

This isn't quite what you're suggesting, but tangentially related. In my attempt at an MMC maximizing model, I have a penalty term in my loss function for correlation with example predictions. I got a reasonably low correlation (0.4243) with example predictions, which could've perhaps be pushed lower if it wasn't for early stopping.