KDpredictor: identifying Kawasaki disease based on regular blood sample features

Tzu-Hsien Yang+, Ying-Hsien Huang+, Yuan-Han Lee, Ling-Sai Chang, Kuang-Den Chen, Wei-Sheng Wu*, Ho-Chang Kuo*

Supplementary Materials

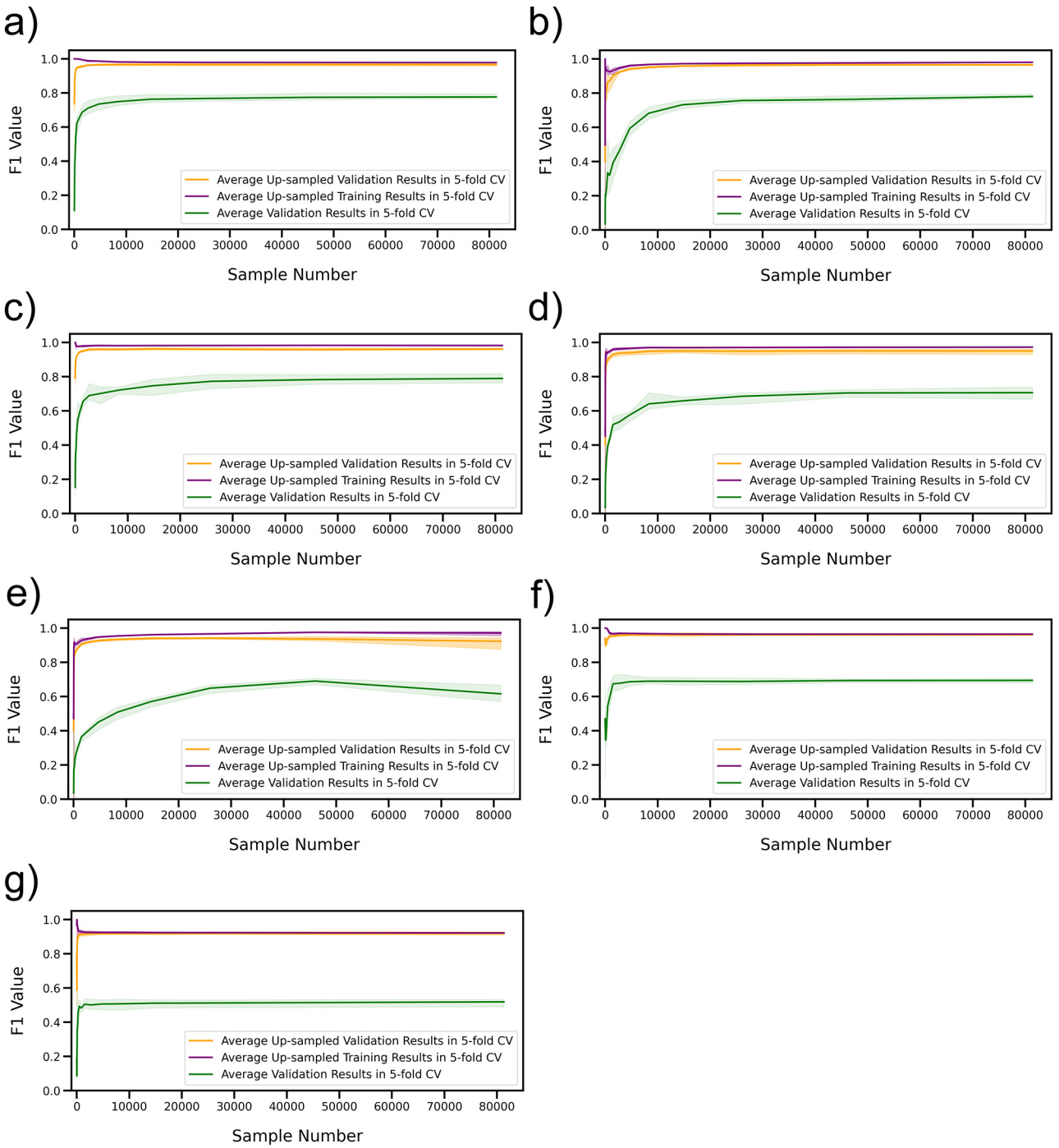

Figure S1: The learning curves of the designed pipeline using different learning algorithms. Notice that in logistic regression, the underlining learning function is not powerful enough to fully identify KD samples, contributing to under-fitting results. (a) AdaBoost. (b) XGBoost. (c) MLP. (d) Random Forest. (e) SVM. (f) Logistic Regression. (g) LDA.

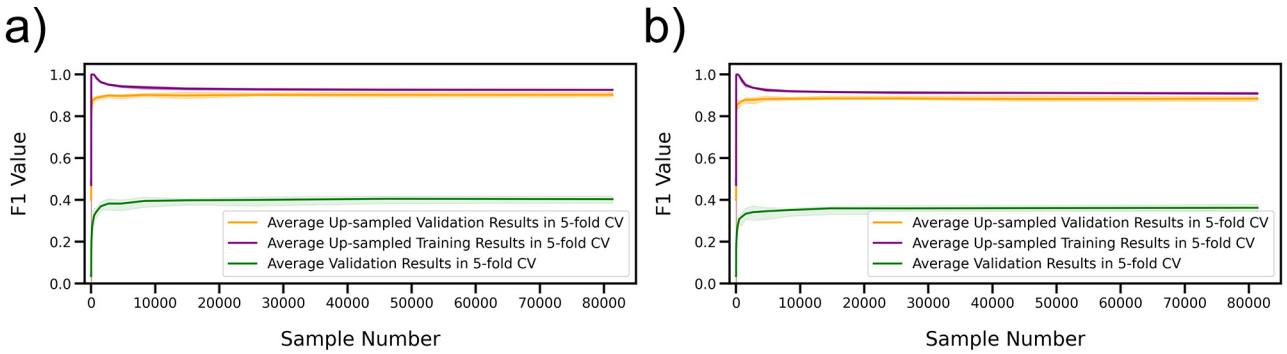

Figure S2: Learning curves for the LightGBM models based on different selected blood sample features. (a) Features utilized by Lam *et al.* [1]. (b) Features utilized by Ling *et al.* [2].

