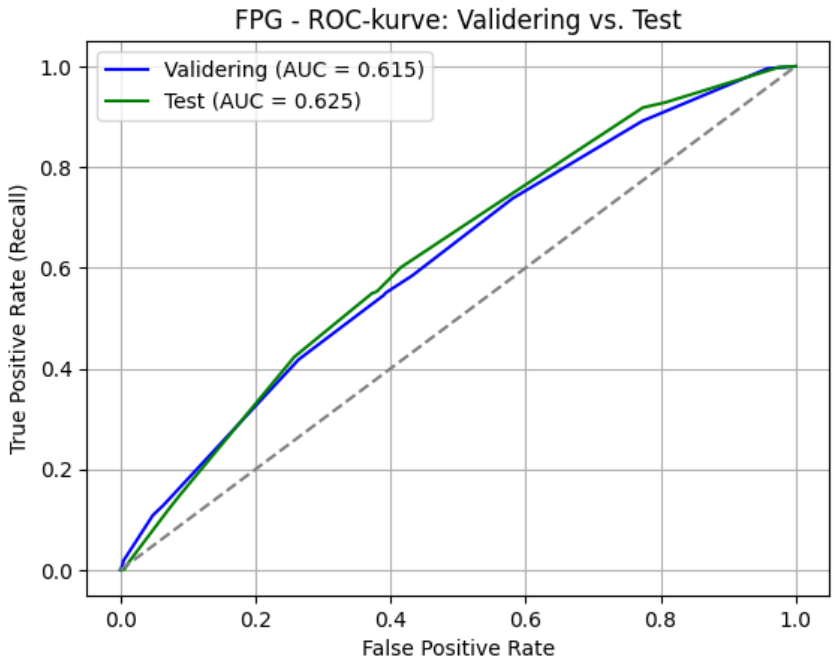
3.5 Performance

3.5.1 ROC-kurver og AUC

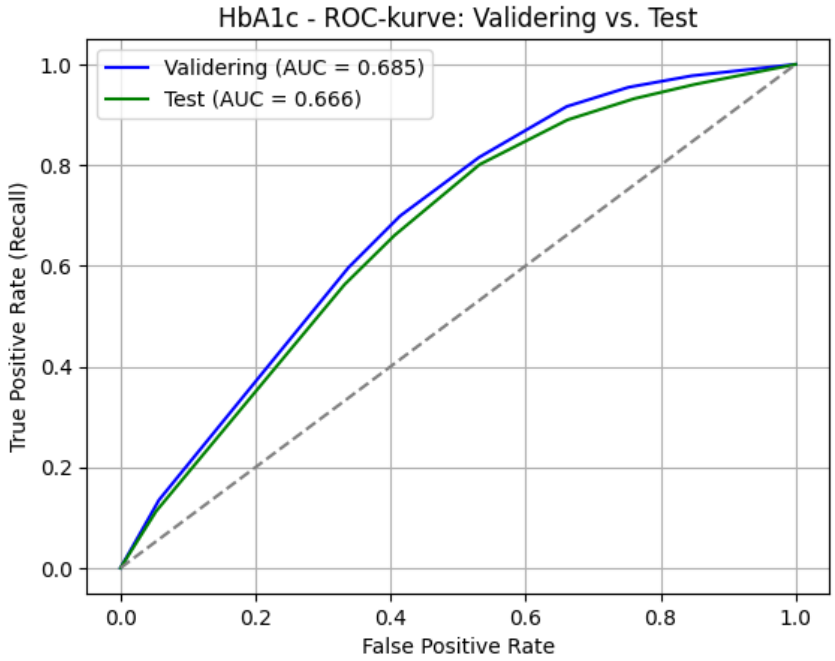
Modellernes evne til at skelne mellem deltagere med og uden prædiabetes blev vurderet ved hjælp af ROC-kurver og tilhørende AUC. Hver kurve viser forholdet mellem false positive rate (x-akse) og true positive rate (recall, y-akse) ved varierende thresholds.

For hver af de tre modeller er ROC-kurverne vist i figur 9-11. Både validerings- og testkurver er præsenteret i samme plot for at muliggøre vurdering af overfitting. AUC-værdierne fremgår af figurerne og er opsummeret i tabel 16.

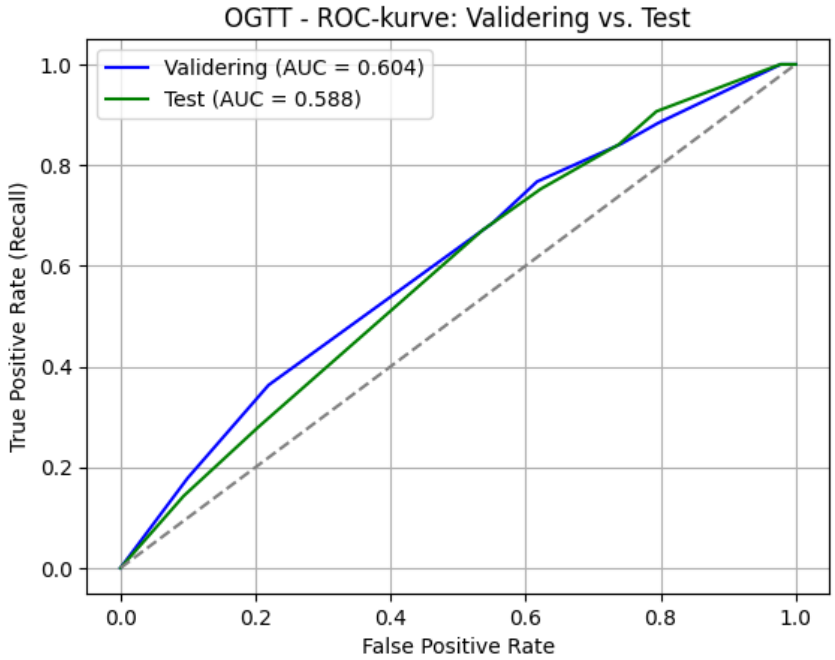
Forskellen mellem AUC på validerings- og testdatasættet var lav (0.01–0.012) for alle tre modeller. Da ROC-kurverne også lå tæt, tydede det ikke på overfitting i nogen af modellerne.



Figur 9: *ROC-kurve for FPG-modellen. Figuren viser forholdet mellem false positive rate og true positive rate (recall) ved varierende thresholds. AUC = 0.625.*



Figur 10: *ROC-kurve for HbA1c-modellen. Figuren viser forholdet mellem false positive rate og true positive rate (recall) ved varierende thresholds. AUC = 0.666.*



Figur 11: *ROC-kurve for OGTT-modellen. Figuren viser forholdet mellem false positive rate og true positive rate (recall) ved varierende thresholds. AUC = 0.588.*

3.5.2 Klassifikationsresultater

Modelperformance blev vurderet på uafhængige testdatasæt for hver af de tre modeller. For hver model blev den threshold anvendt, der resulterede i den højeste F1-score i testdatasættet, hvorefter performance blev opsummeret i tilhørende confusion matrix i tabel 13-15. En systematisk afprøvning af thresholds fra 0.1 til 0.9 og de tilhørende performancemetrikker fremgår af bilag 10 og 11.

