



Universiteit
Leiden
The Netherlands

Outcome after anterior cervical discectomy: from inferential statistics to Machine Learning

Goedmakers, C.M.W.

Citation

Goedmakers, C. M. W. (2023, December 20). *Outcome after anterior cervical discectomy: from inferential statistics to Machine Learning*. Retrieved from <https://hdl.handle.net/1887/3674247>

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/1887/3674247>

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

Summary in Dutch (Samenvatting)

Progressieve, degeneratieve veranderingen in de cervicale wervelkolom kunnen leiden tot nekpijn, verminderde functie, geleidelijk verlies van normale structuur en aandoeningen zoals scheefstand, radiculopathie en myelopathie. Anterieure cervicale discectomie en fusie (ACDF) is de standaard chirurgische behandeling voor degeneratieve cervicale pathologie, echter rapporteert ongeveer 25% van de patiënten na de operatie een lage tevredenheid, gerelateerd aan aanhoudende pijn en beperkingen. Het wordt verondersteld dat het verschil wordt veroorzaakt door meerdere factoren, waaronder; type chirurgische ingreep, degeneratieve patronen in de wervelkolom en geestelijke gezondheid van de patiënt. Het preoperatief identificeren van het onderliggende pathofysiologische proces dat de aanhoudende beperkingen na de operatie veroorzaakt, zou het mogelijk maken om vooraf de patiënten te identificeren die mogelijk geen profijt hebben van chirurgisch ingrijpen.

In dit proefschrift werden klassieke statistische methoden toegepast op klinische gegevens en Machine Learning op medische beeldvorming, met als doel de factoren te bepalen die bijdragen aan de variatie in de functionele uitkomsten na anterieure cervicale discectomie en om te onderzoeken hoe kunstmatige intelligentie kan worden gebruikt om het diagnostische en prognostische proces te verbeteren.

In de literatuur wordt aangenomen dat het plaatsen van een prothese na een anterieure discectomie (ACDA) leidt tot een reductie in het voorkomen van klachten passend bij Adjacent Segment Disease (ASD). Dergelijke studies vergeleken fusie (ACDF) en prothese (ACDA) bij de behandeling van cervicale radiculopathie in een veelal gemengde patiëntenpopulatie: radiculopathie-patiënten met en zonder myelopathie. Met als doel de werkelijke klinische effecten van de prothese te vergelijken met die van fusie, werden in dit proefschrift in de literatuur de klinische uitkomsten na ACDA vergeleken met ACDF bij patiënten die uitsluitend lijden aan radiculopathie. De pijn- en functionaliteitsscores waren bij patiënten na 2 jaar vergelijkbaar, onafhankelijk van het soort chirurgische ingreep. Bovendien waren de verschillen die werden gevonden in de gemengde, myelopathie en radiculopathie, patiëntenpopulatie niet klinisch relevant. Dit werd ook gevonden voor reoperatie aantallen en het ontstaan van ASD.

Het definiëren van een klinisch relevante uitkomst is cruciaal voor patiëntenzorg, besluitvorming en onderzoek. Bij het formuleren van criteria voor een succesvolle uitkomst drie en twaalf maanden na een operatie voor cervicale, degeneratieve radiculopathie, bleek dat een Neck Disability Index (NDI) percentage change score van 35.1, één jaar na de operatie, goed overeenkwam met de zelf gerapporteerde tevredenheid van de patiënten. Daarmee heeft deze parameter het vermogen om het beste onderscheid te maken tussen een klinisch succesvolle en niet succesvolle uitkomst (AUC 0.91). Voor alle gemeten PROMs bleek dat het onderscheidend vermogen van de percentage change scores beter was dan van de absolute change scores.

Na het combineren van data van twee gerandomiseerde, dubbelblinde klinische trials (RCT's) over anterieure decompressie bij cervicale radiculopathie werd aangetoond dat er twee jaar na operatie geen klinisch voordeel kon worden aangetoond voor de cervicale discusprothese, vergeleken

met de gouden standaard; ACDF. Daarnaast kon de voorlopige subgroep analyse twee jaar na de operatie ook geen specifieke patiëntengroep aanwijzen die meer zou profiteren van een prothese.

Echter, bij post-hoc analyse van onze RCT data werd gevonden dat patiënten die op baseline als depressief of angstig waren statistisch significant en klinisch relevant hogere NDI scores hadden één en twee jaar na de operatie. Linear Mixed Models (LMM) maakte het mogelijk om de verminderde geestelijke gezondheid met behulp van Hospital Anxiety and Depression Scale (HADS) op een continue schaal te analyseren, waaruit bleek dat als de NDI tijdens de follow-up verbetert, de score op de HADS ook verbetert. Daarnaast is er een R-Shiny applicatie ontwikkeld om de resultaten van de LMMs visueel weer te geven en daarmee de interpretatie van het model door patiënten tijdens een preoperatieve, poliklinische afspraak te vergemakkelijken.

