OncoCast: an improved interface for survival analysis using genomic data

Axel Martin

Memorial Sloan-Kettering Cancer Center

Motivation

Broad-panel clinical sequencing

It is becoming increasingly routine for patients to have their tumor sequenced as part of their care.

Data biases

Left-truncation or delayed entry is a common bias in such datasets as patients are regularly sequenced at a date post-diagnosis.

Objective

We aim to use statistical machine learning approach for assessing the utility of clinical sequencing data for survival risk stratification, accounting for biases and confounding factors in cohort studies.

Methods

R package

The software we present in order to fulfill our objective is an R package named OncoCast.

Underlying algorithm

OncoCast takes as input (possibly left-truncated) data and a statistical learning method including penalized Cox regression and gradient boosting while performing cross-validation to predict survival outcome. The prediction is base on the ensemble of all models to boost performance.

Prediction

Ensemble learning

In order to maximize prediction accuracy and minimize bias, at each iteration the algorithm will make use of the trained model to assign a predicted risk score to all patients in the test set. After a large number of splits we average the predicted score of each patient to generate an unbiased average predicted risk score.

Going beyond the risk score

The average predicted risk is further rescaled between 0 and 10 to ease interpretation and intuition. It can be used as a surrogate for the aggregated effect of features used to generate the survival models. It can thus be used as a continuous predictor, or dichotomized into risk groups.

