Figure S1 The learning curves of the feature selection process using the REF method applied in random forest (A) and XGBoost algorithms incorporating SHapley Additive exPlanations values for feature importance evaluation (B). The learning curves plot the main metric values and their SDs against the number of features included in the model. Metrics shown include accuracy and Kappa, along with their corresponding SDs. The curves suggest an optimal number of features that maximize model accuracy and Kappa with minimal variability, guiding the feature selection process for model optimization. RFE, recursive feature elimination; XGBoost, eXtreme Gradient Boosting; SD, standard deviation.