

# Improving TCR-pMHC binding prediction algorithms

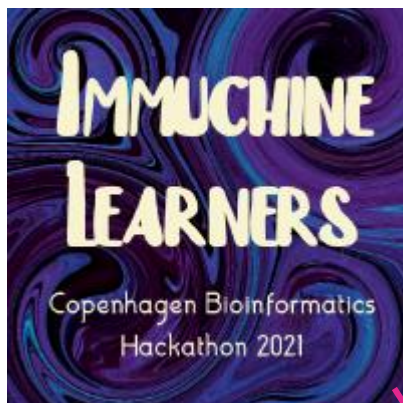
Copenhagen Bioinformatics Hackathon 2021

Team: Immuchine-learners





# Our Amazing Team



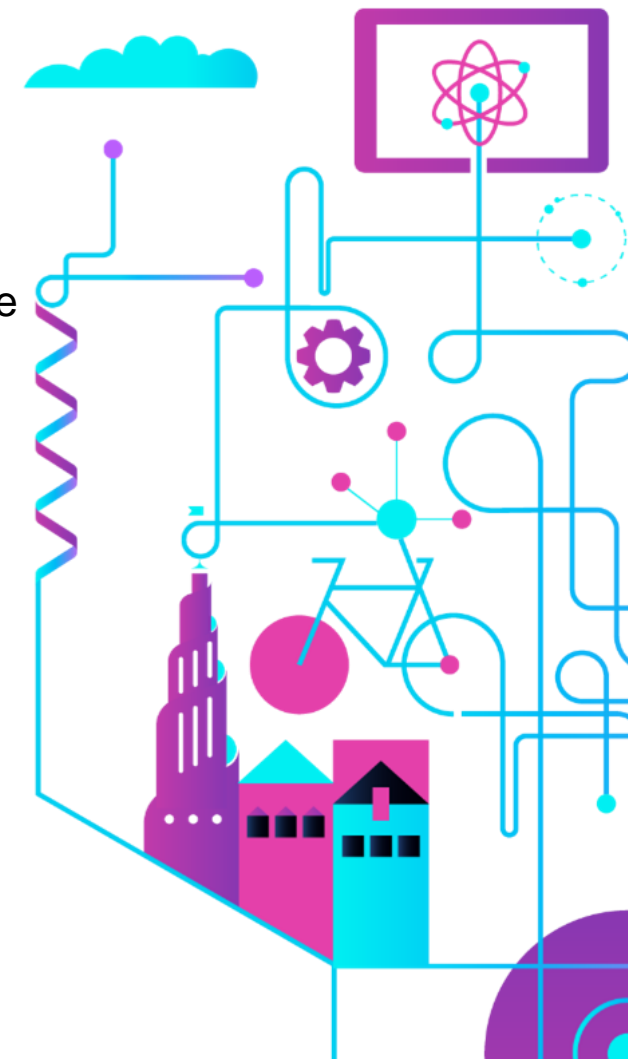
Shashank Yadav  
Research Associate  
@IIT Delhi