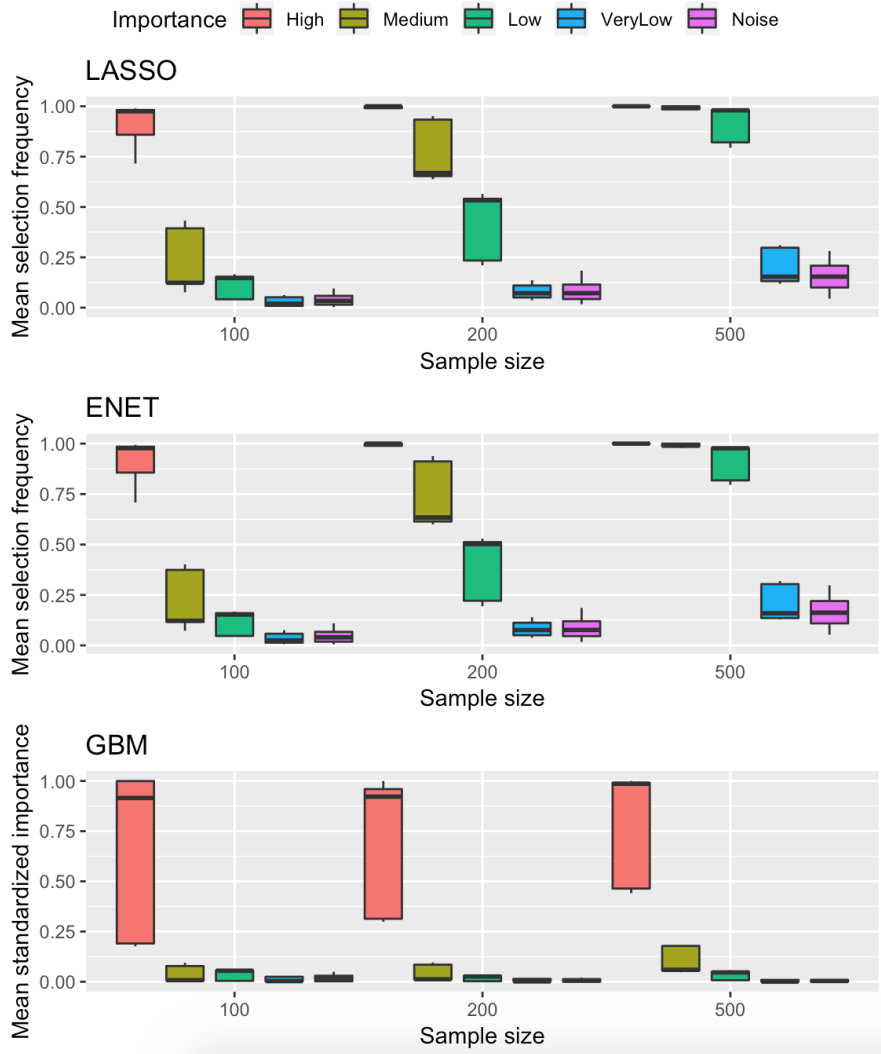
Simulation results

Prognostic

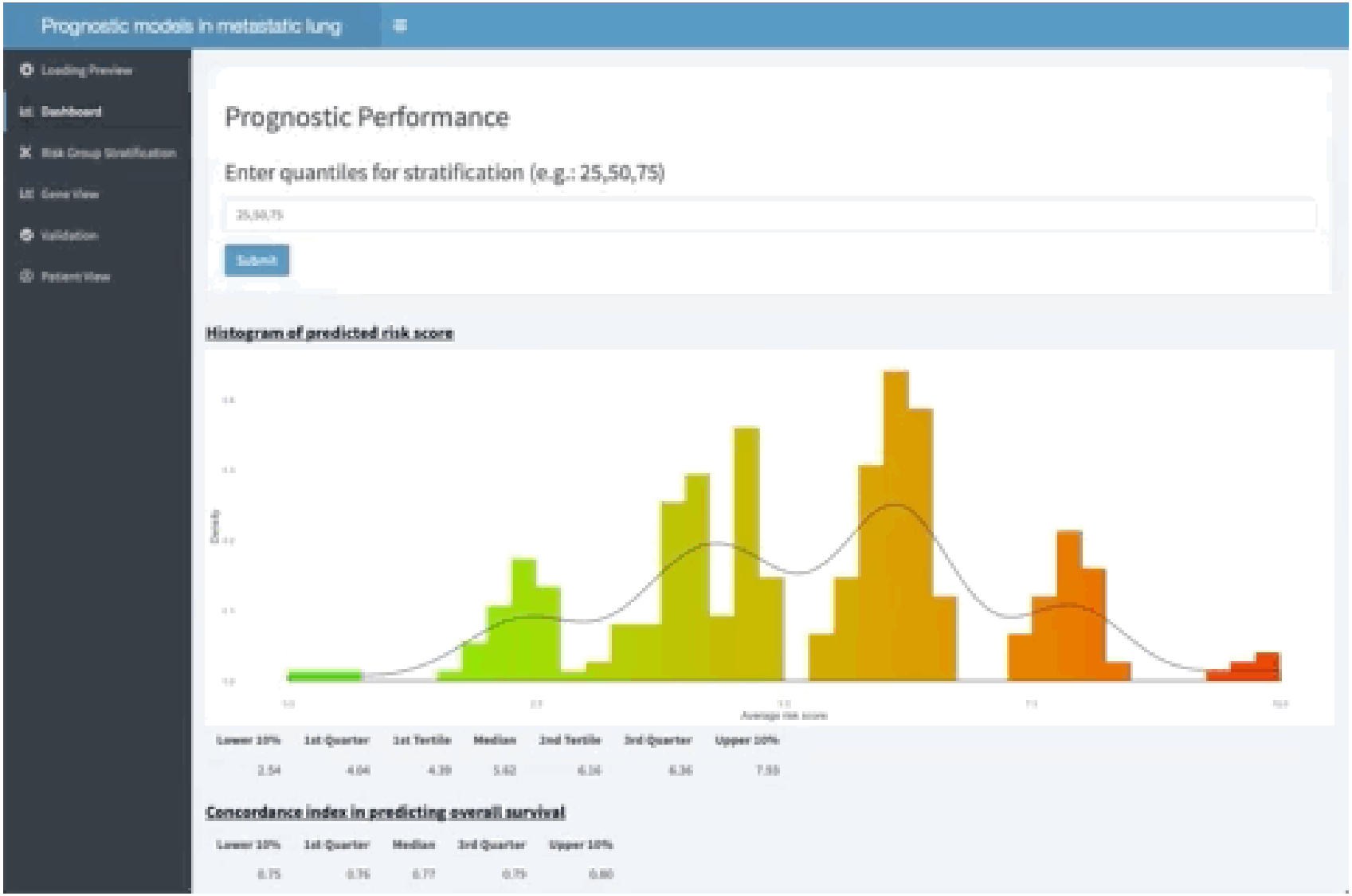
Correlation

	enet	lasso	gbm
0			
100	0.81 (0.04)	0.82 (0.04)	0.725 (0.05)
200	0.87 (0.02)	0.87 (0.02)	0.82 (0.03)
500	0.89 (0.01)	0.89 (0.01)	0.87 (0.01)
0.3			
100	0.84 (0.03)	0.83 (0.04)	0.83 (0.02)
200	0.88 (0.01)	0.89 (0.01)	0.87 (0.01)
500	0.9 (0.01)	0.9 (0.01)	0.89 (0.01)
0.6			
100	0.89 (0.01)	0.88 (0.02)	0.88 (0.01)
200	0.91 (0.01)	0.91 (0.01)	0.9 (0.01)
500	0.92 (0)	0.92 (0)	0.91 (0.01)