Confusion matrix: FPG-modellen		
	Predicted negative	**Predicted positive**
Actual negative	TN = 224	FP = 766
Actual positive	FN = 48	TP = 538

Tabel 13: *Confusion matrix for FPG-modellen baseret på testdatasættet ved threshold 0.30. Forkortelser: TN = True Negative, FP = False Positive, FN = False Negative, TP = True Positive.*

Confusion matrix: HbA1c-modellen		
	Predicted negative	**Predicted positive**
Actual negative	TN = 1119	FP = 1272
Actual positive	FN = 175	TP = 892

Tabel 14: *Confusion matrix for HbA1c-modellen baseret på testdatasættet ved threshold 0.20. Forkortelser: TN = True Negative, FP = False Positive, FN = False Negative, TP = True Positive.*

Confusion matrix: OGTT-modellen		
	Predicted negative	**Predicted positive**
Actual negative	TN = 556	FP = 642
Actual positive	FN = 60	TP = 122

Tabel 15: *Confusion matrix for OGTT-modellen baseret på testdatasættet ved threshold i 0.5. Forkortelser: TN = True Negative, FP = False Positive, FN = False Negative, TP = True Positive.*

3.5.3 Modellernes performance

De samlede performancemetrikker for alle tre modeller, herunder F1-score, accuracy, precision, recall, specificity og AUC, fremgår af tabel 16. Recall og specificity fremhæves, da de begge relaterer til modellernes evne til at klassificere korrekt. FPG-modellen havde en recall på 0.918 og en specificity på 0.226. HbA1c-modellen havde en recall på 0.801 og en specificity på 0.468. OGTT-modellen havde en recall på 0.670 og en specificity på 0.464. Valgte thresholds for klassifikation fremgår også af tabellen.

	Thresholds	F1-score	Accuracy	Precision	Recall	Specificity	AUC
FPG	0.30	0.569	0.484	0.413	0.918	0.226	0.625
HbA1c	0.20	0.494	0.558	0.357	0.801	0.468	0.666
OGTT	0.50	0.258	0.491	0.160	0.670	0.464	0.588

Tabel 16: *Oversigt over performancemetrikker for de tre Decision Trees beregnet på testdatasættet. Tabellen viser threshold, F1-score, accuracy, precision, recall, specificity og AUC for hver model*

4. Resultatdiskussion

4.1 Sammenligning af modellernes performance

4.1.1 Overordnet performance og metrikker

På tværs af de tre modeller blev der observeret forskelle i klassifikationsperformance, selvom modellerne byggede på samme metode og udelukkende ikke-laboratorieafhængige input. Recall var overordnet høj (0.670–0.918), hvilket kunne indikere, at modellerne i høj grad formåede at identificere personer med prædiabetes. Specificity var derimod lavere (0.226–0.468), hvilket tyder på en vis tendens til at klassificere personer uden prædiabetes som falsk positive. Den samlede diskriminationsevne målt ved AUC lå i intervallet 0.588–0.666, hvilket peger på en lav adskillelse mellem de to klasser (87). F1-score varierede fra 0.258 til 0.569 og illustrerer, at modellernes balance mellem precision og recall var begrænset, særligt for OGTT-modellen.

4.1.2 Klinisk perspektiv: Afvejning mellem recall og specificity

I en screeningskontekst kan det argumenteres for, at recall bør tillægges særlig vægt, da falsk negative prædiktioner indebærer risiko for, at personer med prædiabetes forbliver uopdagede. Dette kan have konsekvenser for sygdomsudviklingen, idet ubehandlet prædiabetes øger risikoen for progression til T2D og senkomplikationer (11). En høj recall kan derfor anses som en styrke, også selvom det opnås på bekostning af specificity.

Abbas et al, som udviklede ML-modeller til detektion af prædiabetes baseret på enkle, ikke-invasive variable, fremhæver, at en høj andel falsk positive i screeningskontekst ofte er acceptabel, når det drejer sig om tilstande som prædiabetes, som ikke er livstruende på kort sigt. Bekræftende laboratorietests er relativt billige og minimalt invasive, og omkostningerne ved overdiagnostik vurderes lavere end risikoen ved at overse prædiabetes (65). Dette perspektiv understøtter relevansen af at favorisere modeller med høj recall, også når specificity reduceres. Denne afvejning mellem recall og precision var generelt til stede i specialets modeller og tydeligst i FPG-modellen, som opnåede den højeste recall (0.918), men samtidig den laveste specificity (0.226). Det illustrerer, at valget af threshold og vægtningen af performancemetrikker bør afspejle det kliniske formål med modellen.

Derudover påpeger Abbas et al., at modeller udviklet til prædiabetes også potentielt kan bidrage til opsporing af udiagnosticeret T2D, når højrisikopersoner efterfølgende tilbydes bekræftende blodprøver (65). Selvom specialets modeller ikke er klar til klinisk implementering, peger dette perspektiv på

en mulig sekundær gevinst ved at identificere individer med høj prædiktiv sandsynlighed, selv i tilfælde af falsk positive.

4.1.3 Definition af outcome og modelbegrænsninger

Den lavere performance i OGTT-modellen (F1-score: 0.258; AUC: 0.588; recall: 0.670) kan dels skyldes, at OGTT identificerer en bredere og mere heterogen gruppe af personer med prædiabetes end HbA1c og FPG (8), hvilket kan udfordre modellens evne til at lære konsistente mønstre, særligt i en enkel modelstruktur med én inputvariabel.

Samtidig peger specialets resultater på et bredere metodisk problem: Prædiabetes udgør en klinisk gråzone, hvor det kan være vanskeligt at adskille personer med prædiabetes fra personer med normoglykæmi, især når der udelukkende anvendes selvrapporterede data. Et systematisk review af Barry et al. fremhæver, at de diagnostiske tests til prædiabetes ofte identificerer forskellige undergrupper med begrænset overlap og svingende præcision på tværs af populationer (14). Denne observation bekræftes af Deberneh og Kim, som i en multiclass-model fandt, at prædiabetesklassen havde markant lavere precision (0.61) end både diabetes (0.90) og normoglykæmi (0.70), hvilket blev tilskrevet betydeligt overlap i datastrukturen (100).

Tilsammen understøtter disse fund, at prædiabetes ikke blot er vanskeligt at definere klinisk, men også udgør en særskilt prædiktiv udfordring i ML-modeller. Dette kan være en væsentlig forklaring på den begrænsede modelperformance i dette speciale, særligt for OGTT-modellen.

4.2 Sammenligning med eksisterende modeller

Efter de indledende resultater og kliniske overvejelser præsenteres her en samlet vurdering af, hvordan specialets modeller placerer sig i forhold til eksisterende ML-studier til detektion af prædiabetes. De eksisterende modeller indikerer generelt højere klassifikationsperformance end i dette speciale, hvilket rejser spørgsmålet om hvilke forhold, der kan forklare forskellene.

Choi et al. udviklede ML-modeller til detektion af prædiabetes baseret på data fra The Korea National Health and Nutrition Examination Survey (KNHANES) og ikke-laboratorieafhængige variable (67), hvilket gør studiet metodisk sammenligneligt med dette speciale. De anvendte FPG som outcome og opnåede bedre klassifikationsperformance (AUC: 0.712–0.731) end specialets modeller (AUC: 0.588–0.666), men benyttede black box-modeller, som generelt har lav explainability og begrænset klinisk transparens (67). Trods dette var deres performanceforbedring moderat, hvilket antyder, at prædiktionsopgaven i sig selv kan være kompleks, også når der anvendes mere avancerede algoritmer.

Dette underbygges delvist af Dinh et al., som udviklede ML-modeller til detektion af prædiabetes og diabetes baseret på NHANES-data fra 1999–2014. Studiet anvendte en række black box-algoritmer

(herunder ensemblemetoder) og inkluderede 123 inputvariable (57), hvilket adskiller sig væsentligt fra dette speciales metodiske valg om et lille, explainable feature-set. Deres outcome omfattede både prædiabetes og diabetes som én samlet klasse (57), hvilket kan have reduceret opgavens kompleksitet, da kontrasten mellem klasserne blev større. På trods af brugen af avancerede modeller opnåede de kun moderat performance (AUC: 0.731–0.737) på niveau med simpel logistisk regression. Forfatterne konkluderede, at yderligere modelkompleksitet ikke nødvendigvis forbedrer klassifikationsevnen ved denne type opgave (57), hvilket understøtter observationerne i dette speciale. Forskelle i definering af outcome (kombineret prædiabetes og diabetes vs. prædiabetes alene), antallet af inputvariable (123 vs. 1–2) og datagrundlag (1999–2014 vs. 2011–2016) (57) kan samlet bidrage til at forklare, hvorfor Dinh et al. opnåede højere klassifikationsperformance end i dette speciale.