Aangezien ASD vaak pas na langere follow-up wordt gerapporteerd, is het essentieel om de langetermijnresultaten te bestuderen van patiënten die een anterieure discectomie ondergaan voor de behandeling van cervicale radiculopathie. In de langetermijn follow-up van de RCT waarbij ACDA, ACDF en ACD werden vergeleken, ontbrak het bewijs voor klinische superioriteit van de cervicale discusprothese ook vijf jaar na de operatie. In het bijzonder werd er door het implanteren van een prothese niet voorkomen dat er klinisch relevante ASD optrad. Echter werd wel duidelijk dat ACD, op één niveau, zonder implantaat resulteerde in een slechter klinisch resultaat, waarbij werd verondersteld dat dit mogelijk veroorzaakt wordt door vertraagde fusie.

Er is enorm potentieel voor de gezondheidszorg om te profiteren van de twee grootste technologische innovaties van de eenentwintigste eeuw: big data en kunstmatige intelligentie (AI). Machine Learning, een fundamentele drijfveer van de huidige generatie AI, gebruikt big data om onderliggende patronen te identificeren om de prestaties van het model op vooraf gespecificeerd taken te verbeteren. Met een explosie aan data in de gezondheidszorg, variërend van basaal wetenschappelijke, klinische, numerieke en beeldgegevens tot enorme hoeveelheden administratieve en economische gegevens, zijn kunstmatige intelligente computerprogramma's goed in staat om deze gegevens te verzamelen, analyseren en te interpreteren om de gezondheidszorg te optimaliseren. De opkomst van big data en gerelateerde technologieën heeft echter geleid tot een toegenomen afhankelijkheid van computeralgoritmen en een afname van directe menselijke betrokkenheid en toezicht. Deze verschuiving brengt nieuwe ethische uitdagingen met zich mee. Voor de ethische toepassing van Machine Learning in de gezondheidszorg moeten klinici het evenwicht vinden tussen de principes van welzijn, niet schaden, rechtvaardigheid en respect voor autonomie om de patiëntveiligheid, betrouwbaarheid van diagnoses en eerlijke behandeling te waarborgen. Om de principes van welzijn en niet-schaden te waarborgen, moet AI vooral nauwkeurig en betrouwbaar zijn. De gegevens die worden gebruikt om AI-algoritmen te trainen, moeten zorgvuldig worden geselecteerd om bias te voorkomen die tot schade of onrecht kan leiden. Bovendien moet de zorgverlener de autonomie van patiënten bewaren door de nodige informatie over de onderliggende technologie te verstrekken alvorens te komen tot een informed consent.

Het automatiseren van delen van radiologische beeldanalyse met behulp van Machine Learning kan een meer accurate en consistente beoordeling bieden, waarbij de variabiliteit tussen waarnemers

wordt verminderd, terwijl de efficiëntie verbetert. Er is veel variatie in Machine Learning methoden in de literatuur, afhankelijk van het doel van de toepassing en de beschikbare beeld modaliteiten. In segmentatiemodellen laten Deep Learning methoden veelbelovende resultaten zien bij de toepassing van (volledig automatische) Convolutional Neural Network (CNN) modellen met behulp van röntgen-, CT- of MR-beeldvorming. Wat betreft de analyse van de cervicale wervelkolom, worden de biomechanische eigenschappen meestal bestudeerd met behulp van finite elementmodellen. De toepassing van artificial neural networks en support vector machine-modellen ziet er veelbelovend uit voor andere classificatiedoelinden.

In navolging van de bovenstaande literatuurstudie werd een CNN ontwikkeld om preoperatief patiënten te identificeren die postoperatief ASD gaan ontwikkelen. Dit model doet voorspellingen uitsluitend op basis van de preoperatieve cervicale MRI's, bij patiënten die anterieure cervicale decompressie chirurgie ondergaan op één niveau. Het voorspellen van ASD blijft een uitdaging voor artsen. In 344 ACDF patiënten presteerde het CNN beter dan twee klinische experts in het voorspellen van nieuwe symptomen postoperatief op aangrenzende niveaus, uitsluitend op basis van de preoperatieve MRI (95% versus 58% nauwkeurigheid, respectievelijk). Bij het identificeren van patiënten die symptomen zouden ontwikkelen, bleek de compressie van ruggenmerg en zenuwweefsel de meest belangrijke factor te zijn voor het algoritme.

Het preoperatief identificeren van patiënten die geen baat zullen hebben bij anterieure cervicale discectomie is klinisch lastig. In de laatste studie opgenomen in dit proefschrift is daarom een CNN ontwikkeld om het klinische succes te voorspellen een jaar na anterieure cervicale discectomie, op basis van uitsluitend preoperatieve laterale röntgenfoto's. Het doel was om te identificeren welke kenmerken op de röntgenfoto een belangrijke rol speelden in het besluitproces van het algoritme. Het model toonde een hoog discriminerend vermogen op het gebied van nauwkeurigheid, sensitiviteit en AUC. De Grad-CAM heatmaps toonden een substantiële invloed van de facetgewrichten op het classificatieproces, de enige synoviale gewrichten in de wervelkolom.