**Supplemental Statistical Analysis Text**

Since the Mklaren models are not a sklearn object, we wrote an in-house python script to perform 5-fold nested cross-validation that first saved all hyperparameters, model validation area under the curve (AUC), and root mean square error (RMSE) from inner loops to a data frame, in cases where model validation AUCs were above 0.55. Then these model hyperparameters and associated metrics were filtered to give the top 10 validation AUCs and subsequently, the set giving the lowest RMSE was chosen.

AUC calculates the probability that a classifier will rank a positive instance higher than a negative one in a randomised fashion, with the assumption that “positive” instances would rank higher than “negative” instances. Thus, it is possible for a classifier with high AUC to perform worse in different regions of the receiver operating characteristic curve than that with a low AUC [1]. As a matter of fact, our method requires the use of the optimal AUC value as a threshold for binarising the output of the model as a way to convert the regression output into classes. In this case, choosing the hyperparameters giving a model with the highest AUC could give an overall model with lower performance.

Since we apply a regression model (Mklaren) for a classification task, we further use RMSE to assess how well the model fits the training data, prior to binarising the predicted values. RMSE shows the differences between predicted values and the observed (true) values in a regression model [2]. Hence, rather than only depending on a probability of classification (AUC value), we think the regression model fit (RMSE) should be accounted for in selecting the best hyperparameters. Therefore, in our study, the hyperparameters for a regression model with the highest discriminatory performance based on averaged thresholds and the best fit were applied to the test data (outer loop).

References

1. **Fawcett T**. An introduction to ROC analysis. *Pattern Recognit. Lett.* 2006; **27**(8):861–874.

2. **Chai T**, **Draxler RR**. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 2014; **7**(3):1247–1250.