Dhvani Vora  
PhD candidate  
@IIT Delhi

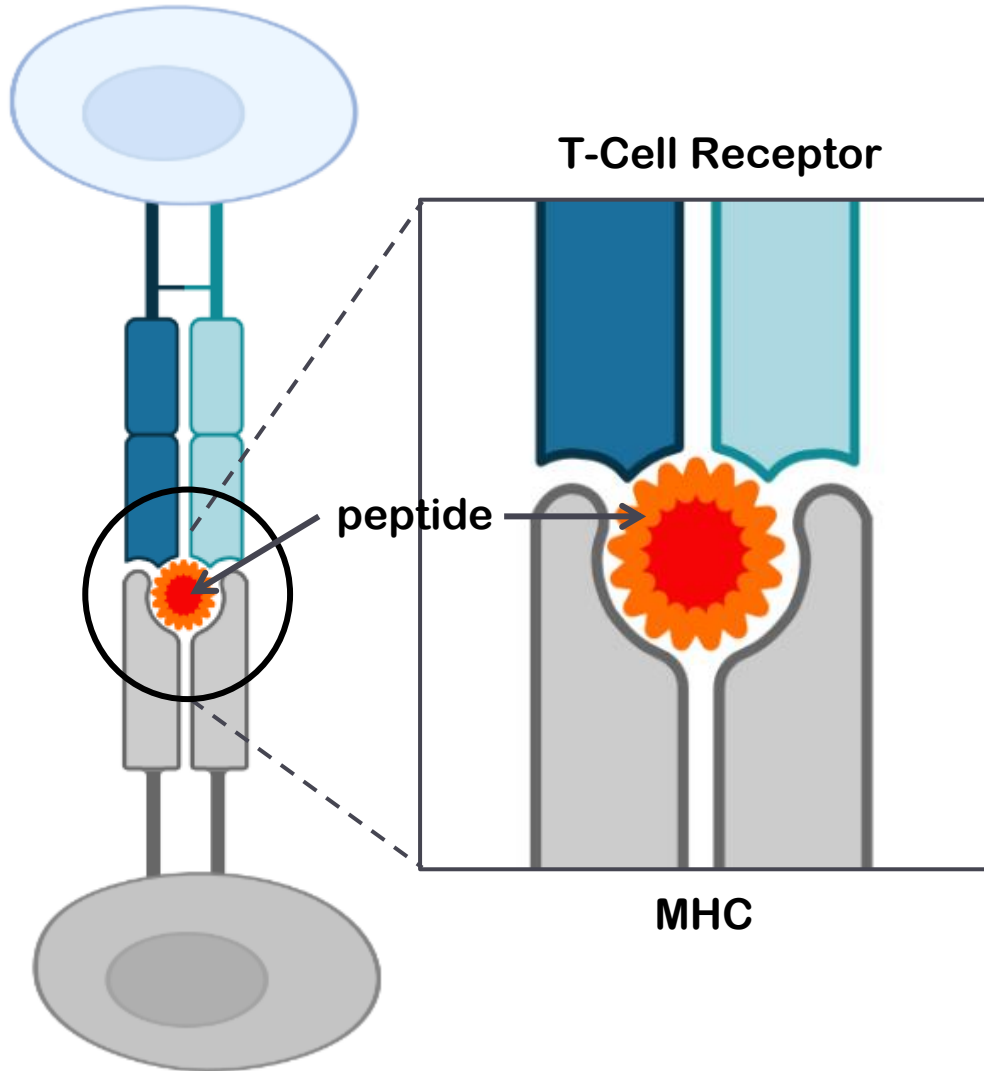


Anu Oswal  
Undergrad student  
@MIT-ADT Pune

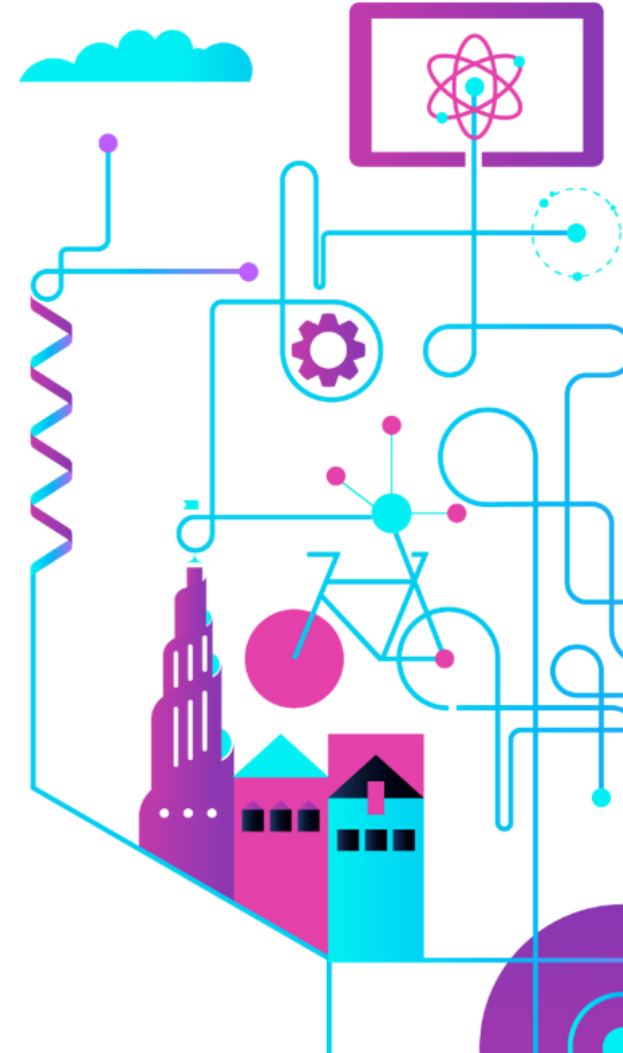




# Background



- T-cells look for (and destroy) suspicious entities in the body
- Antigenic peptides are displayed by the MHC
- TCR-pMHC: keystone interaction of the adaptive immune system



