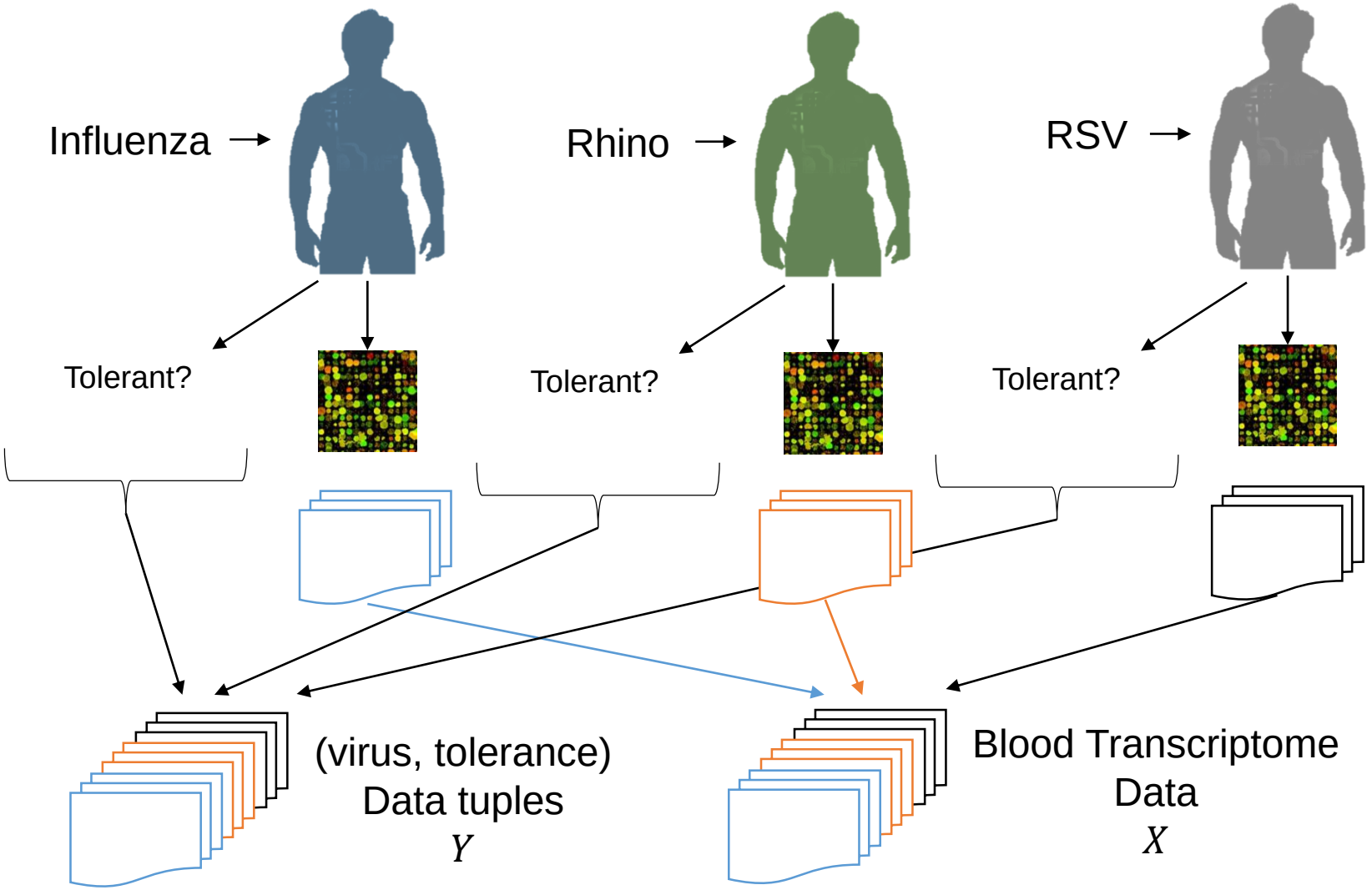
A Hierarchical Meta-Classifier

For Multi-output Tolerance Prediction

SAHIL LOOMBA
Wyss Institute for Biologically Inspired Engineering

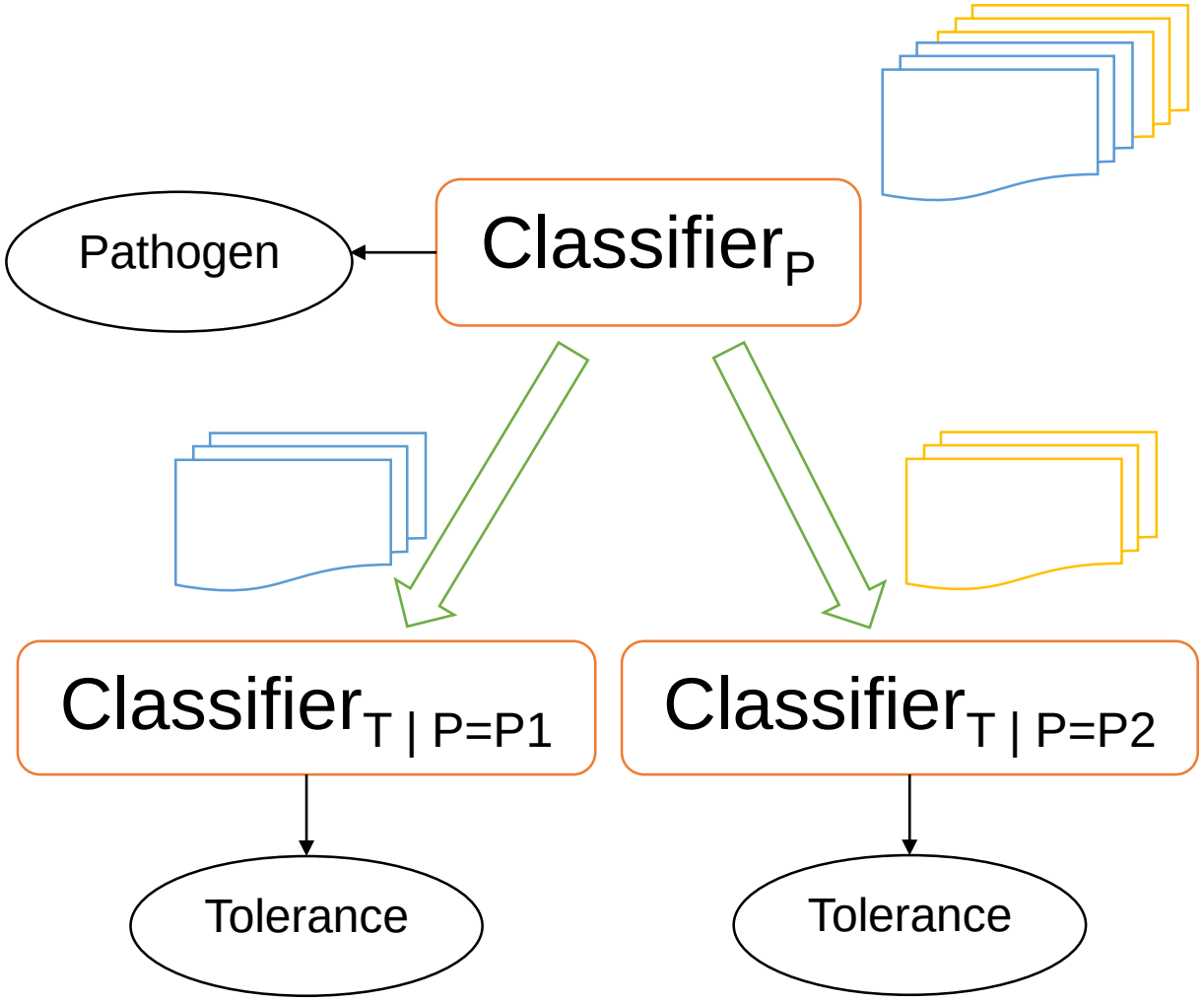
Problem Overview