4.2.1 Sammenligning af prædiktive variable

Blandt de udvalgte variable via forward feature selection i dette speciale var alder den eneste, der indgik i både HbA1c- og OGTT-modellen. At alder blev identificeret som prædiktiv i to af modellerne og tillige fremhæves i tidligere studier (57,65,66), tyder på en høj grad af reproducerbarhed og generaliserbarhed for netop denne prædiktor.

Birk et al. udviklede prædiktive modeller til detektion af prædiabetes med henblik på anvendelse i screeningssammenhæng. De fandt, at alder alene var en så stærk prædiktor, at en simpel model baseret på alder opnåede en AUC > 0.70, hvilket understøtter generaliserbarheden af netop denne variabel (66). Dinh et al. inkluderede alder blandt de mest centrale variable i deres modeller baseret på NHANES data (57,65), og Abbas et al. fandt, at alder ≥55 år var den stærkest vægtede faktor (65).

Derimod blev flere variable, som ellers er veldokumenterede risikofaktorer for prædiabetes, ikke inkluderet i de endelige modeller i dette speciale, trods deres tilstedeværelse i den oprindelige feature set. Det gælder eksempelvis BMI, alkoholforbrug, rygning og blodtryk, som alle blev fravalgt under forward feature selection, da de ikke forbedrede AUC tilstrækkeligt i specialets modeller. Dette kan indikere, at modellernes prædiktive kapacitet blev begrænset af det smalle og selekterede feature set.

Det er dog bemærkelsesværdigt, at netop disse variable er blevet anvendt i tidligere studier. Abbas et al. inkluderede BMI, taljemål og blodtryk (65), Choi et al. anvendte alkoholforbrug og BMI (67) og Birk et al. anvendte alkohol og rygning (66).

At de samme variable har bidraget til forbedret performance i disse studier, kan skyldes, at de alle benyttede black box-modeller, som har større kapacitet til at identificere ikke-lineære relationer og interaktioner mellem variable. Decision Trees, som blev anvendt i dette speciale, kan være begrænsede i deres evne til at modellere sådanne sammenhænge. Samtidig kan fravalget af disse variable også være knyttet til dataspecifikke forhold, for eksempel lav svag sammenhæng med outcome i den undersøgte

periode (NHANES 2011–2016). Det understreger, at fravær af prædiktiv værdi i én model ikke nødvendigvis afspejler en variabels manglende kliniske relevans, men snarere dens kontekstafhængige nytte i netop denne modeltype og datastruktur.

5. Metodediskussion

5.1 Begrænsninger ved datagrundlaget

5.1.1 Vægtning go repræsentativitet

NHANES er designet til at være repræsentativ for den ikke-institutionaliserede, civile befolkning i USA gennem et komplekst stikprøvedesign med stratificering, klyngeudvælgelse og oversampling af subgrupper som personer med lav indkomst, ældre og visse etniske minoriteter (101).

Oversampling muliggør mere præcise estimater for disse grupper, men det indebærer samtidig, at analyser bør vægtjusteres for at afspejle populationens faktiske sammensætning (101). Da formålet i dette speciale var klassifikation og detektion af prædiabetes snarere end estimering af populationsparametre, blev fravalget af vægtning vurderet som metodisk acceptabelt. Det skal dog understreges, at fravalget kunne medføre en skævhed i modellens læring, da visse grupper, for eksempel lavindkomst eller specifikke etniciteter, kunne indebære overrepræsentation i træningsdata. Dette kunne have påvirket modellens generaliserbarhed til populationer med en anden demografisk sammensætning.

5.1.2 Ekstern validitet og kontekstafhængighed

Modeller udviklet i én kontekst bør testes i andre populationer for at vurdere ekstern validitet og sikre generaliserbarhed (74). En central metodisk overvejelse er, hvorvidt målet med ML-modeller bør være én model, der generaliserer bredt, eller snarere metoder, der kan anvendes til at udvikle lokale modeller på kontekstspecifikke data.

Wiens og Shenoy, som i et studie præsenterer potentialet for anvendelse af ML i sundhedsfaglig kontekst argumenterer for det sidstnævnte (88). At der bør tilstræbes generaliserbare metoder frem for universelle modeller, da sundhedsdata ofte er præget af institutions- og kontekstafhængige variationer.

I dette speciale blev der tilstræbt en enkel, explainable modelstruktur og udelukkende ikke-laboratoriebaserede input, og kan dermed gentrænes på lokale datasæt, hvilket øger modellens anvendelighed uden for en amerikansk kontekst.

5.1.3 Dataperiode og retrospektivitet

Der blev anvendt laboratoriedefinerede outcomevariable fra perioden 2011–2016 for at sikre et tilstrækkeligt stort og stabilt datagrundlag samt mulighed for at inkludere OGTT. Ifølge Johnson et al., som beskriver design og udvælgelsesmetoder i NHANES, herunder principper for datasammenlægning og stikprøveudvælgelse, bør data kombineres over mindst fire år for at muliggøre robuste analyser

(101). Den valgte periode i specialet blev derfor vurderet som et hensigtsmæssigt kompromis mellem datamængde og aktualitet.

Det er dog væsentligt at anerkende, at ML-modeller baseret på retrospektive sundhedsdata kan miste validitet over tid, hvis klinisk praksis, patientkarakteristika eller dataindsamlingsmetoder ændrer sig (102). Da NHANES ikke opdateres i realtid, og dette speciale anvendte data fra 2011–2016, afspejler modellerne derfor sundhedsforhold i denne periode. De bør derfor anvendes med forsigtighed i nyere eller ændrede kontekster.

5.1.4 Fravær af kliniske risikofaktorer

Et væsentligt metodisk forhold med potentiel betydning for modellernes prædiktive potentiale er fraværet af flere velkendte risikofaktorer for prædiabetes. Talje-hofte-ratio, som i flere studier har vist høj diskriminativ styrke (28,29), kunne ikke inkluderes, da NHANES ikke indeholder hoftemål i de relevante årgange. Fysisk inaktivitet (39) blev udeladt, da variablen ikke blev indsamlet i 2013–2014, hvilket medførte et stort datatab. Kostrelaterede variable (30,42) og familiær disposition (42) blev fravalgt, da det ikke kunne operationaliseres på en meningsfuld måde.
Dette fravalg genfindes i andre ML-studier. Abbas et al. undlod at inkludere familiær disposition og livsstilsvariable som fysisk aktivitet, idet disse ofte er præget af recall bias, lav datakvalitet eller manglende viden hos deltagerne (65). Studiet fremhæver, at mange individer ikke kender deres familiemedlemmers sygdomshistorik, og at visse adfærdsdata kan være kulturelt følsomme og dermed upålidelige (65). På tilsvarende vis blev moderat depression (44) i dette speciale udeladt, da data var baseret på selvrapportering snarere end klinisk diagnosticering. Udeladelserne blev truffet for at bevare sammenlignelighed på tværs af årgange og sikre robusthed, men kan samtidig have reduceret modellernes evne til at identificere komplekse risikoprofiler.