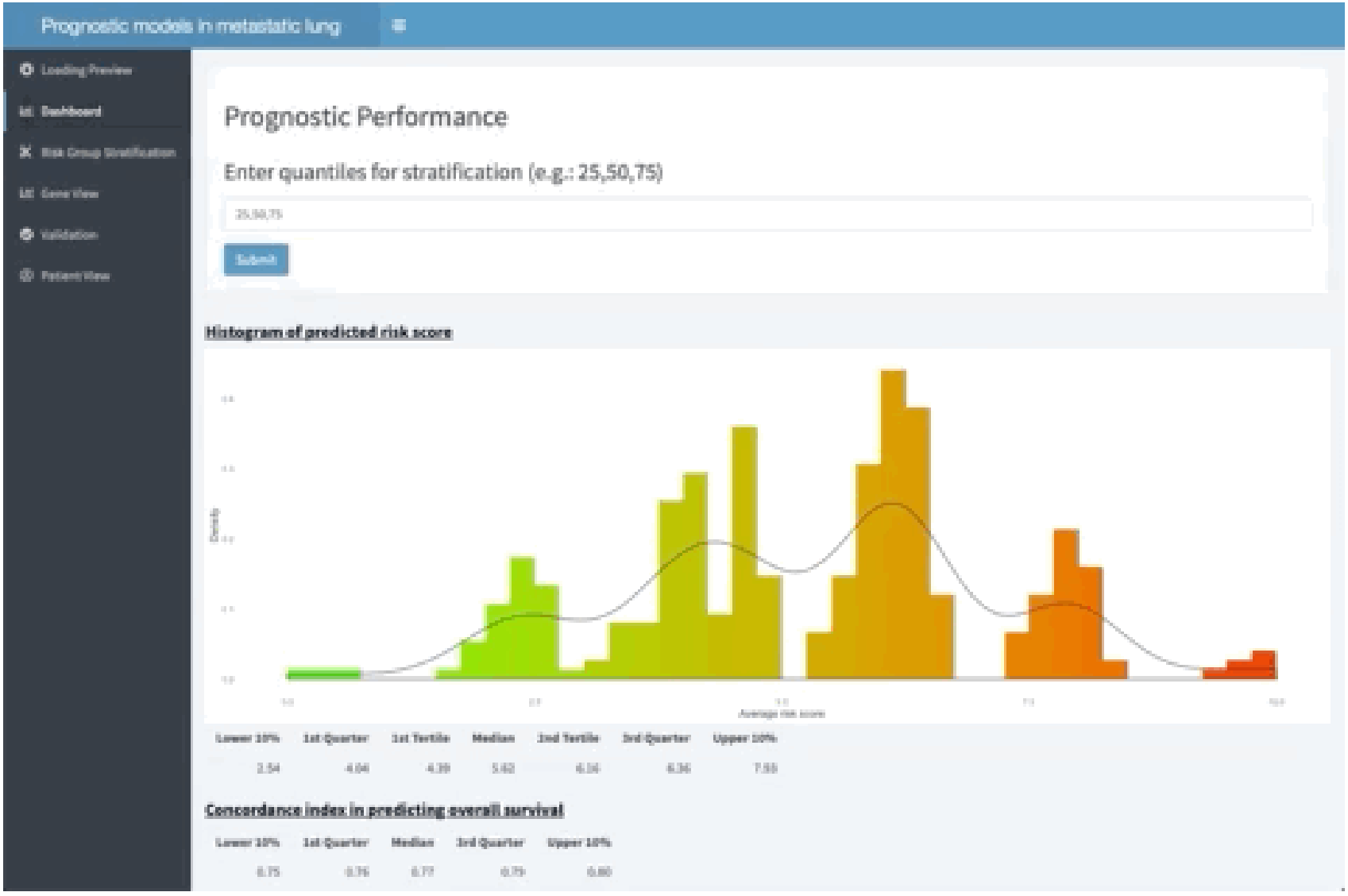
Feature selection



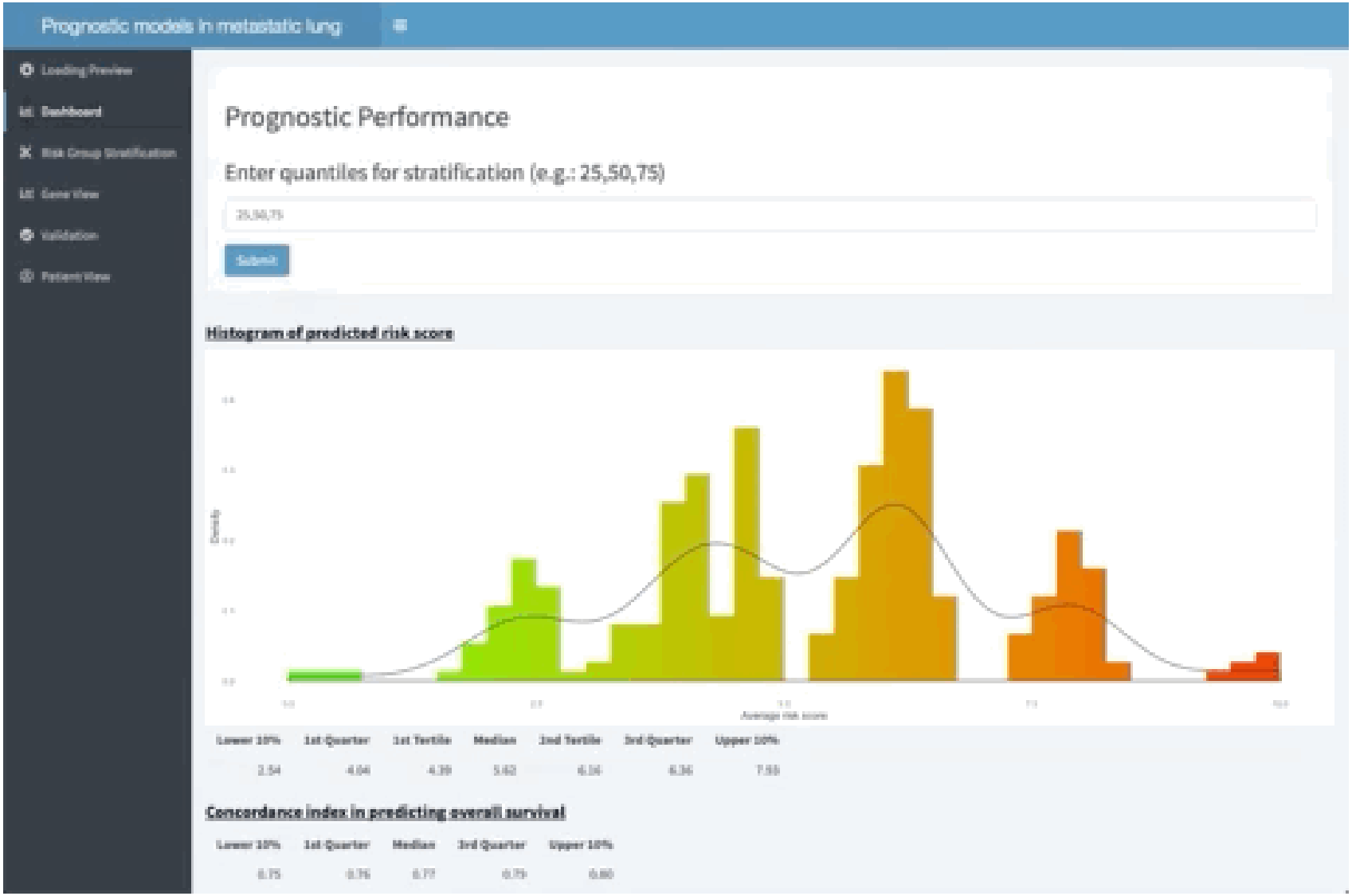
An efficient exploring tool



An efficient exploring tool



