A very good paper!

"Predictive accuracy on the tests is the criteria for how good the model is" .....

I've struggled to separate models for inference and prediction. As an engineer I've believed that models should describe the data generating process (inference) such that it gives us an understanding of how the blackbox functions; to the  point that the inference/knowledge/understanding is solid enough to hold water to predict future (out of sample). Therefore I am able to wrap my mind around predictive accuracy being the criteria for how good the model is.

The statistical community has fit probability models (distributions) on data and used the models to design and use methods to fit models (say linear regression) to characterize the data generation process. We've seen in PREDICT 410 so far that t-tests power depend on the normality of the response/residuals to be meaningful in determining if the regression coefficients (or the predictor) is contributing significantly in characterizing the variation in the response variable. If the data is not normally distributed, we'd have to know the distribution to use the Maximum Likelihood Estimates (MLE) to find the regression coefficients. The challenge is in knowing the correct distribution. Our tests for distributional conformity is such that we "fail to reject" the assumption or hypothesis that real life data is from a particular distribution. We'd never have enough evidence to say the data is from a particular distribution. (For example, data from Burr distribution is akin to normal distribution that the goodness of fit test would fail to reject normality (Wheeler, D. J. (2000). Normality And the Process Behavior Chart. Knoxville, Tenn: SPC PRESS).

Also another learning from the paper is "...goodness of fit test... lacks the power in more than few dimensions". The more the number of predictors being considered, the less likely are we detect lack of goodness of fit in the model. Therefore one can draw misleading conclusions from the model (the inference is questionable). If the model is at the best suspect in characterizing the data generation process, I am not too sure of how well it'd perform in a predictive realm.

The lesson for me is, one needs to be careful and understand how the model was generated and what are the limitations of the model before running with the model for application.

Leo Breiman has provided a view into other methods for modeling from the area of computer science and other non-statistical fields that focuses on the strength of prediction. An approach that focuses on finding an algorithm that learns from the data iteratively.

Whats interesting is predictive strength can be increased by using the plurality of predictions made by several algorithms. (Boosting random forrests). Unlike being caught in the muck of multiple statistical models, not knowing which model tells the correct story on the data generating process.