Lifestyle and behaviors: predicting clinical signs and symptoms with machine learning

Lin MC, Graham AD, Kothapalli T, Wang J, Ding J, Tse V, Yu SX

Introduction: The goal of this study is to use machine learning to predict anterior segment clinical endpoints and symptoms from subject characteristics, lifestyle choices and behaviors.
Methods: Clinical signs, symptoms, and subject data were collected from both eyes of 458 subjects (n=916). Clinical endpoints included assessments of the eyelids and Meibomian glands, tear film stability and ocular surface integrity. Validated symptom questionnaires included the Berkeley DEFC, SPEED II, OSDI, VAS ratings of severity and frequency of discomfort and dryness, and CLDEQ-8 for CL wearers. Predictive features included subject demographics, contact lens wear and lifestyle parameters such as consumption of alcohol and time in near work, exercise and outdoors. The model was trained on 733 records and validated on 183 records to classify predicted outcomes into clinically relevant categories. Models were evaluated by 5-fold cross-validation accuracy.
Results: Age was a heavily weighted predictor for eyelid notching, Line of Marx displacement, blepharitis, shorter FTBUT among Asians, and VAS dryness severity and frequency. Sex and race were not important predictors for any outcomes. More hrs/day of near work was a predictor for erythema and better meibum quality. Poor meibum quality was predicted by more alcohol consumption as was a higher SPEED score, more frequent VAS discomfort, and debilitating dryness symptoms. More hrs/wk exercising predicted reduced lid wiper epitheliopathy (LWE), less frequent VAS discomfort and asymptomatic CL wear. More hrs/day outdoors predicted less corneal staining and lower CLDEQ-8 score. Hrs/day CL wear was not a useful predictor for any outcome, however more hrs/day comfortable CL wear was a predictor for lower OSDI score, better VAS severity and frequency of discomfort and dryness, lower CLDEQ-8 score and asymptomatic CL wear. Predictions for clinical endpoints were achieved with accuracies ranging from 80% to 99%. Symptom prediction accuracies ranged from 61% to 87%.
Conclusions: The machine learning prediction methodology in this study has the potential to complement and extend relationships among signs, symptoms and subject factors established through associative modeling approaches. Age and alcohol consumption are important predictors for several clinical signs and symptoms. More time exercising and outdoors predict beneficial outcomes.