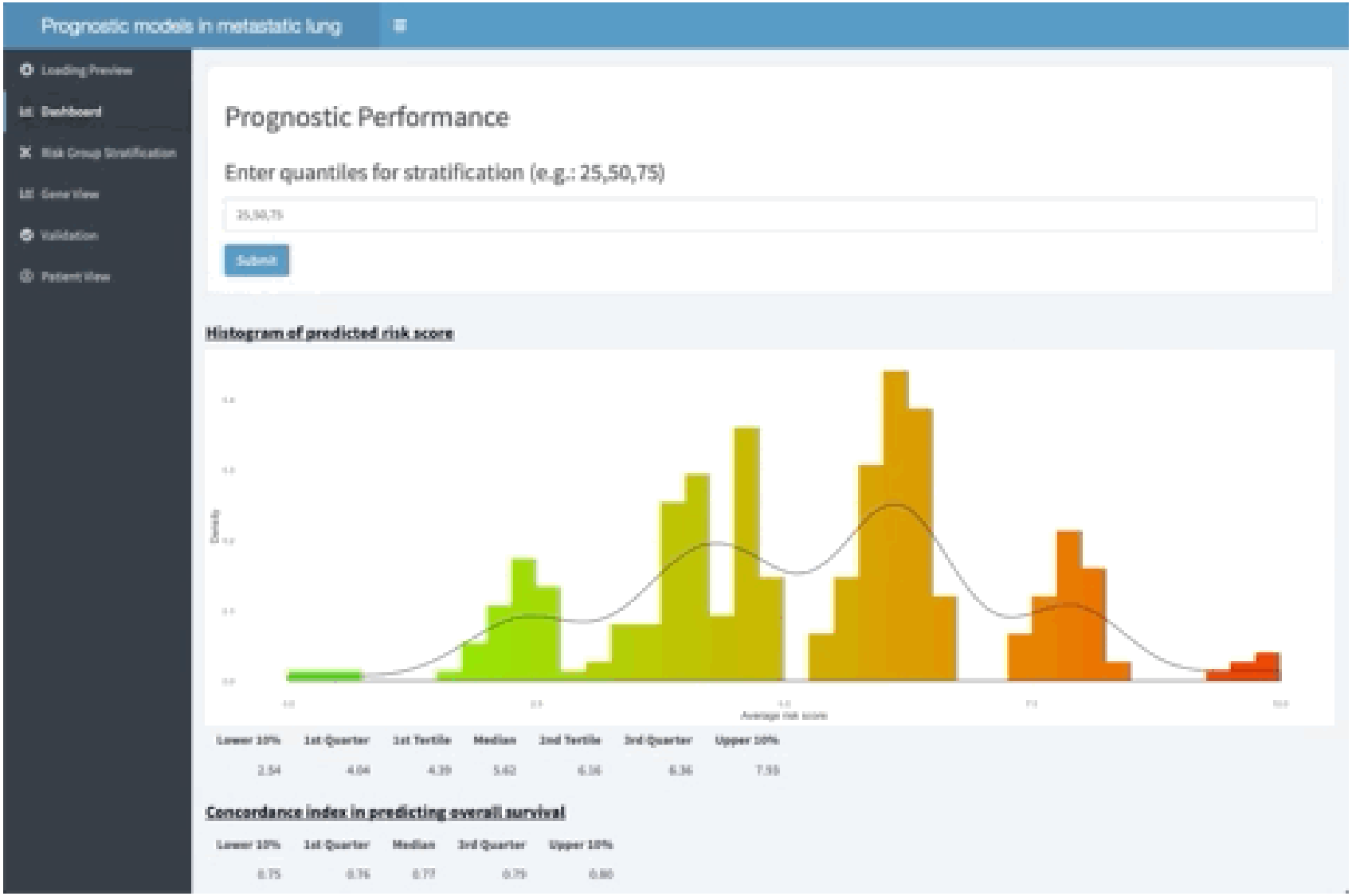
An efficient exploring tool



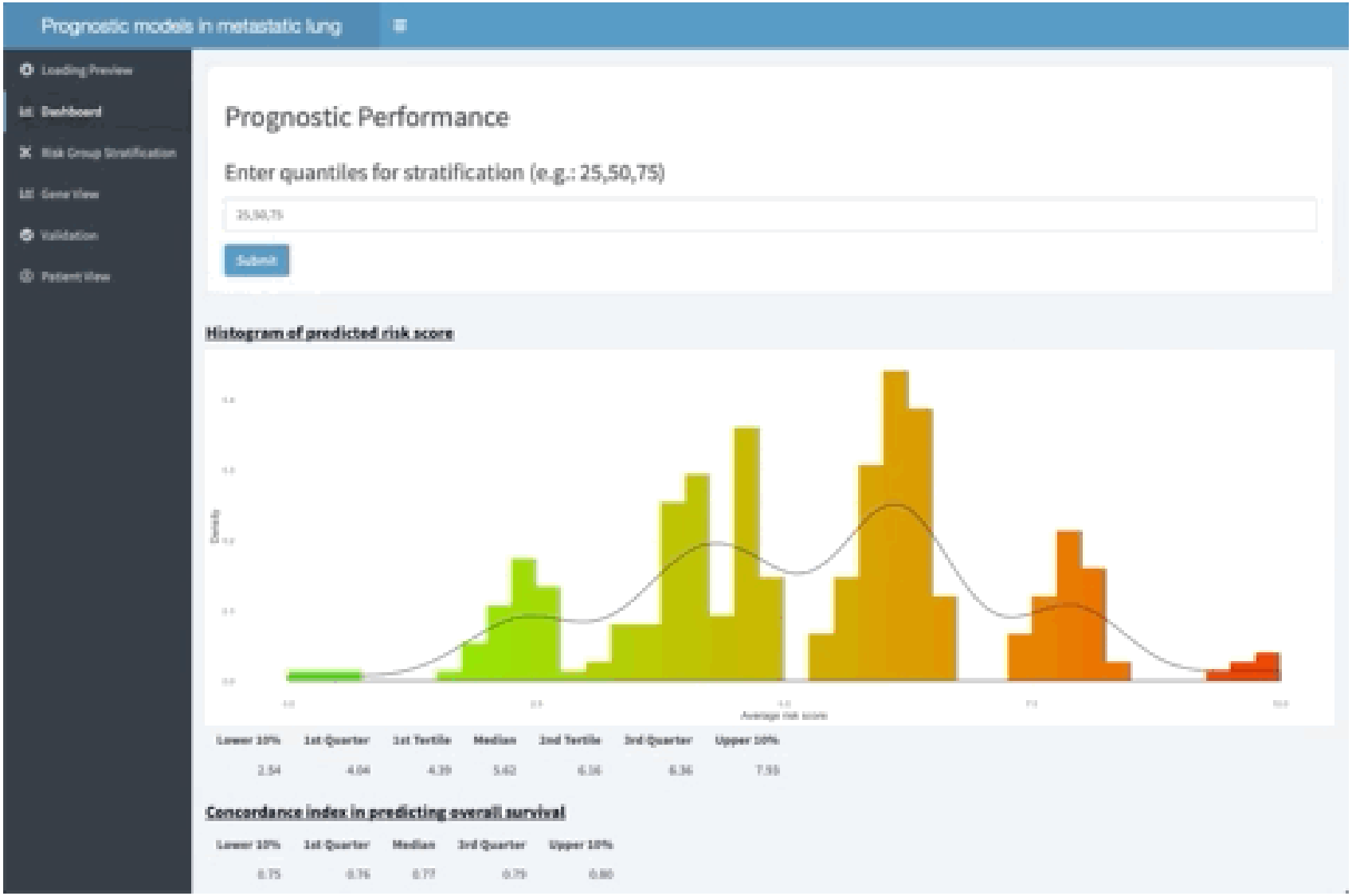
An efficient exploring tool

