Machine Learning Assisted Prediction of Progression in SSc-ILD

This abstract is titled Machine Learning Assisted Prediction of Progression in Systemic Sclerosis Patients: An approach to Concise, Tailored Model Construction Using Outpatient Clinical Data. The lead author is Nina van Leeuwen from Leiden University Medical Center in the Netherlands. There's great heterogeneity in systemic sclerosis with respect to disease severity. Guidelines for follow-ups are mainly based on expert consensus and advocate for annual assessments. The overall aim of this abstract was to develop a prediction model to guide annual assessment in individual systemic sclerosis patients in line with their particular manifestation and disease course. Machine learning approach was used to develop a model that can identify patients without disease progression. 492 systemic sclerosis patients were analyzed with a range of follow-up from two to 10 years. 52% of these patients had disease progression after a median of four years, with 29% having progression in the myocardial region, 23% had lung progression, and 16% had skin progression, and 12% died.

Now, the flip side of this is that 48% did not show progression, and that's where they utilized the probability plots, an area under the curve of receiver operating curves with the aim of optimizing the negative predictive value of the cutoffs, and they were able to identify a low risk group for disease progression that comprised 29% of patients. So, the conclusions from the authors were that... they reported that their data confirmed the severe nature of systemic sclerosis with cumulative progression in 52% of their cohort. Now, although precise risk stratification for individual patients is difficult, a machine learning approach enabled them to classify 29% of patients at low risk.

In this group, annual assessment programs might be less extensive. Now, indeed, this study is a sobering reminder of how severe this disease can be in many patients, and in order to identify patients at risk, the authors used the machine learning approach, using 90 independent patient variables to predict progression, but it really depends on which variables they used, the quality of them, the accuracy, the availability of the variables, and this will obviously depend on many factors, including the users who enter the variables, the source, the type of health system and the local processes involved.

So, obviously external validation is needed, which may need tailoring of their approach by country, region, and how data is captured and used by the system, but an objective tailored approach is needed and using machine learning is a promising way to do so. The authors analyzed data from patients with complete data available at three time points, so adherence and loss of follow-up may be important factors that are not necessarily captured in this study, and this is important, especially in health systems like that in the US where there are issues of healthcare coverage and inequities of access, and lack of follow-up or disrupted follow-up could be incorporated in future external models to define better who's at lower and higher risk.

Now, experts assess test characteristics to assess cutoffs for low, intermediate and high risk for disease progression, and these risk levels, therefore, were still subjective. So one question I had was, could there be a more agnostic approach to determining cutoffs based on the outcomes? So, extend that machine learning if you will, in terms of evaluating these cutoffs, but it makes sense to focus the model on those

who did not progress, although, could we also utilize this approach to inform interventions and follow-up for the intermediate and high risk groups as well?