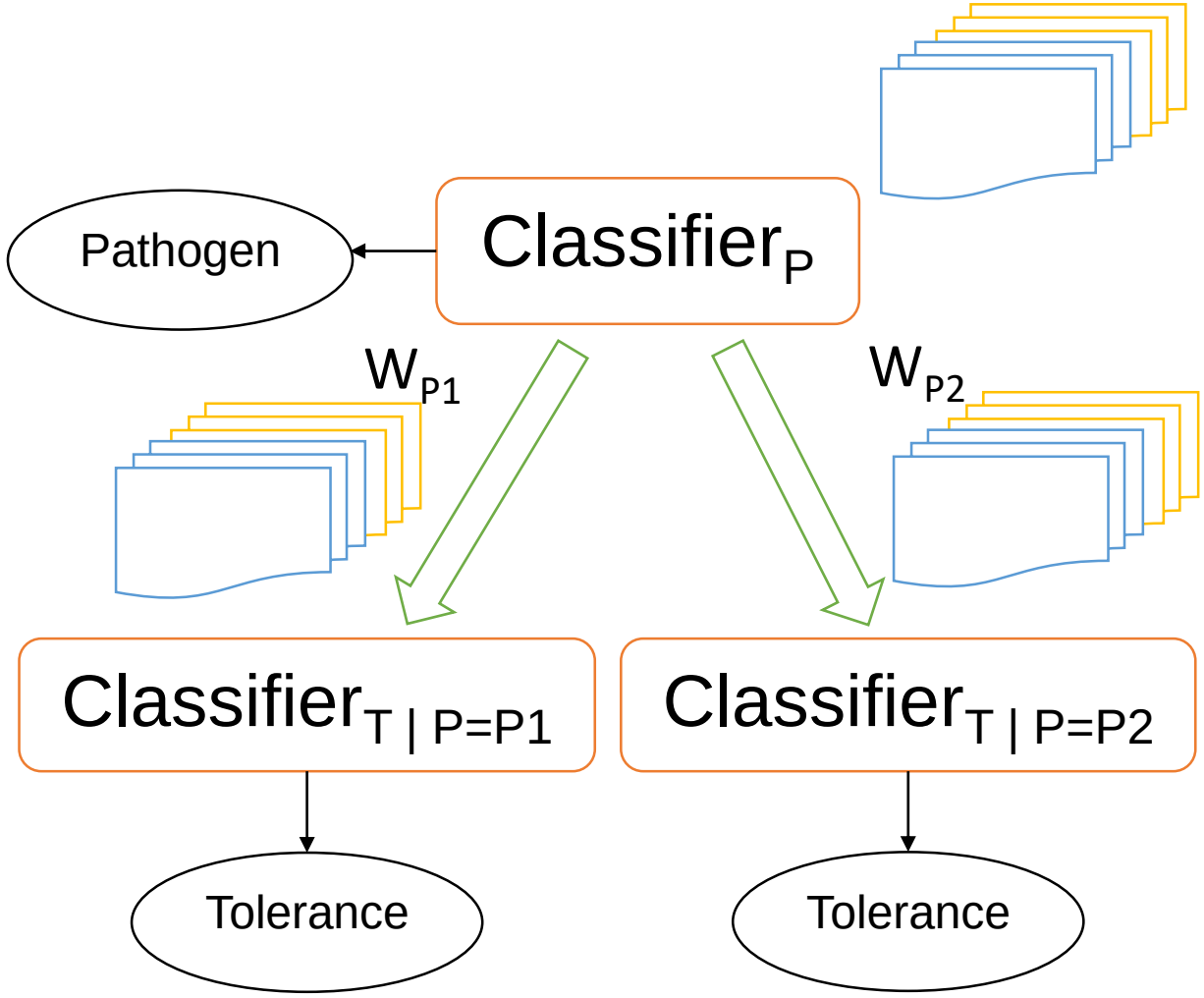
Models Overview

- Single-output version: tolerance_only, pathogen_only classifiers
- Multi-output version:
  - both_together classifier
  - both_separate meta-classifier
  - hierarchical meta-classifier
    - pathogen_given_tolerance meta-classifier
    - tolerance_given_pathogen meta-classifier

Hierarchical Meta-classifier Overview



“Hard” Data subsampling



“Soft” Data weighting

Scoring Classifiers

- Accuracy?
  - Not the best way
- Negative log loss score:

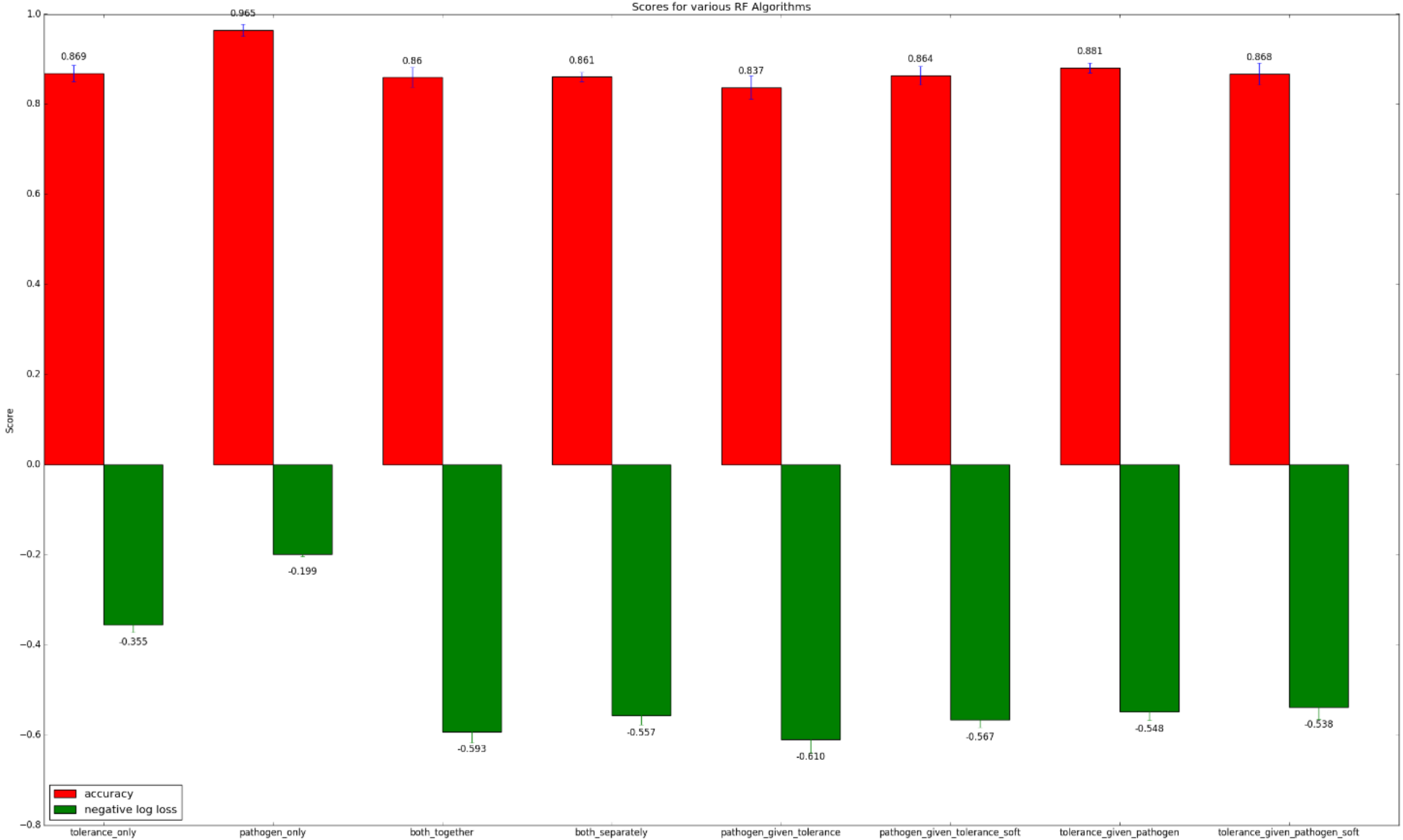
$$nll = -\sum_{i}^{m} 1\{Y_i = y\}\log(P(Y_i = y)) \qquad Y_i \in \{tolerance, pathogen\}$$