5.2 Valg af outcome og klinisk anvendelighed

Modellerne i dette speciale er trænet til at forudsige resultater af specifikke laboratorietests og ikke nødvendigvis den samlede kliniske tilstand prædiabetes. Som Wiens og Shenoy pointerer, lærer en ML-model at forudsige præcis det outcome, den trænes på, og dens anvendelighed kan derfor være følsom over for ændringer i testpraksis (88). Dette er særligt relevant i en dansk kontekst, hvor HbA1c anvendes som primær screeningsmetode for T2D i almen praksis, mens OGTT kun benyttes ved specifik indikation (103,104). Der findes ikke nationale retningslinjer for systematisk screening for prædiabetes i Danmark, hvilket yderligere understreger behovet for definitioner af outcome med høj klinisk relevans og overførbarhed. På den baggrund vurderes HbA1c-modellen at have størst metodisk og klinisk relevans, da

den allerede anvendes i klinisk praksis til både screening og monitorering, hvilket gør det outcome mere robust over for kortsigtede variationer. Samtidig bør det bemærkes, at selv laboratoriebaserede outcomes som FPG ikke er uden metodologiske udfordringer. Birk et al. fremhæver, at FPG-målinger kan være forbundet med målefejl, særligt hvis fasteforholdene ikke er overholdt, og at FPG alene ikke nødvendigvis giver et fuldstændigt billede af den glykæmiske tilstand (66). Dette kan have betydning for FPG-modellens træningsgrundlag og klassifikationspræcision, og peger yderligere på fordelene ved HbA1c som outcome.

5.3 Feature selection

Decision Trees foretager implicit feature selection ved at vælge de variable, der bedst reducerer impurity ved hver split (105). Forud for træningen blev der dog gennemført en separat forward feature selection for at identificere de mest prædiktive variable og samtidig sikre et lavt antal inputvariable, hvilket kan styrke modellernes explainability og kliniske anvendelighed.

Forward selection er en beregningsmæssigt effektiv metode, som rangerer variable ud fra deres univariat AUC og tilføjer dem sekventielt, så længe de forbedrer modelperformance. Denne tilgang kan overse potentielle synergier mellem variable, hvilket kan føre til udeladelse af klinisk relevante kombinationer (96,106). I dette speciale inkluderede HbA1c- og OGTT-modellerne kun én variabel hver, hvilket umiddelbart kan styrke explainability. Det er dog vigtigt at understrege, at et lavt antal inputvariable ikke nødvendigvis medfører høj explainability i praksis, særligt hvis Decision Trees stadig indeholder mange noder og komplekse splits. I dette tilfælde resulterede de endelige Decision Trees, trods få inputvariable, i strukturer med flere lag og forgreninger, hvilket begrænser deres praktiske transparens. Det stiller spørgsmål ved, hvorvidt modellerne reelt er lette at fortolke for klinikere, hvilket ellers er et af de primære argumenter for valg af Decision Trees i sundhedsfaglige kontekst.

Et metodisk alternativ kunne være backward elimination, som starter med den fulde model og fjerner én variabel ad gangen baseret på dens marginale bidrag. Til forskel fra forward selection tager metoden højde for interaktioner, hvilket kan være særlig relevant i sundhedsdata, hvor risikoprofiler ofte er multifaktorielle. Da prædiabetes ofte karakteriseres af kombinationer af let forhøjede værdier snarere end ekstreme værdier på én markør (107,108), kan modeller med bredere inputgrundlag være bedre egnet til at fange disse mønstre.

5.4 Modellens struktur og træning

Decision Trees er tilbøjelige til overfitting, særligt ved mange inputvariable. Risikoen herfor blev søgt reduceret gennem en hold-out-valideringsstrategi med en 80/20 opdeling i trænings- og testdata,

hvilket giver en robust vurdering af generaliserbarhed (88). Eftersom træningen er baseret på en samlet periode uden tidsopdeling, vurderes den interne validitet som god, mens den tidsmæssige robusthed forbliver uafprøvet. Dette understreger behovet for ekstern validering eller opdatering af modellerne på nyere datasæt.

5.5 Valg af model

I dette speciale blev et enkeltstående Decision Tree anvendt som modelstruktur. Valget blev truffet for at øge explainability og transparens, hvilket anses som centralt ved udvikling af kliniske beslutningsstøtteværktøjer. I modsætning til mere komplekse modeller gør et enkelt Decision Tree det muligt at følge den prædiktive logik trin for trin og dermed identificere præcise thresholds og beslutningsveje. Decision Trees er dog kendt for deres følsomhed over for overfitting og har ofte lavere prædiktiv styrke end ensemblemodeller. Random Forest er et eksempel på en ensemblemetode, som kombinerer output fra mange træer for at forbedre robusthed og performance. Typisk performer ensemblemetoder bedre end enkeltstående træer i sundhedsdomæner (51). På trods af dette blev de ikke valgt i dette speciale, da formålet var at udvikle en model, der ikke blot havde høj precision, men også kunne anvendes og forstås i klinisk praksis uden black-box elementer. Fravalget af ensemblemetode afspejler dermed en metodisk prioritering af explainability over maksimal performance.

6. Konklusion

Dette speciale undersøgte, hvordan explainable ML kan anvendes til at detektere prædiabetes udelukkende baseret på ikke-laboratorieafhængige risikofaktorer. Med udgangspunkt i data fra NHANES 2011–2016 blev der udviklet tre separate Decision Tree modeller, hvor prædiabetes blev defineret ved henholdsvis HbA1c, FPG og OGTT. Valget af Decision Tree som algoritme var motiveret af ønsket om høj explainability og klinisk anvendelighed, og feature selection blev gennemført med henblik på at identificere de mest informative og robuste inputvariable.

Resultaterne viste, at modellernes prædiktive evne varierede afhængigt af definitionen af outcome. Performance var generelt lav målt ved AUC, men sensitiviteten var høj i FPG- og HbA1c-modellerne, hvilket er centralt i en screeningskontekst, hvor formålet er at identificere så mange risikopersoner som muligt. Den lave AUC kan blandt andet tilskrives begrænset inputdata, anvendelsen af simple modeller samt den kliniske heterogenitet, der kendetegner prædiabetes, især når outcome defineres ud fra forskellige laboratorietests.

På trods af modellernes begrænsede prædiktive styrke bidrager specialet med flere vigtige indsigter. For det første understøtter resultaterne, at enkelte ikke-laboratorieafhængige variable som alder og talje-højde-ratio konsekvent blev udvalgt som prædiktivt relevante på tværs af definitioner af prædiabetes. For det andet illustrerer specialet, at en models præcision er tæt knyttet til definitionen af outcome og datagrundlagets karakteristika.

Samlet set indikerer specialets resultater, at explainable ML kan anvendes til at identificere individer med øget risiko for prædiabetes på baggrund af tilgængelige, ikke-invasive data. Dog bør modellernes præcision forbedres, før de kan anbefales til bred implementering. Fremtidige studier med større datamængder, mere komplette variable og alternative algoritmer kan bidrage til at styrke modellernes kliniske anvendelighed og prædiktive præcision.

7. Perspektivering

Selvom de udviklede modeller i dette speciale ikke er implementeringsklare pga. lav prædiktiv performance, er det relevant at overveje perspektiver for fremtidig anvendelse, forudsat at modellerne videreudvikles og valideres.

7.1 Implementering i praksis

For at en explainable ML-model til opsporing af prædiabetes kan implementeres effektivt i klinisk praksis, er det nødvendigt at sikre en meningsfuld og brugervenlig integration. Dette kræver mere end blot høj prædiktiv præcision (108). Modellen skal være intuitiv at anvende og let at fortolke, samtidig med at dens kliniske relevans anerkendes af sundhedsprofessionelle.

For at fremme anvendelsen kan det være hensigtsmæssigt at etablere støttestrukturer. Herunder teams af specialister, som kan bistå ved implementering, oplæring og drift (109,110). Sådanne teams bør inkludere både klinikere, dataloger og sundhedsteknologiske eksperter, da udvikling og anvendelse af ML-modeller ikke bør foregå isoleret inden for én faglighed. Som beskrevet af Wiens et al. kræver meningsfuld anvendelse af ML i sundhedsvæsenet tæt samarbejde på tværs af discipliner, og det bør ledsages af træning og organisatorisk forankring, hvis potentialet skal realiseres (88). Derudover er det centralt at kortlægge og tage højde for sundhedsprofessionelles præferencer og teknologiske vaner for at reducere modstand mod forandring (109,110).

