USING A HYBRID-ML MODEL TO PREDICT IMMUNOGENICITY OF MOLECULES

<Q|Team>

1M+

The Dataset + Classical Limitations

- Had SMILES ids for different small molecules as identifiers, as well as other molecular data (number of different functional groups, valence electrons, etc.)
- Our classical ML models needed to be highly complicated in order to increase AUC to anywhere significantly over 60%, making them less generalizable and at risk of overfitting.
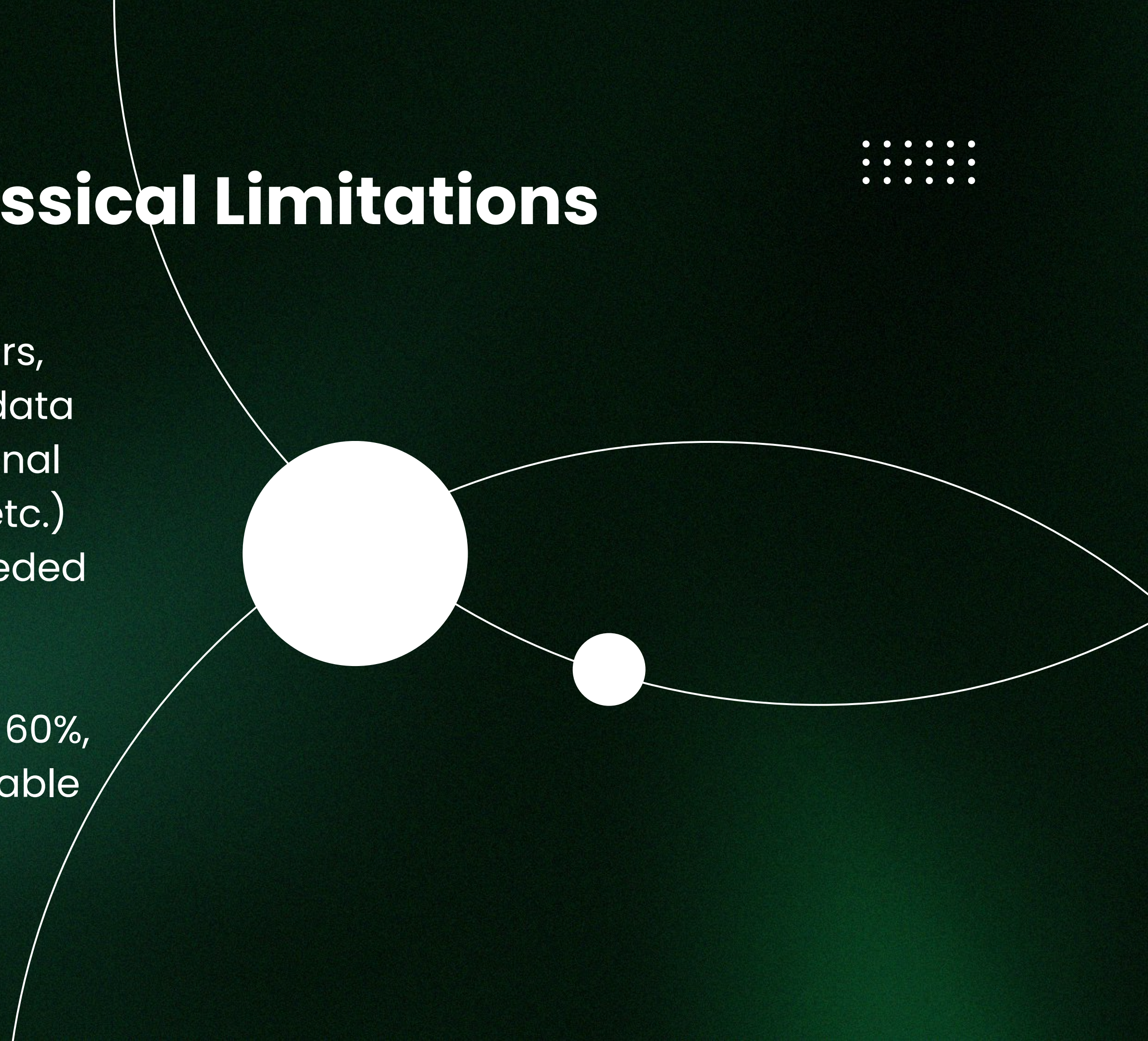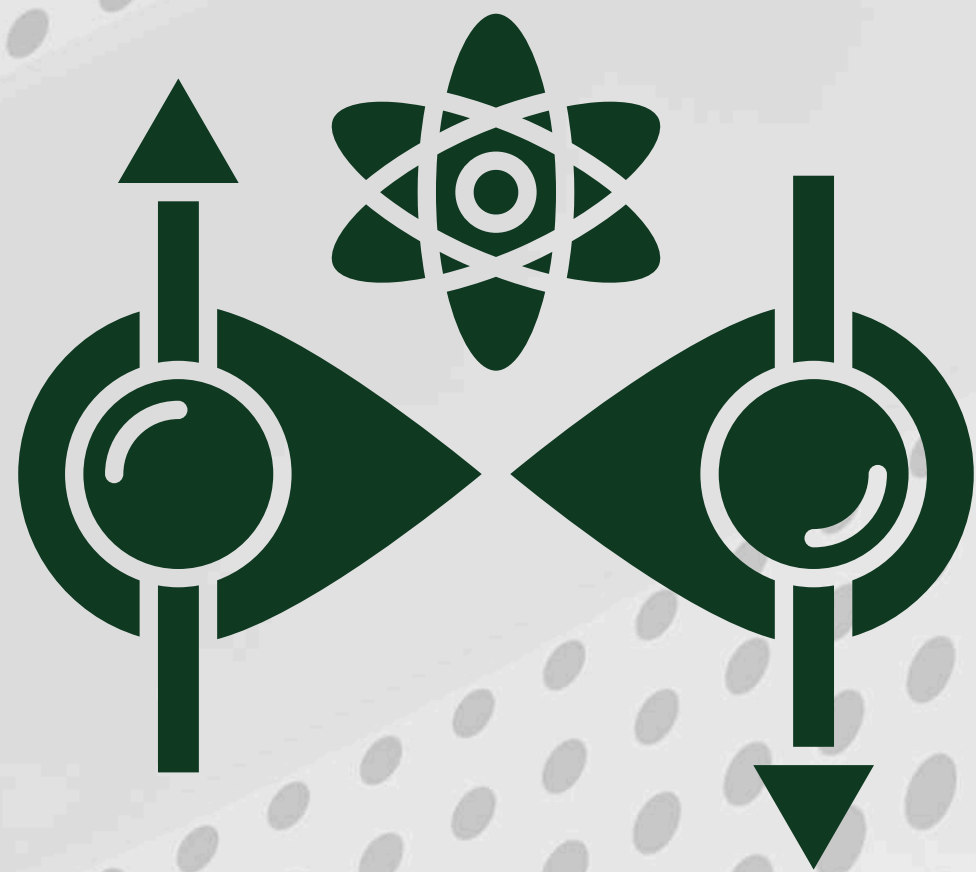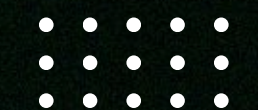
3-Qubit Circuit