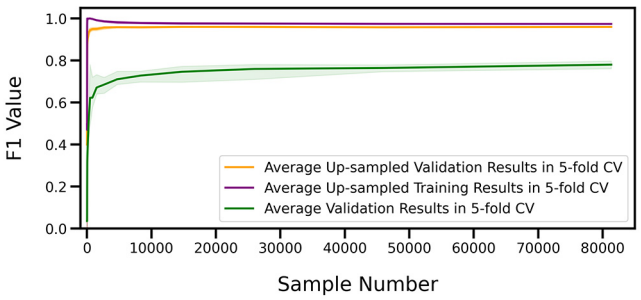

Figure S3: The learning curve for the reduced KDpredictor model based on only the top five identifying features.

References

[1] J. Y. Lam, C. Shimizu, A. H. Tremoulet, E. Bainto, S. C. Roberts, N. Sivilay, M. A. Gardiner, J. T. Kanegaye, A. H. Hogan, J. C. Salazar *et al.*, "A machine-learning algorithm for diagnosis of multisystem inflammatory syndrome in children and Kawasaki disease in the USA: a retrospective model development and validation study," *The Lancet Digital Health*, vol. 4, no. 10, pp. e717–e726, 2022.

[2] X. B. Ling, J. T. Kanegaye, J. Ji, S. Peng, Y. Sato, A. Tremoulet, J. C. Burns, and H. J. Cohen, "Point-of-care differentiation of Kawasaki disease from other febrile illnesses," *The Journal of Pediatrics*, vol. 162, no. 1, pp. 183–188, 2013.