An efficient exploring tool

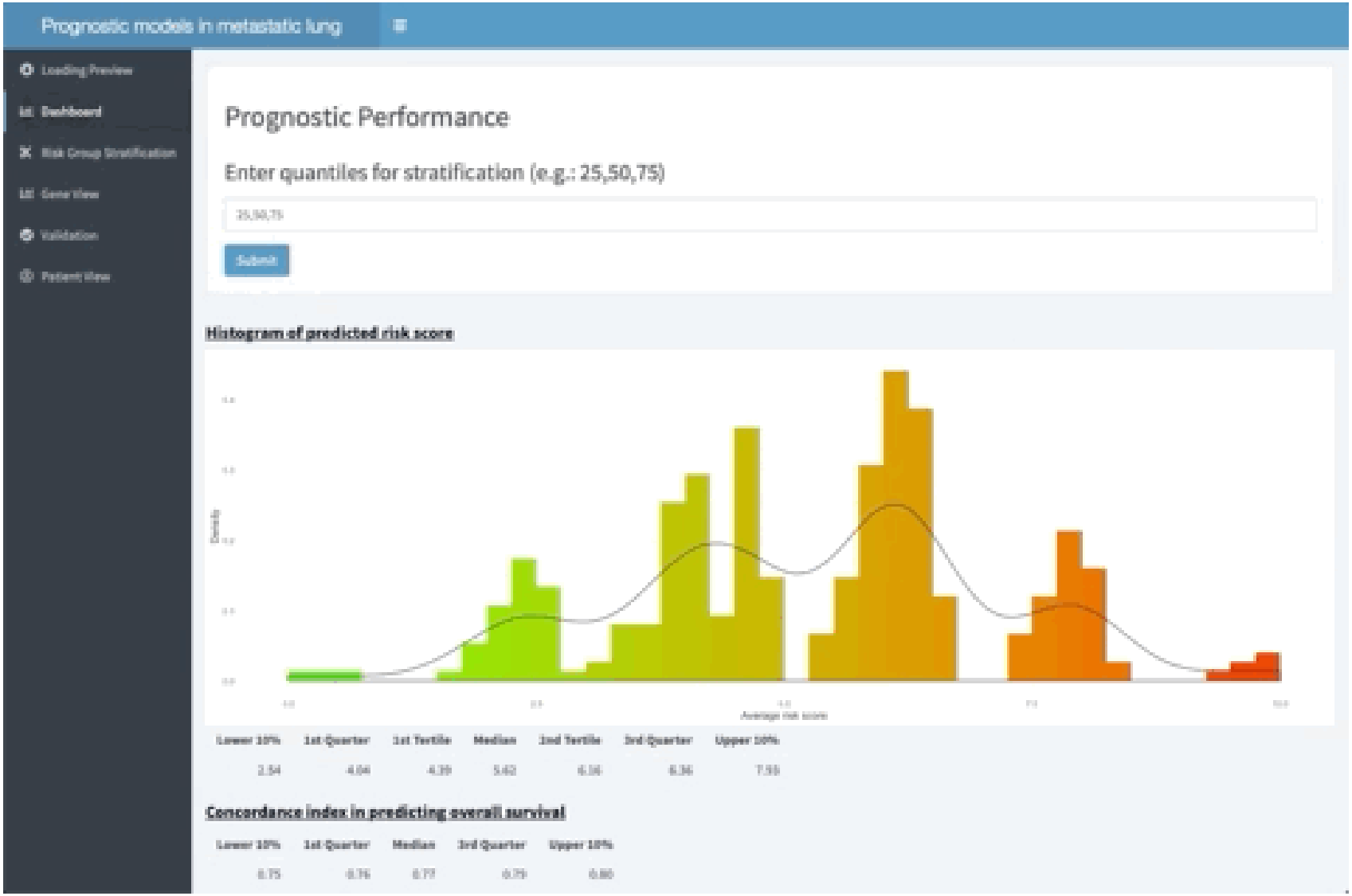
An efficient exploring tool



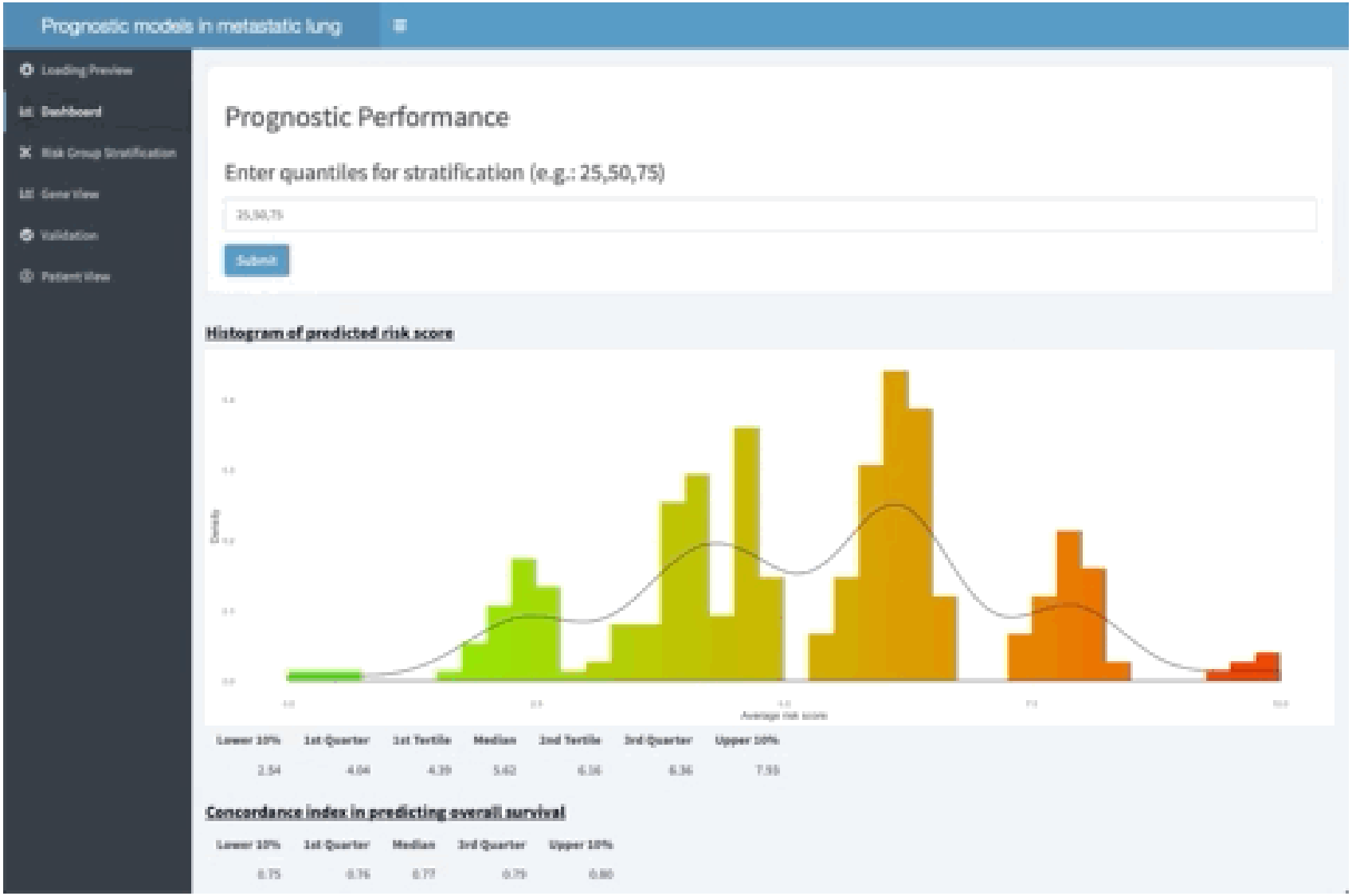