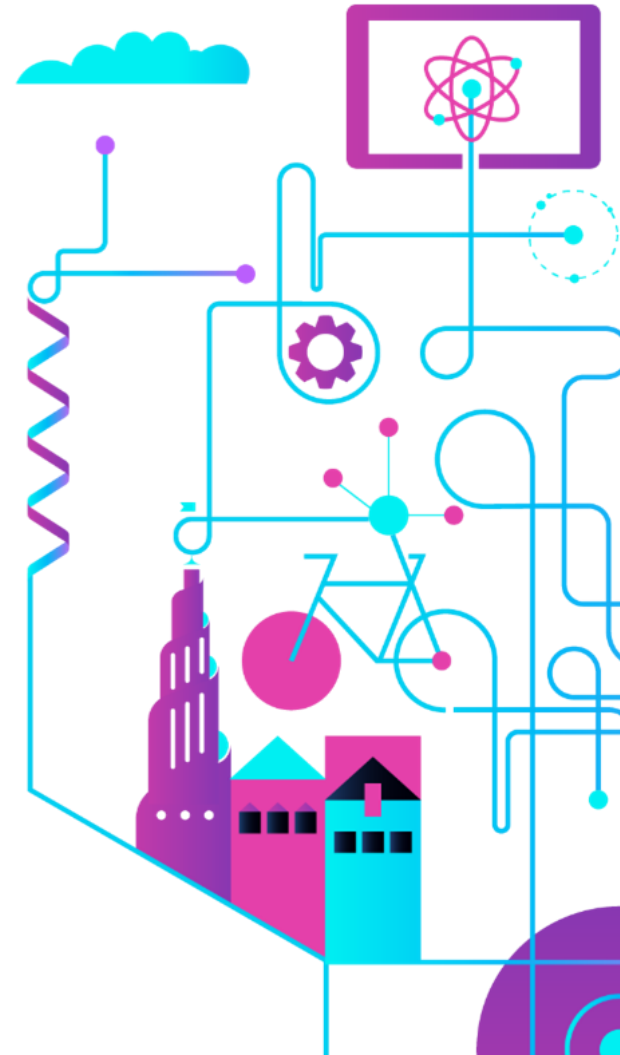


# The Challenge

## Dataset:

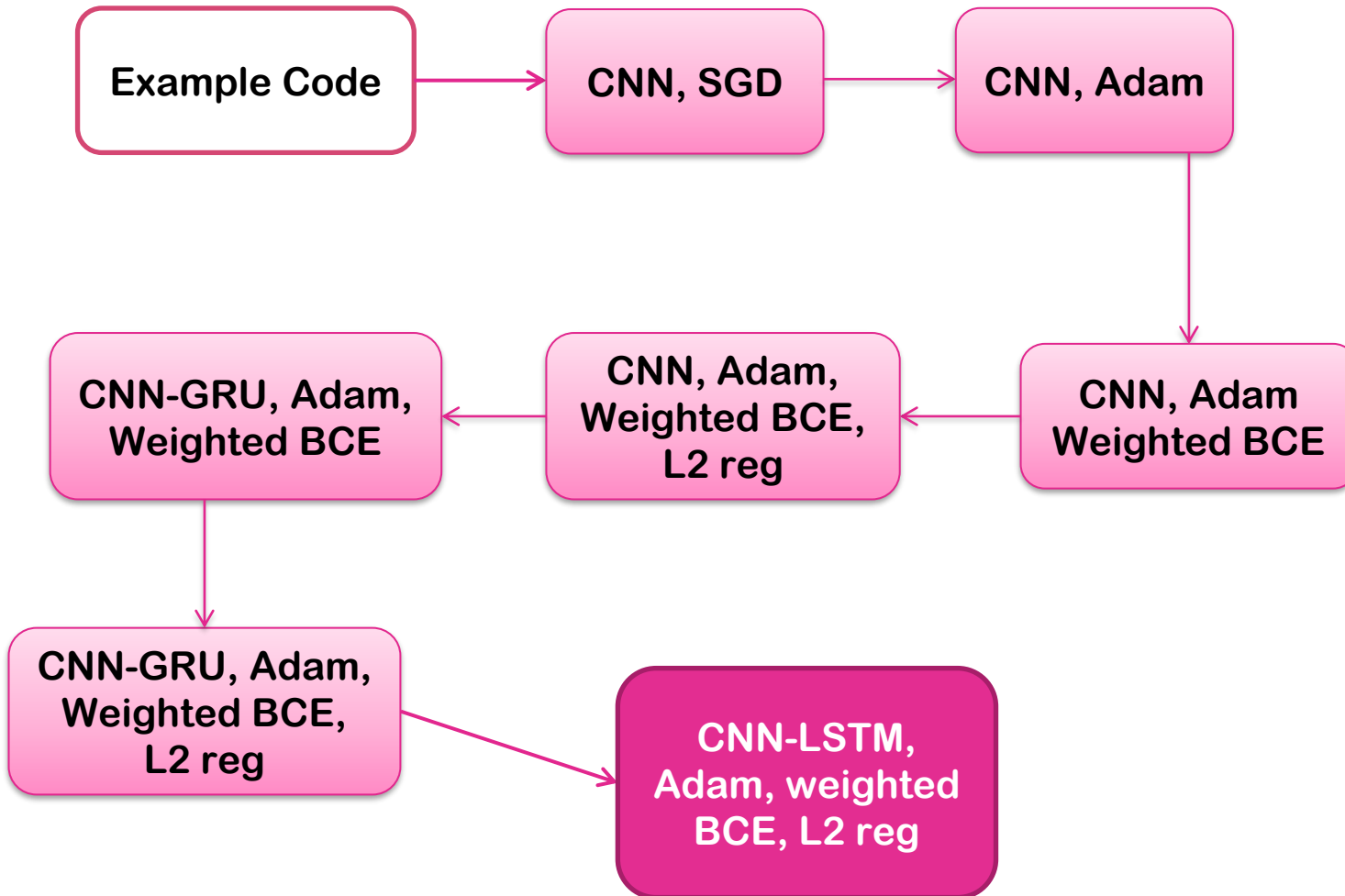
- ~8000 Peptide sequence fragments
- 25-75 Class imbalanced data
- Sequence and energy parameters

Make a generalizable model to accurately predict both classes: improve test MCC score