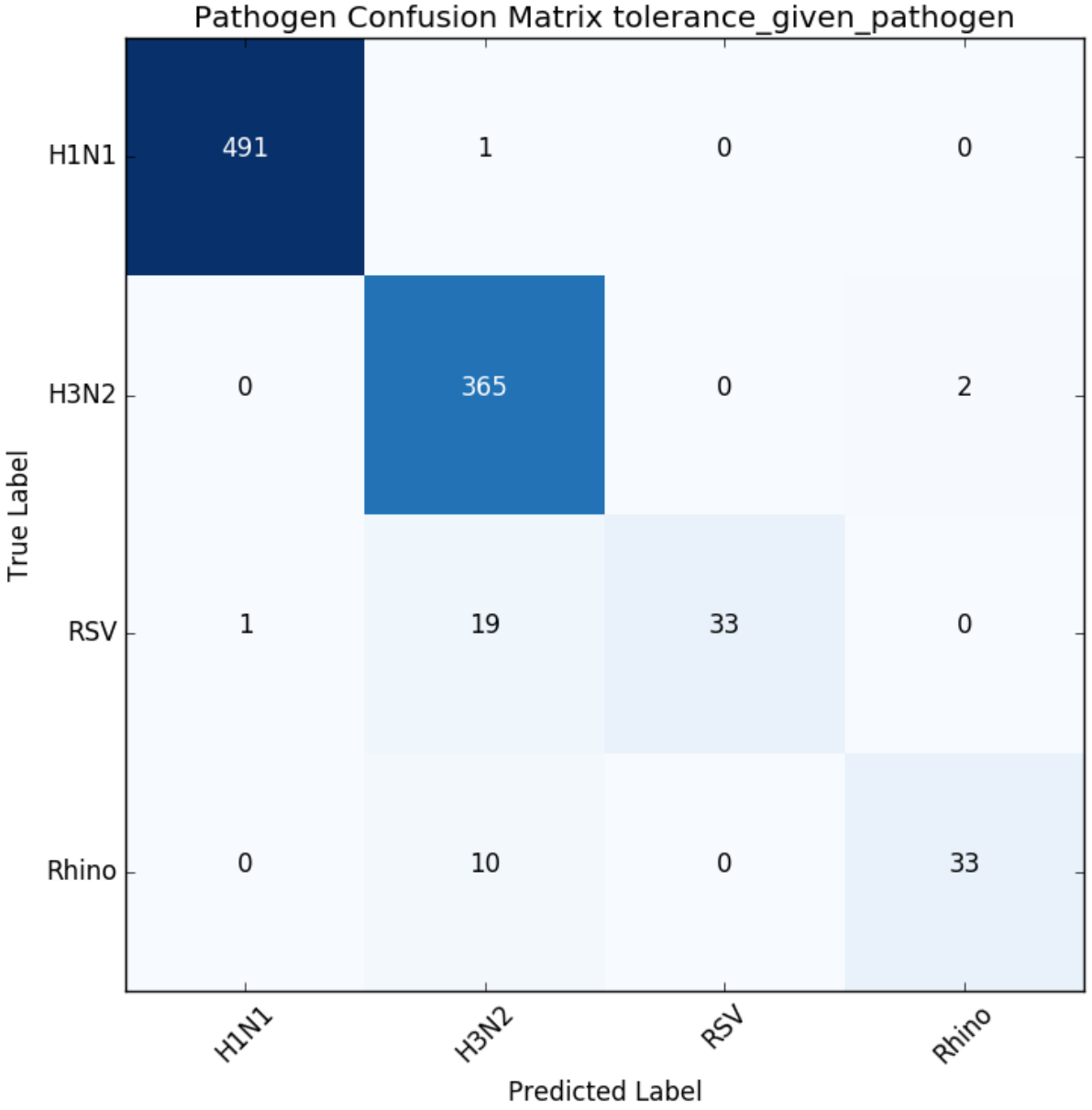
- NLL for both_separate:

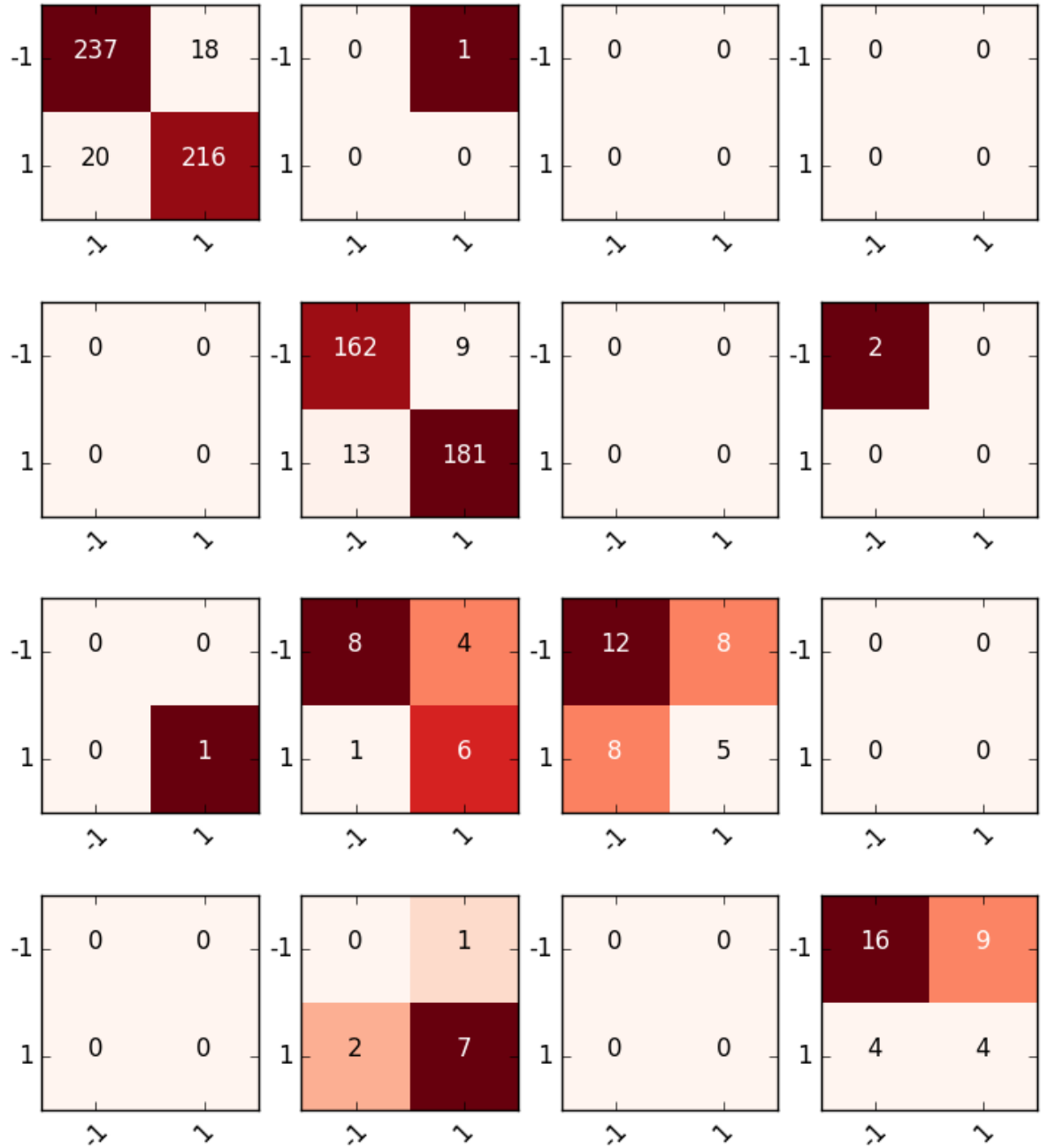
$$nll = -\sum_{i}^{m} 1\{Y_i = y\}\left[\log\left(P(T_i = t)\right) + \log(P(P_i = p))\right] \qquad P(Y) = P(T)P(P)$$

- NLL for tolerance_given_pathogen:

$$nll = -\sum_{i}^{m} 1\{Y_i = y\}\log\left(P(P_i = p)\right) + \log\left(\sum_{p} P(P_i = p)\,P(T_i = t|P_i = p)\right) \qquad P(Y) = P(P)\sum_{p} P(P)P(T|P)$$


Scores for various RF Algorithms
Score
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.869
0.965
0.86
0.861
0.837
0.864
0.881
0.868
-0.355
-0.199
-0.593
-0.557
-0.610
-0.567
-0.548
-0.538
accuracy
negative log loss
tolerance_only
pathogen_only
both_together
both_separately
pathogen_given_tolerance
pathogen_given_tolerance_soft
tolerance_given_pathogen
tolerance_given_pathogen_soft