En differentieret præsentation af modellen, eksempelvis både i digitalt og papirbaseret format, kan tilbydes således at brugernes behov kan tilpasses efter kontekst. Dette er særligt relevant i primærsektoren, hvor behandlingsindsatsen ofte foregår tæt på borgerens hverdag og i kontekster med begrænsede ressourcer (111). Her kunne modellen fungere som et relevant støtteværktøj i fravær af ressourcekrævende redskaber eller speciallægefaglig supervision. Alternativt kunne det indgå i et mobilt sundhedstilbud, der opsøger borgere og tilbyder frivillig screening udenfor etablerede kliniske rammer.

7.2 Vedligeholdelse og evaluering

Efter implementering bør modellen indgå i en systematisk strategi for løbende evaluering og vedligeholdelse for at understøtte dens langsigtede driftssikkerhed (112). I takt med ændringer i datagrundlag, kliniske retningslinjer eller organisatoriske strukturer i sundhedsvæsenet, kan det være nødvendigt at kalibrere og opdatere modellen løbende. Der bør opretholdes uafhængige test- og evalueringer, som anvendes på nye- og kontekstspecifikke data, for at vurdere modellens ydeevne og bevare dens kliniske relevans (88).

For at en fremtidig model kan overvejes til implementering, kræver det en væsentlig forbedring i prædiktiv præcision, ekstern validering og dokumenteret klinisk effekt. Først da vil modellen kunne vurderes som et reelt supplement til eksisterende screeningsmetoder.

8. Referenceliste

1. International Diabetes Federation [Internet]. [henvist 7. april 2025]. Facts and Figures. Tilgængelig hos: https://idf.org/about-diabetes/diabetes-facts-figures/

2. Diabetesforeningen [Internet]. [henvist 7. april 2025]. Diabetestal. Tilgængelig hos: https://www.diabetestal.nu/

3. Carstensen B, Rønn PF, Jørgensen ME. Prevalence, incidence and mortality of type 1 and type 2 diabetes in Denmark 1996-2016. BMJ Open Diabetes Res Care [Internet]. 1. maj 2020 [henvist 7. april 2025];8(1). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/32475839/

4. Carstensen B, Rønn PF, Jørgensen ME. Components of diabetes prevalence in Denmark 1996-2016 and future trends until 2030. BMJ Open Diabetes Res Care [Internet]. 11. august 2020 [henvist 7. april 2025];8(1). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/32784246/

5. Lima JEBF, Moreira NCS, Sakamoto-Hojo ET. Mechanisms underlying the pathophysiology of type 2 diabetes: From risk factors to oxidative stress, metabolic dysfunction, and hyperglycemia. Mutat Res Genet Toxicol Environ Mutagen [Internet]. 1. februar 2022 [henvist 7. april 2025];874–875. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/35151421/

6. Steno Diabetes Center. https://sdco.dk/om-os/presse/fakta-om-diabetes. [henvist 31. maj 2025]. Fakta om diabetes. Tilgængelig hos: https://sdco.dk/om-os/presse/fakta-om-diabetes

7. Magliano DJ, Boyko EJ, committee IDA 10th edition scientific. IDF DIABETES ATLAS, 10th edition. IDF DIABETES ATLAS [Internet] Brussels [Internet]. 2021 [henvist 7. april 2025];1–141. Tilgængelig hos: https://www.ncbi.nlm.nih.gov/books/NBK581934/

8. Committee ADAPP, ElSayed NA, Aleppo G, Bannuru RR, Bruemmer D, Collins BS, m.fl. 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2024. Diabetes Care [Internet]. 1. januar 2024 [henvist 27. maj 2025];47(Supplement_1):S20–42. Tilgængelig hos: https://dx.doi.org/10.2337/dc24-S002

9. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. Lancet [Internet]. 2012 [henvist 7. april 2025];379(9833):2279–90. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/22683128/

10. American Diabetes Association [Internet]. [henvist 7. april 2025]. Understanding Diabetes Diagnosis. Tilgængelig hos: https://diabetes.org/about-diabetes/diagnosis

11. Echouffo-Tcheugui JB, Selvin E. Prediabetes and What It Means: The Epidemiological Evidence. Annu Rev Public Health [Internet]. 1. april 2021 [henvist 7. april 2025];42:59–77. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/33355476/

12. Bansal N. Prediabetes diagnosis and treatment: A review. World J Diabetes [Internet]. 2015 [henvist 7. april 2025];6(2):296. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/25789110/

13. Punthakee Z, Goldenberg R, Katz P. Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. Diabetes Canada [Internet]. 2018 [henvist 7. april 2025]; Tilgængelig hos: https://guidelines.diabetes.ca/cpg/chapter3

14. Barry E, Roberts S, Oke J, Vijayaraghavan S, Normansell R, Greenhalgh T. Efficacy and effectiveness of screen and treat policies in prevention of type 2 diabetes: Systematic review and meta-analysis of screening tests and interventions. BMJ (Online) [Internet]. 2017 [henvist 29. maj 2025];356. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/28052845/

15. McGraw K, Lee J. Glucose Intolerance. The 5-Minute Clinical Consult Standard 2016: Twenty Fourth Edition [Internet]. 8. august 2023 [henvist 30. april 2025]; Tilgængelig hos: https://www.ncbi.nlm.nih.gov/books/NBK499910/

16. Echouffo-Tcheugui JB, Perreault L, Ji L, Dagogo-Jack S. Diagnosis and Management of Prediabetes: A Review. JAMA [Internet]. 11. april 2023 [henvist 31. maj 2025];329(14):1206–16. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/37039787/

17. Meijnikman AS, De Block CEM, Dirinck E, Verrijken A, Mertens I, Corthouts B, m.fl. Not performing an OGTT results in significant underdiagnosis of (pre)diabetes in a high risk adult Caucasian population. Int J Obes (Lond) [Internet]. 1. november 2017 [henvist 7. april 2025];41(11):1615–20. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/28720876/

18. National Diabetes Statistics Report [Internet]. 2014. Tilgængelig hos: http://diabetes.niddk.nih.gov/dm/pubs/comparingtests/index.aspx

19. Jørgensen ME, Ellervik C, Ekholm O, Johansen NB, Carstensen B. Estimates of prediabetes and undiagnosed type 2 diabetes in Denmark: The end of an epidemic or a diagnostic artefact? Scand J Public Health [Internet]. 1. februar 2020 [henvist 7. april 2025];48(1):106–12. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/30222048/

20. Diabetesforeningen [Internet]. 2024 [henvist 7. april 2025]. Diabetes i tal. Tilgængelig hos: https://diabetes.dk/forskning/viden-om-diabetes/diabetes-i-tal-2024

21. Dall TM, Yang W, Gillespie K, Mocarski M, Byrne E, Cintina I, m.fl. The Economic Burden of Elevated Blood Glucose Levels in 2017: Diagnosed and Undiagnosed Diabetes, Gestational Diabetes Mellitus, and Prediabetes. Diabetes Care [Internet]. 1. september 2019 [henvist 7. april 2025];42(9):1661–8. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/30940641/

22. Jiang L, Johnson A, Pratte K, Beals J, Bullock A, Manson SM. Long-term Outcomes of Lifestyle Intervention to Prevent Diabetes in American Indian and Alaska Native Communities: The Special Diabetes Program for Indians Diabetes Prevention Program. Diabetes Care [Internet]. 1. juli 2018 [henvist 7. april 2025];41(7):1462–70. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/29915128/

