

Hype of Trend? Vergelijking van 'klassieke' multi-variate logistische regressie en 'contemporary' machine-learning gederiveerde voorspellingsmodellen in orthopedie

Auteurs

J.H.F. Oosterhoff

Namens Traumaplatform AI Consortium : A.E.J. Bulstra, G. Bain, M. vd Bekerom, M. Bhandari, M. Bongers, G.A. Buijze, S. Bzovsky, N. Chen, C.M. Court-Brown, D. van Deurzen, J.N. Doornberg, A.D. Duckworth, D. Eygendaal, O. Groot, Gosl

Werkplek spreker

Amsterdam UMC, Flinders Medical Center Adelaide, Massachusetts General Hospital Boston

j.h.oosterhoff@amsterdamumc.nl

Inleiding

Voorspelling van een uitkomst is fundamenteel bij besluitvorming ten aanzien van behandeling per unieke patiënt. Robuustheid van een voorspellingsmodel is een complex vraagstuk en afhankelijk van gekozen algoritme, dimensionaliteit van data en methodologie voor ontwikkeling van een model. Doel van deze studie is exploratie van statistische voordelen en uitdagingen bij het ontwikkelen van een voorspellingsmodel gebaseerd op machinaal leren (ML) en logistische regressie (LR) op eerder gepubliceerde datasets in orthopedie.

Methode

In totaal zijn 10 datasets geïncorporeerd van eerder gepubliceerde studies (o.a. complicatie na distale biceps ruptuur, (occulte) posterieur malleolus fractuur, (occulte) scaphoid fractuur, distale radius fractuur, infectie na operatieve fractuurbehandeling, infectie na orthopedische-oncologie chirurgie, mortaliteit bij fractuur door metastase).

Per dataset zijn parameters uiteengezet: voorspellers (retrograde variabelen selectie voor LR, variabelen selectie per ML), dimensionaliteit van data, sample- en event size. Interne validatie is uitgevoerd middels 5-maal 80:20 split en 10-fold kruisvalidatie.

Uitkomstmaten zijn discriminatie (AUC) en calibratie. Nauwkeurigheid van LR en ML werd vergeleken middels de Long's test.

Resultaten

Ontwikkelde ML-algoritmes zijn Boosted-Decision Tree, Support Vector Machine, Neural Networks en Bayes Point Machine. De Long's test liet geen significant verschil zien in nauwkeurigheid in LR- en ML-gederiveerde voorspellingsmodellen.

Conclusie

We presenteren de eerste studie in orthopedie die ML en LR-gederiveerde voorspellingsmodellen vergeleken heeft. Resultaten zijn veelbelovend en tonen geen significant verschil. Focus van de studie is exploratie van variabelen die bijdragen aan robuust ML-voorspellingsmodel. Klinische meerwaarde van ML is een feedback-loop in EPD waarin hij continu ge-update en nauwkeuriger wordt: richting gepersonaliseerde geneeskunde. Een methodologisch framework wordt gepresenteerd voor ontwikkeling van ML-gederiveerde voorspellingsmodellen in orthopedie.