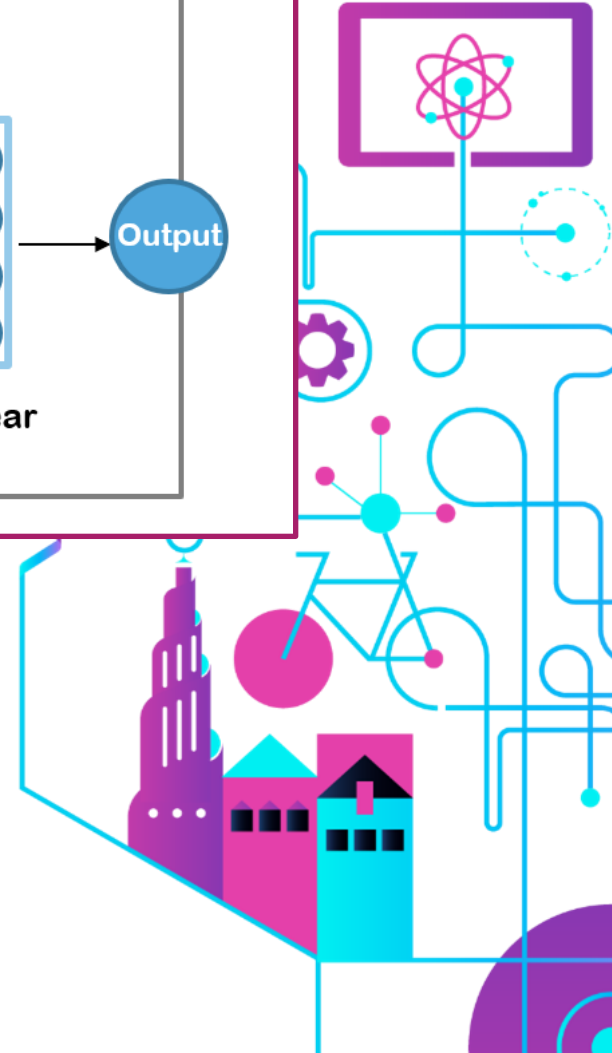
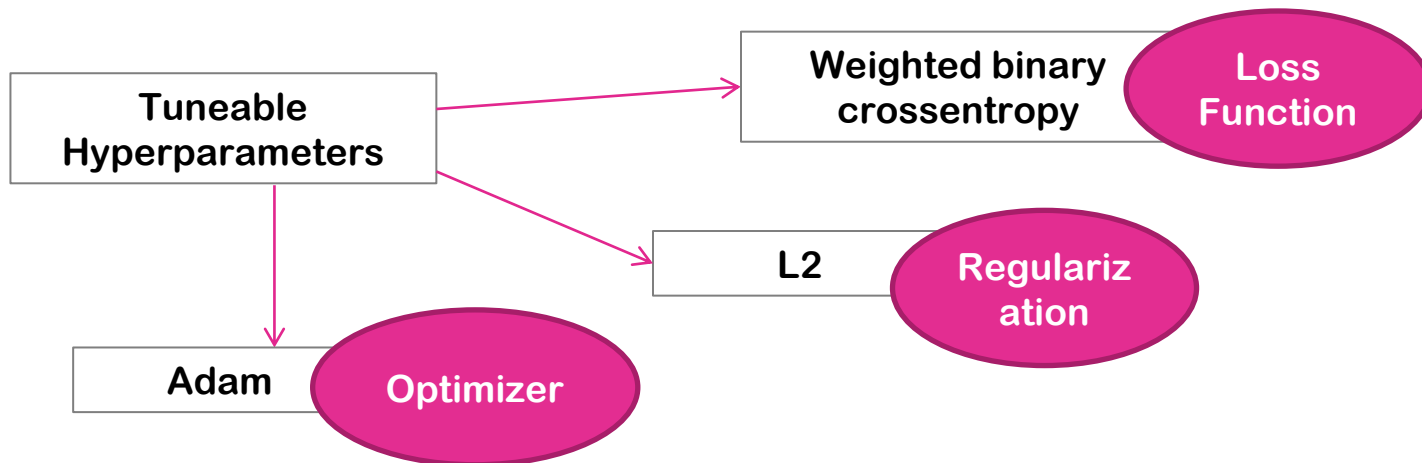