An efficient exploring tool



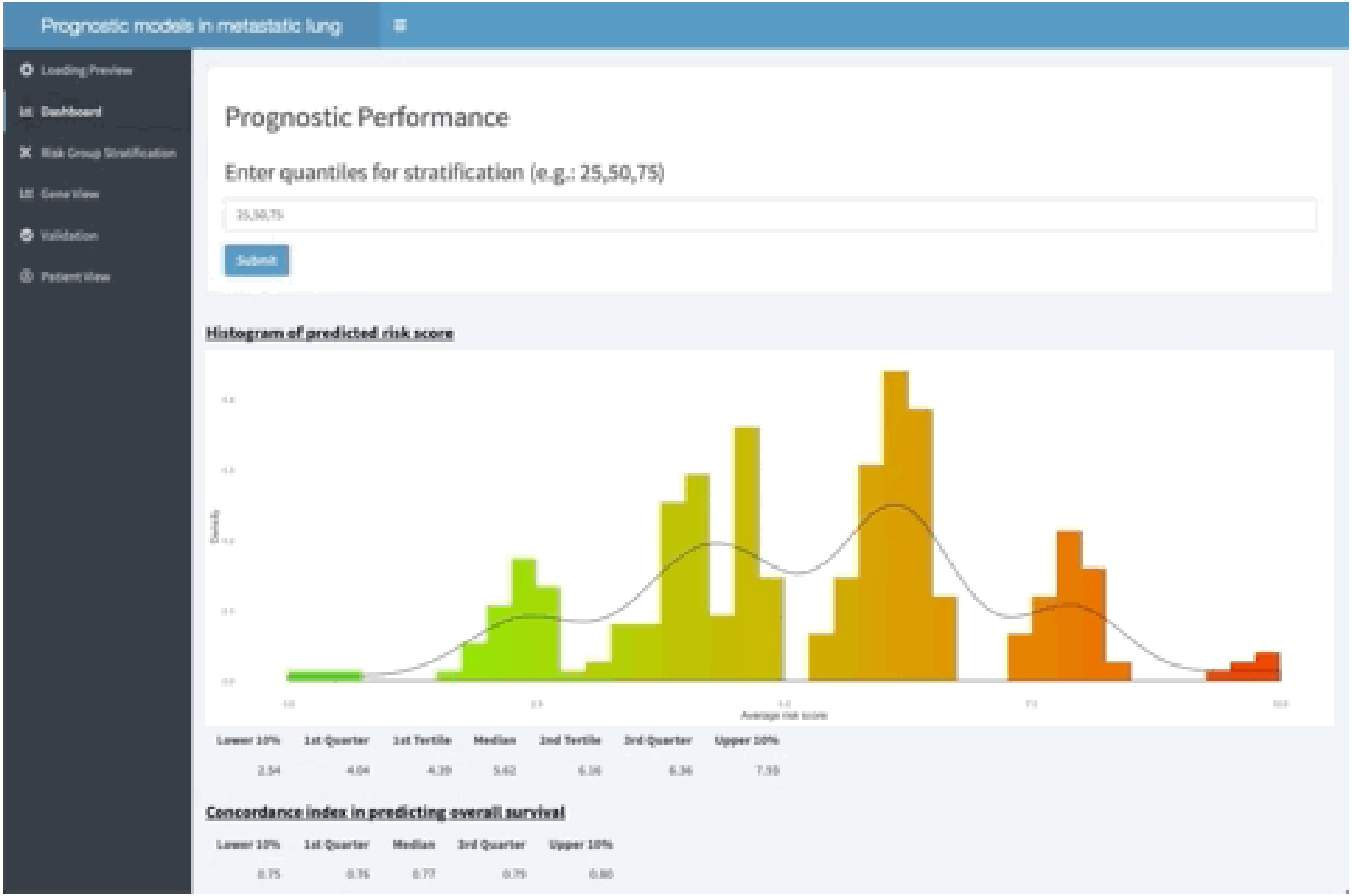
An efficient exploring tool



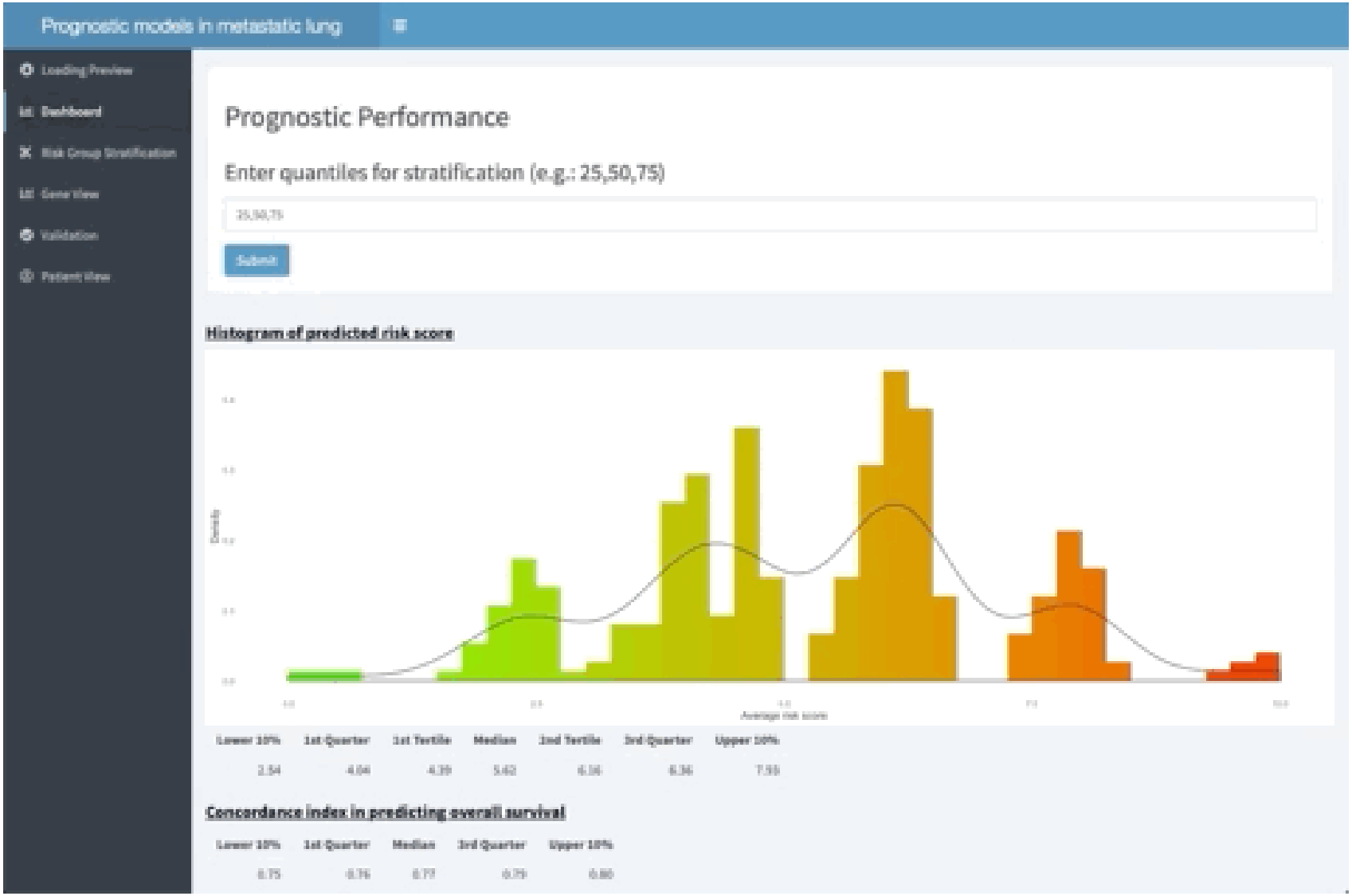
An efficient exploring tool