23. Yuen A, Sugeng Y, Weiland TJ, Jelinek GA. Lifestyle and medication interventions for the prevention or delay of type 2 diabetes mellitus in prediabetes: a systematic review of randomised controlled trials. Aust N Z J Public Health [Internet]. april 2010 [henvist 7. april 2025];34(2):172–8. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/23331362/

24. Nathan DM, Davidson MB, DeFronzo RA, Heine RJ, Henry RR, Pratley R, m.fl. Impaired fasting glucose and impaired glucose tolerance: implications for care. Diabetes Care [Internet]. marts 2007 [henvist 7. april 2025];30(3):753–9. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/17327355/

25. Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, Pedersen L. Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark. BMJ Open Diabetes Res Care [Internet]. 16. september 2022 [henvist 7. april 2025];10(5). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/36113888/

26. Lee CMY, Colagiuri S, Woodward M, Gregg EW, Adams R, Azizi F, m.fl. Comparing different definitions of prediabetes with subsequent risk of diabetes: an individual participant data meta-analysis involving 76 513 individuals and 8208 cases of incident diabetes. BMJ Open Diabetes Res Care [Internet]. 29. december 2019 [henvist 7. april 2025];7(1). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/31908797/

27. Tourkmani AM, Alharbi TJ, Bin Rsheed AM, Alotaibi AF, AlEisaa M, Youzghadli IM, m.fl. Characteristics and risk factors associated with developing prediabetes in Saudi Arabia. Ann Med [Internet]. 1. december 2024 [henvist 7. april 2025];56(1):2413922. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/39392033/

28. Zhao X, Zhu X, Zhang H, Zhao W, Li J, Shu Y, m.fl. Prevalence of diabetes and predictions of its risks using anthropometric measures in southwest rural areas of China. BMC Public Health [Internet]. 2012 [henvist 7. april 2025];12(1). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/22998969/

29. Hou X, Chen S, Hu G, Chen P, Wu J, Ma X, m.fl. Stronger associations of waist circumference and waist-to-height ratio with diabetes than BMI in Chinese adults. Diabetes Res Clin Pract [Internet]. 1. januar 2019 [henvist 7. april 2025];147:9–18. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/30144478/

30. Nwatu CB, Ofoegbu EN, Unachukwu CN, Young EE, Okafor CI, Okoli CE. Prevalence of prediabetes and associated risk factors in a rural Nigerian community. Int J Diabetes Dev Ctries. 1. juni 2016;36(2):197–203.

31. Liberty IA, Septadina IS, Rizqie MQ, Ananingsih ES, Hasyim H, Sitorus RJ. Predictors of Prediabetes Among Communities Without a Family History of Type 2 Diabetes Mellitus: A Case-Control Study. Cureus [Internet]. 26. august 2023 [henvist 7. april 2025];15(8). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/37753042/

32. Alomari A, Al Hisnah S. Prevalence of Prediabetes and Associated Risk Factor Assessment Among Adults Attending Primary Healthcare Centers in Al Bahah, Saudi Arabia: A Cross-Sectional Study. Cureus [Internet]. 22. september 2022 [henvist 7. april 2025];14(9). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/36299950/

33. U.S. Centers for Disease Control and Prevention [Internet]. [henvist 24. april 2025]. National Diabetes Statistics Report. Tilgængelig hos: https://www.cdc.gov/diabetes/php/data-research/index.html

34. Vatcheva KP, Fisher-Hoch SP, Reininger BM, McCormick JB. Sex and age differences in prevalence and risk factors for prediabetes in Mexican-Americans. Diabetes Res Clin Pract [Internet]. 1. januar 2020 [henvist 7. april 2025];159. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/31805354/

35. Kontochristopoulou AM, Karatzi K, Karaglani E, Cardon G, Kivelä J, Wikström K, m.fl. Sociodemographic, anthropometric, and lifestyle correlates of prediabetes and type 2 diabetes in europe: The Feel4Diabetes study. Nutrition, Metabolism and Cardiovascular Diseases. 1. august 2022;32(8):1851–62.

36. Cheng YJ, Kanaya AM, Araneta MRG, Saydah SH, Kahn HS, Gregg EW, m.fl. Prevalence of Diabetes by Race and Ethnicity in the United States, 2011-2016. JAMA [Internet]. 24. december 2019 [henvist 23. april 2025];322(24):2389–98. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/31860047/

37. Chakkalakal RJ, Galaviz KI, Thirunavukkarasu S, Shah MK, Venkat Narayan KM. Test and Treat for Prediabetes: A Review of the Health Effects of Prediabetes and the Role of Screening and Prevention. Annu Rev Public Health [Internet]. 20. maj 2024 [henvist 7. april 2025];45(1):151–67. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/38109519/

38. Formagini T, Brooks JV, Roberts A, Bullard KMK, Zhang Y, Saelee R, m.fl. Prediabetes prevalence and awareness by race, ethnicity, and educational attainment among U.S. adults. Front Public

Health [Internet]. 2023 [henvist 7. april 2025];11:1277657. Tilgængelig hos: https://pmc.ncbi.nlm.nih.gov/articles/PMC10758124/

39. Mahat RK, Singh N, Arora M, Rathore V. Health risks and interventions in prediabetes: A review. Diabetes Metab Syndr [Internet]. 1. juli 2019 [henvist 7. april 2025];13(4):2803–11. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/31405710/

40. Duan Y, Fu H, Gao J, Wang S, Chen C, Zhao Y, m.fl. Associations of prediabetes and sleep duration, and inflammation as a mediator in the China Health and Retirement Longitudinal Study. Sleep Health [Internet]. 1. august 2024 [henvist 7. april 2025];10(4):470–7. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/38749824/

41. Huang Z, Deng J, Li H, Fang S, Wei Y, Lei W, m.fl. Prediabetes and sleep patterns: Linking poor sleep to adverse outcomes through metabolic syndrome. Diabetes Res Clin Pract [Internet]. 1. marts 2025 [henvist 7. april 2025];221. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/39956456/

42. Wagner R, Thorand B, Osterhoff MA, Müller G, Böhm A, Meisinger C, m.fl. Family history of diabetes is associated with higher risk for prediabetes: a multicentre analysis from the German Center for Diabetes Research. Diabetologia [Internet]. oktober 2013 [henvist 7. april 2025];56(10):2176–80. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/23979484/

43. Goergen AF, Ashida S, Skapinsky K, De Heer HD, Wilkinson A V., Koehly LM. What You Don't Know: Improving Family Health History Knowledge among Multigenerational Families of Mexican Origin. Public Health Genomics [Internet]. 1. april 2016 [henvist 7. april 2025];19(2):93–101. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/26854931/

44. Zhou Id J, Yang X. Association between depression and the prevalence and prognosis of prediabetes: Data from National Health and Nutrition Examination Survey (NHANES) 2013–2018. Klisic A, redaktør. PLoS One [Internet]. 13. januar 2025 [henvist 7. april 2025];20(1):e0304303. Tilgængelig hos: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0304303

45. Haile T. Literature review of Diabetes noninvasive screening tools for previously undiagnosed. Scoping review. medRxiv [Internet]. 29. december 2023 [henvist 7. april 2025];2023.12.19.23300192. Tilgængelig hos: https://www.medrxiv.org/content/10.1101/2023.12.19.23300192v2

46. National Institute of Diabetes and Digestive and Kidney Diseases [Internet]. [henvist 7. april 2025]. Prediabetes Screening: How & Why - NIDDK. Tilgængelig hos: https://www.niddk.nih.gov/health-information/professionals/clinical-tools-patient-management/diabetes/game-plan-preventing-type-2-diabetes/prediabetes-screening-how-why

47. Elsayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, m.fl. 4. Comprehensive Medical Evaluation and Assessment of Comorbidities: Standards of Care in Diabetes-2023. Diabetes Care [Internet]. 1. januar 2023 [henvist 7. april 2025];46(Suppl 1):S49–67. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/36507651/

48. Lindström J, Tuomilehto J. The Diabetes Risk Score - A practical tool to predict type 2 diabetes risk. Diabetes Care [Internet]. 1. marts 2003 [henvist 7. april 2025];26(3):725–31. Tilgængelig hos: https://dx.doi.org/10.2337/diacare.26.3.725

49. Center for Disease Control and Prevention [Internet]. 2022 [henvist 7. april 2025]. Prediabetes Risk Test. Tilgængelig hos: https://www.cdc.gov/prediabetes/risktest/index.html

50. Diabetesforeningen [Internet]. [henvist 7. april 2025]. Risikotest - test din risiko for type 2-diabetes. Tilgængelig hos: https://diabetes.dk/risikotest

51. Saleem TJ, Chishti MA. Exploring the Applications of Machine Learning in Healthcare. International Journal of Sensors, Wireless Communications and Control. 20. december 2019;10(4):458–72.

