



Universiteit
Leiden
The Netherlands

Leveraging AI-based prediction in perioperative and critical care: from model development to clinical implementation

Meijden, S.L. van der

Citation

Meijden, S. L. van der. (2025, May 6). *Leveraging AI-based prediction in perioperative and critical care: from model development to clinical implementation*. Retrieved from <https://hdl.handle.net/1887/4245255>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4245255>

Note: To cite this publication please use the final published version (if applicable).

Chapter 11

Nederlandse Samenvatting

In dit proefschrift is de potentie van Artificial Intelligence (AI) voor het voorspellen van patiëntuitkomsten onderzocht. Hierbij is met name op toepassingen in de perioperatieve zorg en op de intensive care gefocust. De hoofdstukken van dit proefschrift volgen de fases van het AI ontwikkel- en evaluatietraject. PERISCOPE is de voornaamste AI-toepassing die in dit proefschrift wordt besproken. PERISCOPE voorspelt door middel van AI het risico op postoperatieve infecties binnen 7 en 30 dagen na de operatie.

Hoofdstuk 2 geeft een overzicht van de mogelijkheden en uitdagingen van AI en Machine Learning (ML) in de perioperatieve zorg. De basisconcepten van AI en ML worden besproken, evenals veelbelovende AI-toepassingen in dit klinische domein. De stappen van modelontwikkeling en validatie worden geschetst, inclusief relevante regelgeving. Tot slot worden de uitdagingen en valkuilen besproken waarvan artsen en verpleegkundigen zich bewust moeten zijn voordat ze AI-toepassingen in hun klinische praktijk implementeren.

Fase 0: Voorbereidingen voorafgaand aan modelontwikkeling

Hoofdstuk 3 presenteert een pre-implementatieonderzoek onder 64 intensive care (IC) artsen in twee ziekenhuizen voor de mogelijke implementatie van een AI-model dat het risico op IC heropname voorspelt. Deze enquête was uitgevoerd om informatie te verzamelen over de kennis en houding van de artsen in over de huidige besluitvorming. Respondenten toonden zich zeer bekend met AI en geloofden dat AI hen zou kunnen ondersteunen in hun dagelijkse werk. Er waren echter significante verschillen tussen de twee ziekenhuizen, waar in één ziekenhuis het AI-model was ontwikkeld en in de ander niet. Dit laat zien dat het belangrijk is om, voornamelijk in ziekenhuizen waarbij eindgebruikers nog onbekend zijn met de technologie, een effectief implementatieproces op te zetten met een focus op het informeren van eindgebruikers.

Hoofdstuk 4 geeft een overzicht van de verschillende methoden die worden gebruikt om patiënten met (postoperatieve) infecties te identificeren in predictie- en surveillance onderzoeken. Om voorspelmodellen voor postoperatieve infecties te kunnen ontwikkelen en valideren, moet op basis van gegevens uit het elektronisch patiëntendossier (EPD) betrouwbaar worden bepaald welke patiënten een infectie hadden en welke niet (i.e., het label). In dit onderzoek ontdekten we dat er 75 verschillende methoden en definities zijn om patiënten met infecties te identificeren en dat handmatig statusonderzoek nog steeds de belangrijkste methode is.

Fase 1: AI-model ontwikkeling

Hoofdstuk 5 biedt een haalbaarheidsstudie van PERISCOPE en een validatie van het label van postoperatieve infecties in vergelijking met handmatig statusonderzoek. De automatische labelmethode gebruikt voor PERISCOPE kwam voor 93% (54/60) van de gevallen overeen met de handmatige methode. De maat voor onderscheidend vermogen (AUROC) was 0.81 op de twee meest recente jaren aan patiënten data, waarbij 0.5 niet beter is dan kans en 1.0 een perfect voorspellend model. Op basis van deze resultaten werd PERISCOPE klaar geacht voor externe validatie in andere ziekenhuizen.

Fase 2: Beoordeling van AI-prestaties en betrouwbaarheid

Hoofdstuk 6 presenteert de externe validatie resultaten en de herkalibratie aanpak van PERISCOPE in twee Nederlandse en één Belgisch ziekenhuis. In totaal werden 253.010 procedures meegenomen in dit onderzoek, waarin we aantonen dat de discriminerende prestaties, kalibratie en klinische nut aanzienlijk verbeterden na het herkalibreren van het model per ziekenhuis. Er werd een grondige subgroep analyse uitgevoerd om afwijkingen in het model te kunnen opsporen voor verschillende patiëntengroepen. De in dit hoofdstuk beschreven methodologie kan worden toegepast voor andere op AI-gebaseerde voorspellingsmodellen om lokale validiteit te garanderen.

Fase 3: Klinisch testen van AI

Hoofdstuk 7 evalueert de voorspellingen van PERISCOPE ten opzichte van die van 51 chirurgen en chirurgen in opleiding voor het voorspellen van het risico op postoperatieve infecties. PERISCOPE presteerde beter dan menselijke deelnemers wanneer zij onzeker waren over hun voorspelling (AUROC 0,80 vs. 0,68) en klinici in opleiding (AUROC 0,77 vs. 0,73), al was de studiegroep klein. De kalibratie-eigenschappen waren beter voor het PERISCOPE model in vergelijking met klinici. Deze bevindingen geven aan dat de voorspellingen van PERISCOPE vooral waardevol zullen zijn in deze situaties en ook ondersteuning kunnen bieden in andere situaties.

Fase 4: AI implementeren en besturen

Hoofdstuk 8 geeft een overzicht van methoden voor de evaluatie van *Fairness* (een combinatie van eerlijkheid en rechtvaardigheid), voor AI-gebaseerde voorspellingsmodellen. Daarnaast wordt een kader geschetst voor de praktische toepassing van *Fairness* evaluatie voor AI-toepassingen. Het doel van onderzoek naar AI *Fairness* is ervoor te zorgen dat ongelijkheden in de gezondheidszorg niet groter worden door bevooroordeelde algoritmen. Aan de hand van twee verschillende AI-toepassingen laten we zien dat verschillende *Fairness* evaluatiemethoden van toepassing kunnen zijn, afhankelijk van het beoogde gebruik en het onderliggende ethische kader. Deze praktische richtlijnen zijn cruciaal voor de ontwikkelaars en beoordelaars van AI-gebaseerde voorspellingsmodellen om rechtvaardigere AI-implementaties te bewerkstelligen.

Concluderend kan worden gesteld dat de potentie van AI voor het voorspellen van perioperatieve en IC-uitkomsten geleidelijk wordt gerealiseerd. Hierbij is het belangrijk dat uitdagingen op het gebied van patiëntendata worden geadresseerd en dat robuuste validatie stappen worden geïmplementeerd. Deze stappen werden uitgevoerd en bewezen voor de AI-toepassing PERISCOPE. Er is echter meer klinisch bewijs nodig om de werkelijke impact van AI-gebaseerde voorspellingsmodellen op patiëntuitkomsten, kosten en ongelijkheden in de gezondheidszorg te kunnen beoordelen.