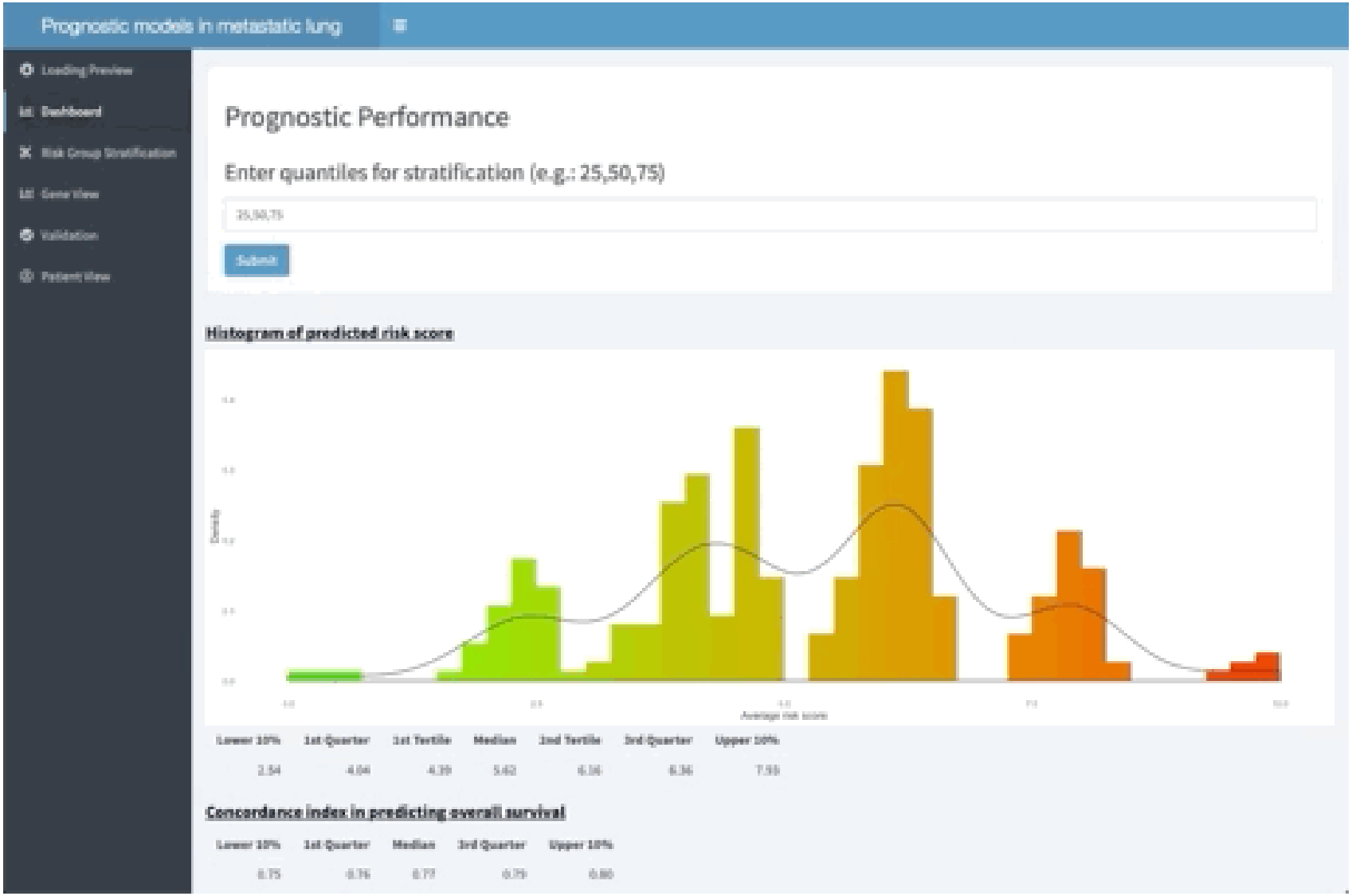
An efficient exploring tool



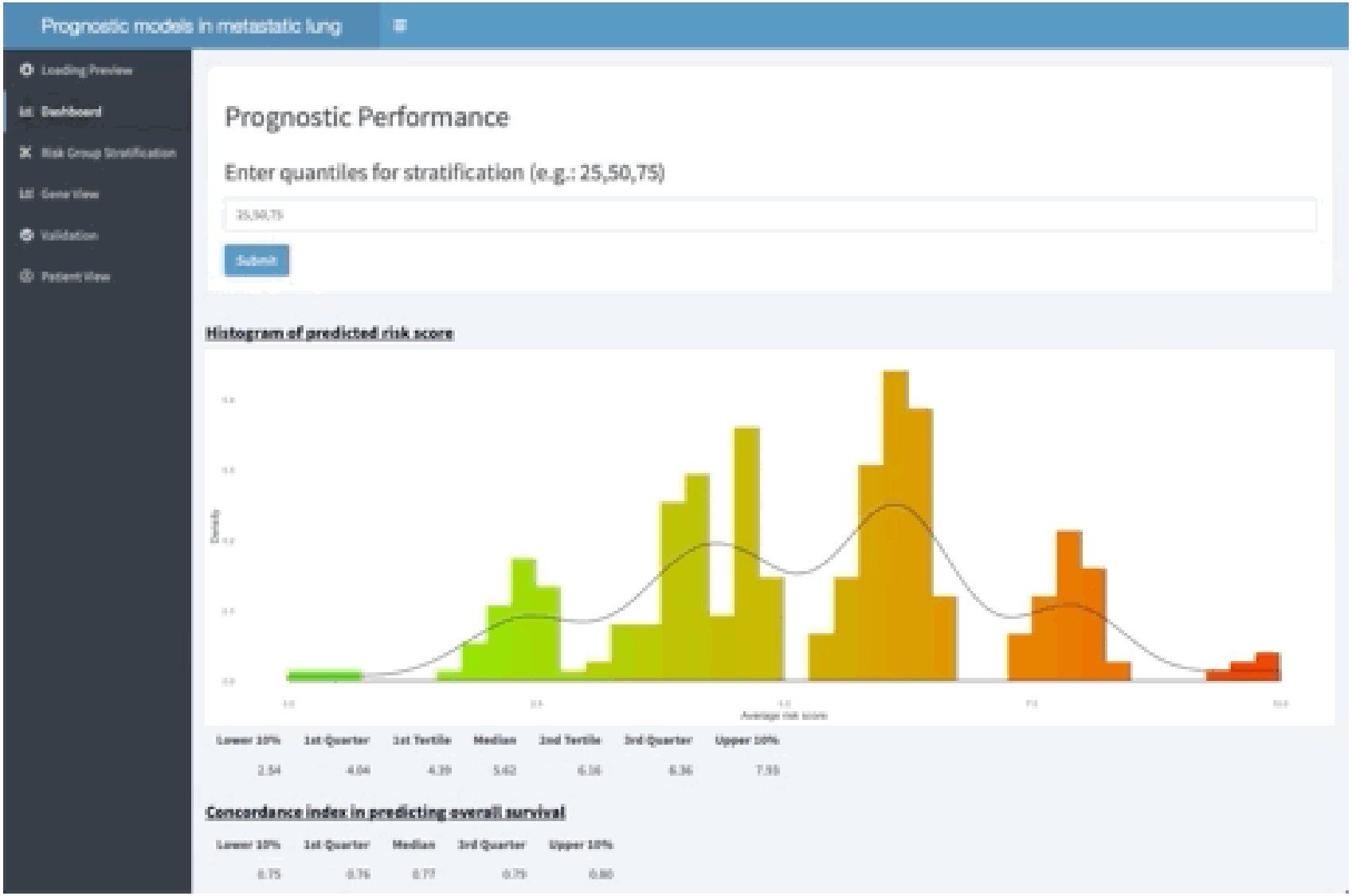
An efficient exploring tool



An efficient exploring tool



An efficient exploring tool



An efficient exploring tool

An efficient exploring tool

An efficient exploring tool