- We used 3 qubits to represent the top 3 principal components, and then augmented our results with the classical model.

Ensemble
Accuracy: 84.38%
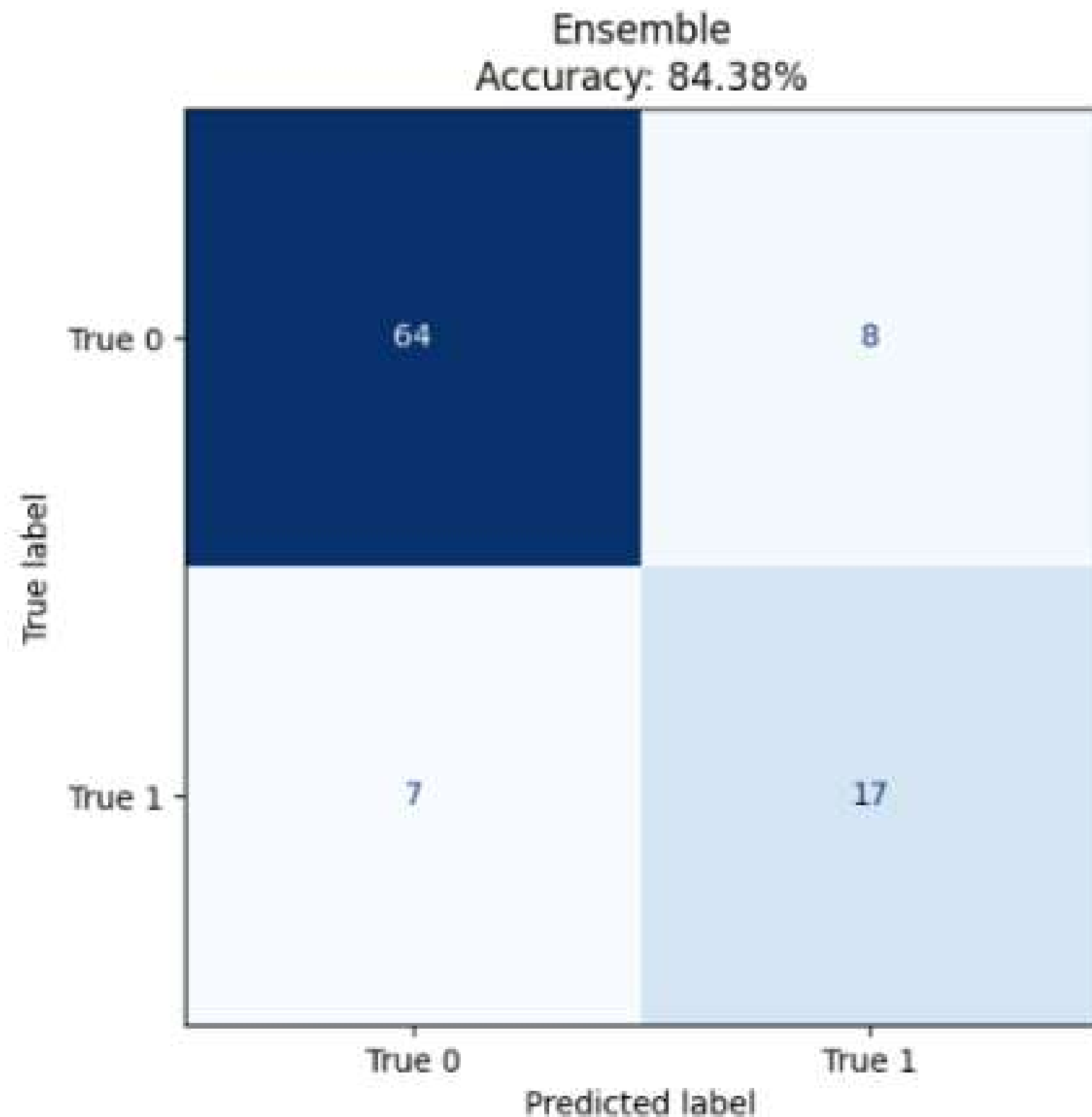
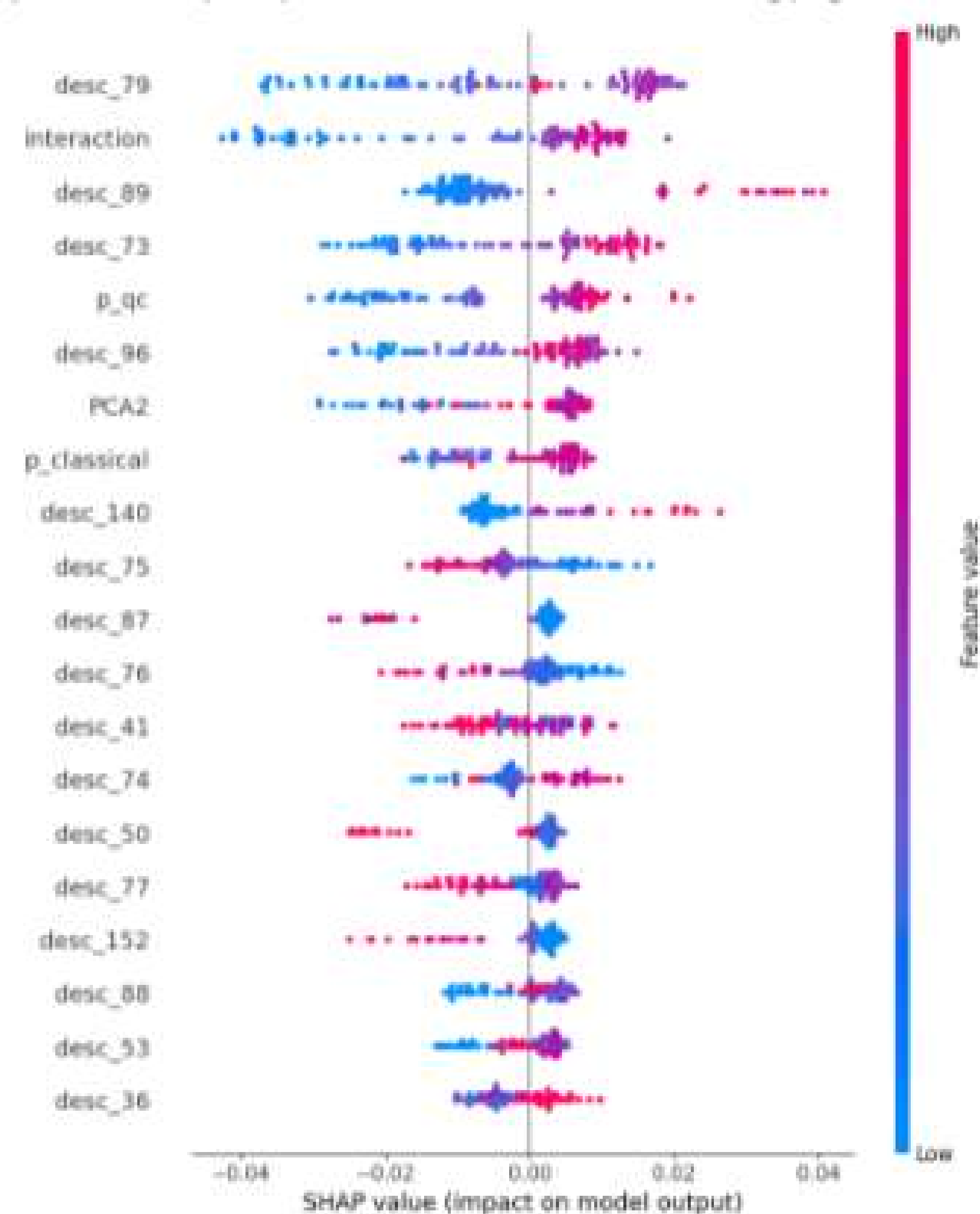# We used RF + XGBoost to get AUC of 0.880.

- False Positive Rate: 0.11
- False Negative Rate 0.29
- Precision 0.68
- Recall 0.71
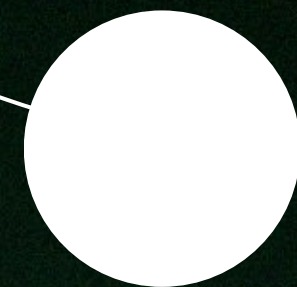- F1 Score 0.69
- Accuracy 0.84

Implications

- Desc_79, desc_89 and desc_73 are MOE MR VSA Descriptors, which reveal information about how much of the molecule's surface is available for van der Waals interactions.
- We theorise, using our biochemical knowledge, that a greater space availability for van der Waals interactions is also correlated with a greater chance of a reaction occuring, which would explain why they are top features.
- PCA2 is the second principle component piece, and p_qc is the component we derived using the quantum circuit. You can see it is more impactful than p_classical.

We have created a simpler, more explanatory and generalizable model for determining trends in molecular data to determine immunogenicity by using the quantum circuit.

THANK YOU FOR YOUR TIME!