

# Expanding P-NET, a multi-purpose biologically informed deep learning framework

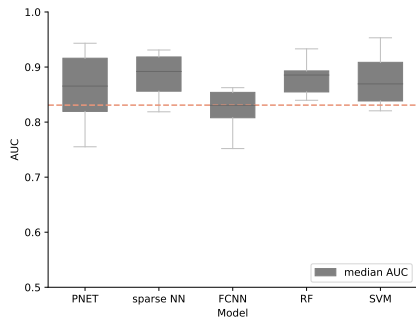
Marc Glettig

Dana Farber Cancer Institute (DFCI)

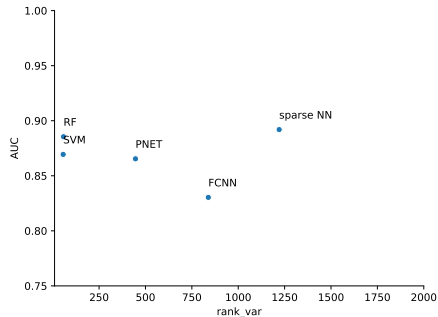
Medical Oncology, Population Sciences

May 13, 2023

# Validating performance

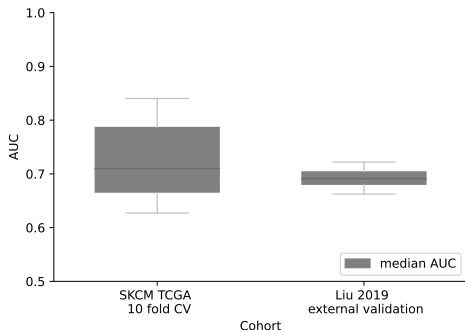


(a) AUC (10 fold CV) for predicting metastatic vs. non-metastatic prostate cancer in the original cohort. We see that sparse deep learning models (PNET & sparse NN) achieve similar performance to traditional machine learning (ML) models like random forests (RF) and support vector machine (SVM). The fully connected neural network is performing slightly worse.



(b) We compare the variance in gene importance rank for the 50 most important genes per model. The rank variance serves as a stability measure for model interpretability (lower scores are better). We see that traditional ML models with low complexity are stable while PNET is the most stable deep learning model.

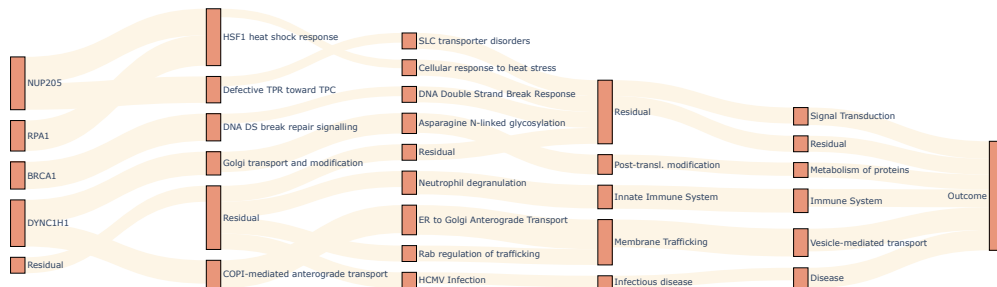
# Ploidy application



(a) AUC (10 fold CV) for predicting whole genome doubling (WGD) in SKCM cancer samples. Training/testing was performed in the TCGA cohort with 373/43 SKCM samples. PNET predicts WGD events well with a mean test AUC of 0.72. The model generalizes well to an external melanoma cohort (Liu 2019) and predicts WGD with a mean AUC of 0.69

# Ploidy application

Importance scores Sankey Diagram



(a) Sankey diagram to show feature importances flow through Pnet in WGD prediction. Due to the sparse connections in Pnet we can attribute the nodes in our neural network the respective Reactome pathway name. Shown are the most important genes & pathways per model layer. Genes such as RPA1 and BRCA known to be involved in DNA repair are important drivers of the model decision. Additionally the pathway for DNA double strand (DS) break repair is involved in the model decision.