52. Griffin SJ, Little PS, Hales CN, Wareham NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. Diabetes Metab Res Rev [Internet]. 2000 [henvist 7. april 2025]; Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/10867715/

53. Mcmahan HB, Streeter M, Mannor S, Srebro N, Williamson RC. Open Problem: Better Bounds for Online Logistic Regression [Internet]. Bd. 23. JMLR Workshop and Conference Proceedings; 2012 [henvist 7. april 2025]. s. 44.1-44.3. Tilgængelig hos: https://proceedings.mlr.press/v23/mcmahan12.html

54. Lu K, Sheth P, Zhou ZL, Kazari K, Guergachi A, Keshavjee K, m.fl. Identifying Prediabetes in Canadian Populations Using Machine Learning. medRxiv [Internet]. 5. februar 2024 [henvist 7. april 2025];2024.02.03.24302301. Tilgængelig hos: https://www.medrxiv.org/content/10.1101/2024.02.03.24302301v1

55. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. Ann Intern Med [Internet]. 18. december 2018 [henvist 7. april 2025];169(12):866. Tilgængelig hos: https://pmc.ncbi.nlm.nih.gov/articles/PMC6594166/

56. Owess MM, Owda AY, Owda M, Massad S. Supervised Machine Learning-Based Models for Predicting Raised Blood Sugar. International Journal of Environmental Research and Public Health 2024, Vol 21, Page 840 [Internet]. 27. juni 2024 [henvist 7. april 2025];21(7):840. Tilgængelig hos: https://www.mdpi.com/1660-4601/21/7/840/htm

57. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak [Internet]. 6. november 2019 [henvist 24. april 2025];19(1). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/31694707/

58. Cardozo G, Pintarelli GB, Andreis GR, Lopes ACW, Marques JLB. Use of Machine Learning and Routine Laboratory Tests for Diabetes Mellitus Screening. Biomed Res Int [Internet]. 2022 [henvist 7. april 2025];2022. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/35392258/

59. Baig MM, Gholamhosseini H, Gutierrez J, Ullah E, Lindén M. Early Detection of Prediabetes and T2DM Using Wearable Sensors and Internet-of-Things-Based Monitoring Applications. Appl Clin Inform [Internet]. 1. januar 2021 [henvist 7. april 2025];12(1):1–9. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/33406540/

60. Zee B, Lee J, Lai M, Chee P, Rafferty J, Thomas R, m.fl. Digital solution for detection of undiagnosed diabetes using machine learning-based retinal image analysis. BMJ Open Diabetes Res Care [Internet]. 22. december 2022 [henvist 7. april 2025];10(6). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/36549873/

61. Dietz B, Machann J, Agrawal V, Heni M, Schwab P, Dienes J, m.fl. Detection of diabetes from whole-body MRI using deep learning. JCI Insight [Internet]. 8. november 2021 [henvist 7. april 2025];6(21). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/34591793/

62. M SF, Kumar JS, Jeya Prabha A, Selvaraj J, Kirubha S P A. Pioneering diabetes screening tool: machine learning driven optical vascular signal analysis. Biomed Phys Eng Express [Internet]. 1.

november 2024 [henvist 7. april 2025];10(6). Tilgængelig hos: https://www.researchgate.net/publication/385138430_Pioneering_diabetes_screening_tool_machine_learning_driven_optical_vascular_signal_analysis

63. Talakey AA, Hughes FJ, Bernabé E. Can periodontal measures assist in the identification of adults with undiagnosed hyperglycaemia? A systematic review. J Clin Periodontol [Internet]. 1. april 2022 [henvist 7. april 2025];49(4):302–12. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/35066921/

64. Li J, Yuan P, Hu X, Huang J, Cui L, Cui J, m.fl. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. J Biomed Inform. 1. marts 2021;115:103693.

65. Abbas M, Mall R, Errafii K, Lattab A, Ullah E, Bensmail H, m.fl. Simple risk score to screen for prediabetes: A cross-sectional study from the Qatar Biobank cohort. J Diabetes Investig [Internet]. 1. juni 2021 [henvist 24. april 2025];12(6):988–97. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/33075216/

66. Birk N, Matsuzaki M, Fung TT, Li Y, Batis C, Stampfer MJ, m.fl. Exploration of Machine Learning and Statistical Techniques in Development of a Low-Cost Screening Method Featuring the Global Diet Quality Score for Detecting Prediabetes in Rural India. Journal of Nutrition [Internet]. 1. oktober 2021 [henvist 24. april 2025];151(12 Suppl 2):110S-118S. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/34689190/

67. Choi SB, Kim WJ, Yoo TK, Park JS, Chung JW, Lee YH, m.fl. Screening for Prediabetes Using Machine Learning Models. Comput Math Methods Med [Internet]. 1. januar 2014 [henvist 24. april 2025];2014(1):618976. Tilgængelig hos: https://onlinelibrary.wiley.com/doi/10.1155/2014/618976

68. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nat Mach Intell [Internet]. 1. januar 2019 [henvist 30. april 2025];1(1):20–3. Tilgængelig hos: https://www.nature.com/articles/s42256-018-0004-1

69. Rasheed K, Qayyum A, Ghaly M, Al-Fuqaha A, Razi A, Qadir J. Explainable, trustworthy, and ethical machine learning for healthcare: A survey. Comput Biol Med [Internet]. 1. oktober 2022 [henvist 30. april 2025];149:106043. Tilgængelig hos: https://www.sciencedirect.com/science/article/pii/S0010482522007569

70. About NHANES [Internet]. [henvist 1. maj 2025]. Tilgængelig hos: https://www.cdc.gov/nchs/nhanes/about/

71. Examination Data - Continuous NHANES [Internet]. [henvist 1. maj 2025]. Tilgængelig hos: https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Examination

72. Questionnaire Data - Continuous NHANES [Internet]. [henvist 1. maj 2025]. Tilgængelig hos: https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire

73. Laboratory Data - Continuous NHANES [Internet]. [henvist 1. maj 2025]. Tilgængelig hos: https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Laboratory

74. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, m.fl. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ [Internet]. 2024 [henvist 23. maj 2025];385. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/38626948/

75. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat [Internet]. 2001 [henvist 30. maj 2025];29(5):1189–232. Tilgængelig hos:

https://www.researchgate.net/publication/2424824_Greedy_Function_Approximation_A_Gradient_Boosting_Machine

76. Zheng A, Casari A. Feature Engineering for Machine Learning - Principles and techniques for data scientists [Internet]. 2018 [henvist 2. maj 2025]. Tilgængelig hos: https://www.oreilly.com/library/view/feature-engineering-for/9781491953235/