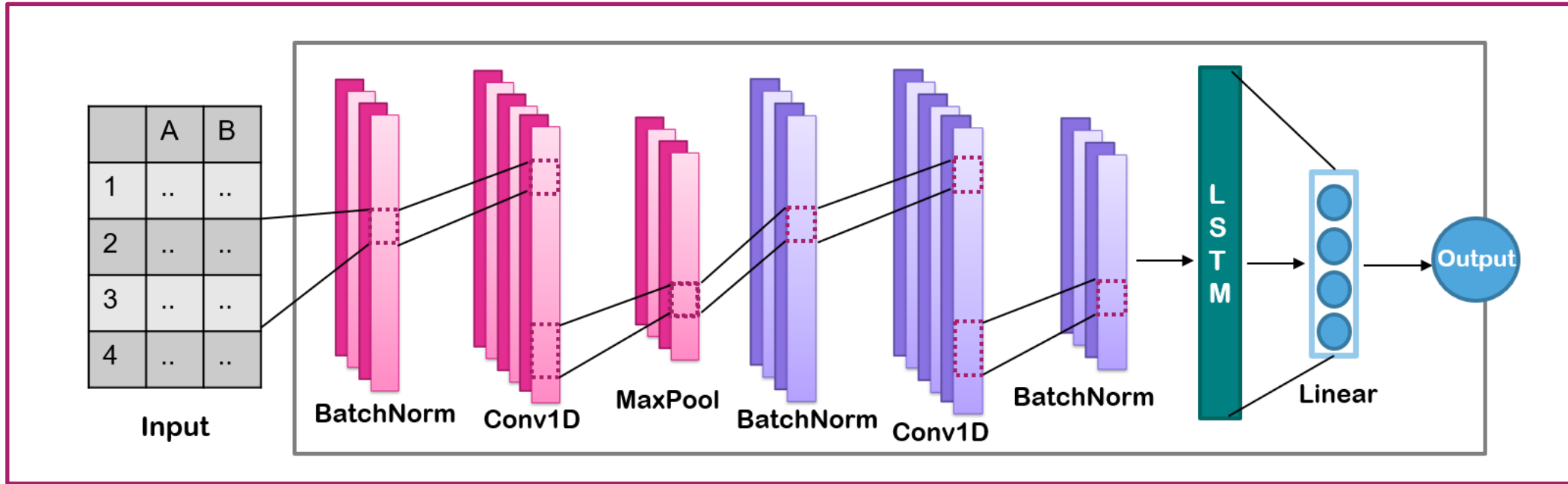
# The Predictinator 2000 (and what we did new)







