Machine learning to detect the Klinefelter syndrome endocrine profile

André Madsen1, Lise Aksglaede2 and Anders Juul2

1The Hormone Laboratory, Department of Medical and Biochemical Pharmacology, Haukeland University Hospital, Bergen, Norway
2Department of Growth and Reproduction, Copenhagen University Hospital – Rigshospitalet, Copenhagen, Denmark

RESULTS CONT’D

Machine learning classification method and performance

A) Reference
B) Prediction

Figure 2. Classification performance, agreement and clinical decision tree.
A) Representative confusion matrix for healthy controls (‘0’) or KS patients (‘1’) obtained from applying the trained ML algorithm to unseen test data (95%CI, 0.89 to 0.94). With repeated simulations, the accuracy was 97% (95%CI, 96% to 98%) and the Cohen’s Kappa of agreement was 0.92 (95%CI, 0.89 to 0.94). B) Clinical decision tree component of the random forest algorithm to classify healthy controls (‘0’) or KS patients (‘1’).

CONCLUSIONS

- Machine learning applied to biochemical data was able to make valid predictions to accurately classify cases of KS in this retrospective pilot study, but the model must be verified prior to clinical use.
- The random forest classification model outperformed all individual markers of KS and may yet be improved by addition of anthropometric and other biochemical markers.
- PCA provided a systematic exploration of biochemical differences between healthy boys and KS patients.

REFERENCES

4. Groth, K.A., Skakkebæk, N.E., Ishii, S., et al. Hormonal determination of prepubertal KS (47,XXY) age 6 - 13 years (total 35 visits). Serum hormone profile data was obtained with respect to testosterone, dehydroepiandrosterone sulfate (DHEAS) and 17-hydroxyprogesterone (17-OHP) and androstenedione [LC-MS/MS], sex hormone-binding globulin (SHBG) and anti-Müllerian hormone (AMH) [Beckman 

RESULTS

Benchmarking the KS endocrine profile

Figure 1. Klinefelter syndrome endocrine profile in relation to normal reference ranges for indicated hormones. Hormone level equivalent z-scores for KS patients (purple dots) overlaid on reference curves made with the ‘gamlss’ R package.