77. Liang Z. Efficient Representations for High-Cardinality Categorical Variables in Machine Learning. 9. januar 2025 [henvist 31. maj 2025]; Tilgængelig hos: https://www.researchgate.net/publication/387952798_Efficient_Representations_for_High-Cardinality_Categorical_Variables_in_Machine_Learning

78. Sundhedsstyrelsen [Internet]. 2025 [henvist 1. maj 2025]. Anbefalinger om alkohol. Tilgængelig hos: https://www.sst.dk/da/Borger/En-sund-hverdag/Alkohol/Anbefalinger-om-alkohol

79. Centers for Disease Control and Prevention [Internet]. 2025 [henvist 1. maj 2025]. About Moderate Alcohol Use. Tilgængelig hos: https://www.cdc.gov/alcohol/about-alcohol-use/moderate-alcohol-use.html

80. Sundhedsstyrelsen [Internet]. 2024 [henvist 1. maj 2025]. Anbefalinger for søvnlængde. Tilgængelig hos: https://www.sst.dk/da/Borger/En-sund-hverdag/Soevn/Anbefalinger-for-soevnlaengde

81. National Heart, Lung and Blood Institute [Internet]. 2022 [henvist 1. maj 2025]. How Much Sleep Is Enough? Tilgængelig hos: https://www.nhlbi.nih.gov/health/sleep/how-much-sleep

82. Centers for Disease Control and Prevention [Internet]. 2024 [henvist 1. maj 2025]. About Sleep. Tilgængelig hos: https://www.cdc.gov/sleep/about/index.html

83. Li P, Stuart EA. Best (but oft-forgotten) practices: missing data methods in randomized controlled nutrition trials. Am J Clin Nutr [Internet]. 1. marts 2019 [henvist 31. maj 2025];109(3):504–8. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/30793174/

84. IterativeImputer — scikit-learn 1.6.1 documentation [Internet]. [henvist 31. maj 2025]. Tilgængelig hos: https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html

85. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? Int J Methods Psychiatr Res [Internet]. marts 2011 [henvist 2. maj 2025];20(1):40–9. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/21499542/

86. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. BMC Med Res Methodol [Internet]. 6. december 2017 [henvist 1. maj 2025];17(1). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/29207961/

87. Gupta P& SNK. Introduction to Machine Learning in the Cloud with Python. 1. Springer Cham; 2021.

88. Wiens J, Shenoy ES. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clinical Infectious Diseases [Internet]. 1. januar 2018 [henvist 18. maj 2025];66(1):149–53. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/29020316/

89. Paper D. Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python [Internet]. Hands-on Scikit-Learn for Machine Learning Applications: Data Science

Fundamentals with Python. Springer; 2019 [henvist 2. maj 2025]. 1–242 s. Tilgængelig hos: https://link.springer.com/book/10.1007/978-1-4842-5373-1

90. Irmanita R, Sri Suryani Prasetiyowati, Yuliant Sibaroni. Classification of Malaria Complication Using CART (Classification and Regression Tree) and Naïve Bayes. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi). 13. februar 2021;5(1):10–6.

91. Saxena R, Sharma SK, Gupta M, Sampada GC. A Novel Approach for Feature Selection and Classification of Diabetes Mellitus: Machine Learning Methods. Comput Intell Neurosci [Internet]. 2022 [henvist 1. maj 2025];2022. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/35463255/

92. What is Python? Executive Summary [Internet]. [henvist 1. maj 2025]. Tilgængelig hos: https://www.python.org/doc/essays/blurb/

93. Ekmekci B, McAnany CE, Mura C. An Introduction to Programming for Bioscientists: A Python-Based Primer. PLoS Comput Biol [Internet]. 1. juni 2016 [henvist 1. maj 2025];12(6). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/27271528/

94. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, m.fl. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods [Internet]. 1. marts 2020 [henvist 1. maj 2025];17(3):261–72. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/32015543/

95. Deepnote: Analytics and data science notebook for teams. [Internet]. [henvist 8. maj 2025]. Tilgængelig hos: https://deepnote.com/

96. Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell [Internet]. 1. december 1997 [henvist 8. maj 2025];97(1–2):273–324. Tilgængelig hos: https://www.sciencedirect.com/science/article/pii/S000437029700043X

97. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry [Internet]. 1. april 2015 [henvist 30. maj 2025];27(2):130–5. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/26120265/

98. Verbakel JY, Steyerberg EW, Uno H, De Cock B, Wynants L, Collins GS, m.fl. ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. J Clin Epidemiol [Internet]. 1. oktober 2020 [henvist 30. maj 2025];126:207–16. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/32712176/

99. Wynants L, Van Smeden M, McLernon DJ, Timmerman D, Steyerberg EW, Van Calster B. Three myths about risk thresholds for prediction models. BMC Med [Internet]. 25. oktober 2019 [henvist 30. maj 2025];17(1):192. Tilgængelig hos: https://pmc.ncbi.nlm.nih.gov/articles/PMC6814132/

100. Deberneh HM, Kim I. Prediction of type 2 diabetes based on machine learning algorithm. Int J Environ Res Public Health [Internet]. 2. marts 2021 [henvist 23. maj 2025];18(6). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/33806973/

101. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National health and nutrition examination survey: sample design, 2011-2014 [Internet]. 2014 [henvist 18. maj 2025]. Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/25569458/

102. Abroshan M, Burkhart M, Giles O, Greenbury S, Kourtzi Z, Roberts J, m.fl. Safe AI for health and beyond -- Monitoring to transform a health service. 2. marts 2023 [henvist 18. maj 2025]; Tilgængelig hos: https://arxiv.org/pdf/2303.01513

103. Almdal T, Kristensen JK. Lægehåndbogen. 2023 [henvist 27. maj 2025]. Type 2-diabetes. Tilgængelig hos: https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/endokrinologi/tilstande-og-sygdomme/diabetes-mellitus/type-2-diabetes/

104. Dansk Selskab for Almen Medicin [Internet]. 2019 [henvist 27. maj 2025]. Type 2-diabetes - opfølgning og behandling. Tilgængelig hos: https://www.dsam.dk/vejledninger/type2/diagnosen

105. 1.10. Decision Trees — scikit-learn 1.6.1 documentation [Internet]. [henvist 31. maj 2025]. Tilgængelig hos: https://scikit-learn.org/stable/modules/tree.html

106. Guyon I, Elisseeff A. An Introduction of Variable and Feature Selection. Journal of Machine Learning Research [Internet]. 2003 [henvist 5. maj 2025];1. Tilgængelig hos: https://www.researchgate.net/publication/221996079_An_Introduction_of_Variable_and_Feature_Selection

107. Videncenter for Diabetes [Internet]. [henvist 7. april 2025]. Prædiabetes. Tilgængelig hos: https://videncenterfordiabetes.dk/viden-om-diabetes/type-2-diabetes/hvad-er-type-2-diabetes/praediabetes

108. Khan RMM, Chua ZJY, Tan JC, Yang Y, Liao Z, Zhao Y. From Pre-Diabetes to Diabetes: Diagnosis, Treatments and Translational Research. Medicina (Kaunas) [Internet]. 1. september 2019 [henvist 31. maj 2025];55(9). Tilgængelig hos: https://pubmed.ncbi.nlm.nih.gov/31470636/

109. Jacobsen D, Thorsvik J. Forandring af organisationer. I: Hvordan organisationer fungerer - en indføring i organisation og ledelse. 4. udg. Hans Reitzels Forlag; 2022.

110. Jacobsen D. Forandringsstrategier. I: Organisationsændringer og forandringsledelse. 3. udg. Samfundslitteratur; 2020.

111. Morville A, Kolsum A, Rasmussen DH. Primærsektor - Det nære sundhedsvæsen. 1. Munksgaard; 2012.

112. Kidholm K, Dinesen B. Evaluering af telerehabilitering. I: Telerehabilitering. 1. udg. Munksgaard; 2017.