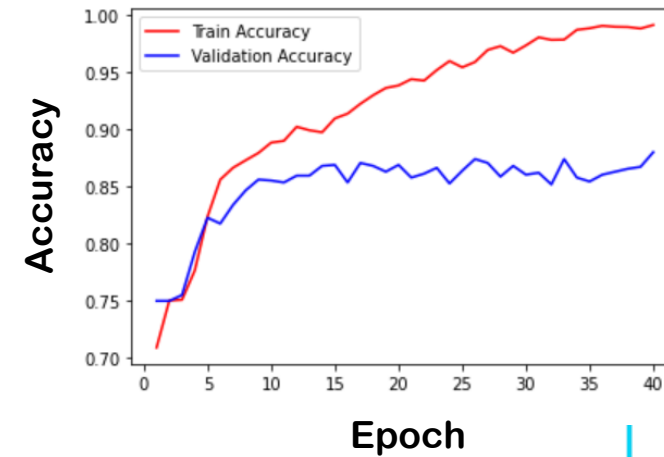
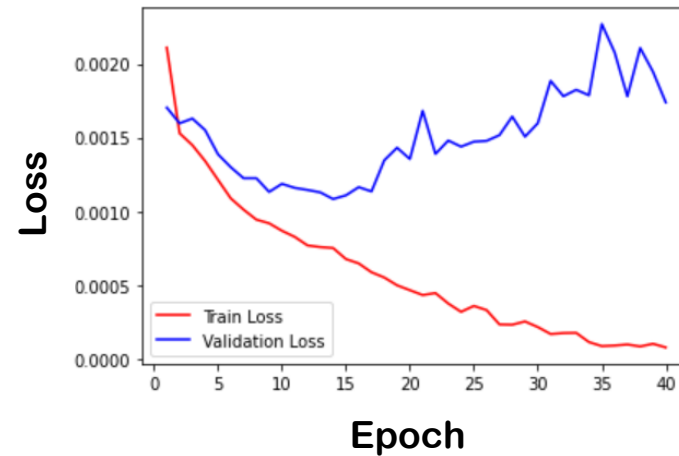
# The Predictinator 2000 (and what we did new)