Pathogen Confusion Matrix tolerance_given_pathogen
True Label
Predicted Label
H1N1
H3N2
RSV
Rhino
491
1
0
0
0
365
0
2
1
19
33
0
0
10
0
33


237
18
20
216
0
1
0
0
0
0
0
0
0
0
0
0
162
9
13
181
0
0
0
0
2
0
0
0
0
0
8
4
1
6
12
8
8
5
0
0
0
1
0
0
0
1
2
7
0
0
0
0
16
9
4
4

Key Transcripts

- Drivers of tolerance “common” to all pathogens: intersection of top-2000 features of level-2 classifiers
- We get 16
- Follow it up with PANTHER

Entrez ID	Function
9693	Rap guanine nucleotide exchange factor 2
2352	Folate receptor
9631	Nucleoporin 155
5413	Septin 5
2812	Glycoprotein Ib platelet beta subunit
3430	Interferon induced protein 35
3431	SP110 nuclear body protein
6772	Signal transducer and activator of transcription 1
53335	B-cell CLL/lymphoma 11A
115207	Potassium channel tetramerization domain containing 12
6925	Transcription factor 4
440026	Transmembrane protein 41B
3119	Major histocompatibility complex, class II, DQ beta 1
23189	KN motif and ankyrin repeat domains 1
26137	Zinc finger and BTB domain containing 20
375346	Transmembrane protein 110
9380	Glyoxylate and hydroxypyruvate reductase

PANTHER Pathway	Function
JAK/STAT signaling pathway	Activated by both cytokines and interferons; allows for rapid and direct transduction of an extracellular signal into the nucleus
Enkephalin release	Opiod peptides that are found at high levels in the brain and endocrine tissues; play an important role in behavior, pain, cardiac function, cellular growth, immunity, and ischemic tolerance
Metabotropic glutamate receptor group I pathway	Found in forebrain and cerebellum; its antagonists have antidepressant-like activity in a variety of preclinical models; potential drug targets for ischemia, schizophrenia, and epilepsy
Histamine H2 receptor mediated signaling pathway	Activation results in many physiological responses: secretion of gastric juices, smooth muscle relaxation, inhibition of antibody synthesis, T-cell proliferation and cytokine production
Interferon-gamma signaling pathway	IFNs are pleiotropic cytokines that mediate anti-viral responses, inhibit proliferation and participate in immune surveillance and tumor; IFN-gamma, that is produced by activated T cells and natural killer cells, activates JAK-STAT pathway
Nicotine pharmacodynamics pathway	Nicotine causes cell depolarization and an influx of calcium through voltage dependent calcium channels, triggering the release of epinephrine from the chromaffin vesicles to the bloodstream, which leads to increase of heart rate and blood pressure, and elevation of blood glucose level
GABA-B receptor II signaling	Stimulate the opening of K^+ channels which hyperpolarizes the neuron; considered inhibitory receptors and decrease the cell's conductance to Ca^{2+}
EGF receptor signaling pathway	Mediate cellular signaling pathways involved in growth and proliferation in response to the binding of a variety of growth factor ligands
CCKR signaling map	Binding of gastrin (responsible for stimulation of acid secretion from the parietal cell) or CCK to their common cognate receptor triggers the activation of multiple signal transduction pathways that relay the mitogenic signal to the nucleus and promote cell proliferation
Cadherin signaling pathway	Involved in many biological processes, such as development, neurogenesis, cell adhesion, and inflammation; implicated to be involved in many disease, such as cancer; cadherin-catenin complexes are important sensors and transmitters of the extracellular cues inside the cell body and into the nucleus
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	G-protein receptor activated pathways; a number of activated receptors can bind to and activate the associated heterotrimeric G-protein consisting of either Gi alpha or Gs alpha
Inflammation mediated by chemokine and cytokine signaling pathway	Upon binding to a family of G-protein coupled seven-transmembrane receptors, chemokines (chemotactic cytokines) control and direct trafficking and migration of immune cells

Future Work

- It is very flexible. Any classifier can be used locally to classify on the output of a particular hierarchy level. Rank-SVM?
- Deeper models, by incorporating more than just two outputs: “severity score” that signifies the viral load.
  - The meta-classifier allows lower outputs to be learnt better by the virtue of highly-discriminating outputs in upper layers
- Multi-omics