# Model performance

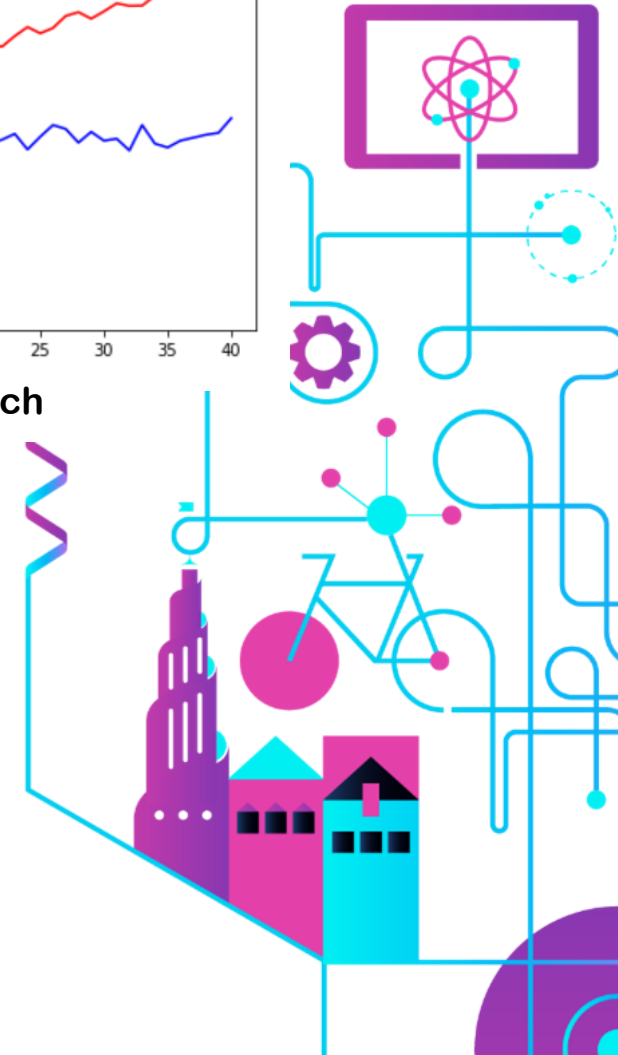
Metric	Train	Val
Accuracy	0.99	0.88
MCC	0.98	0.66
Precision	0.99	0.87
Recall	0.99	0.81



## Scope for improvement:

- Extracting structural features
- Experimenting with different representation schemes
- Solutions for class imbalance: focal loss, sample weights etc
- Interpretability of the model's prediction

(caution: Binding of TCR-pMHC does not always correlate to T-cell activity)



# Thank you very much!

Link to the code:

<https://github.com/biohackathon-immuchine-learners/tcr-pmhc>

