

# Predicting Protein-Protein Interactions



# Overview

**Background:** What are PPIs and why do we want to model them?

**Feature Extraction:** How do we model PPIs?

**Deep Learning:** What model do we use?

**Model Evaluation:** How well did our model do?

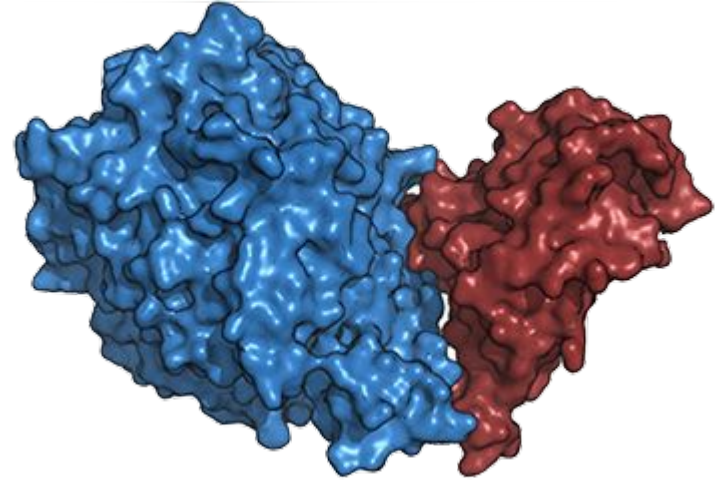


**James Pecore** 6:50 PM

ok i'll bite. what are protein-protein interactions?

# What are protein-protein interactions?

- PPIs drive biological activity
- Examples of PPIs:
  - Signal Transduction
  - Cell Metabolism
  - Membrane Transport
  - Muscle Contraction
- Modeling PPIs will save time and money
- PPIs are caused by intermolecular forces
- Knowing the chemical composition of proteins will help us predict these intermolecular forces



# What data do we have on proteins?

- Amino acid sequences are widely available
- Amino acids determine the structure and function of a protein
- Unstructured data with lots of poorly understood relationships
- UniProt Database has protein structures and binding partners



# Formulating our problem

Given a pair of proteins' amino acid sequences,  
predict whether those two proteins will interact.

# Feature Extraction: Conjoint Triad Method

Technique borrowed from Sun et. al.

Step One: cluster amino acids according to dipole and side chain volumes.

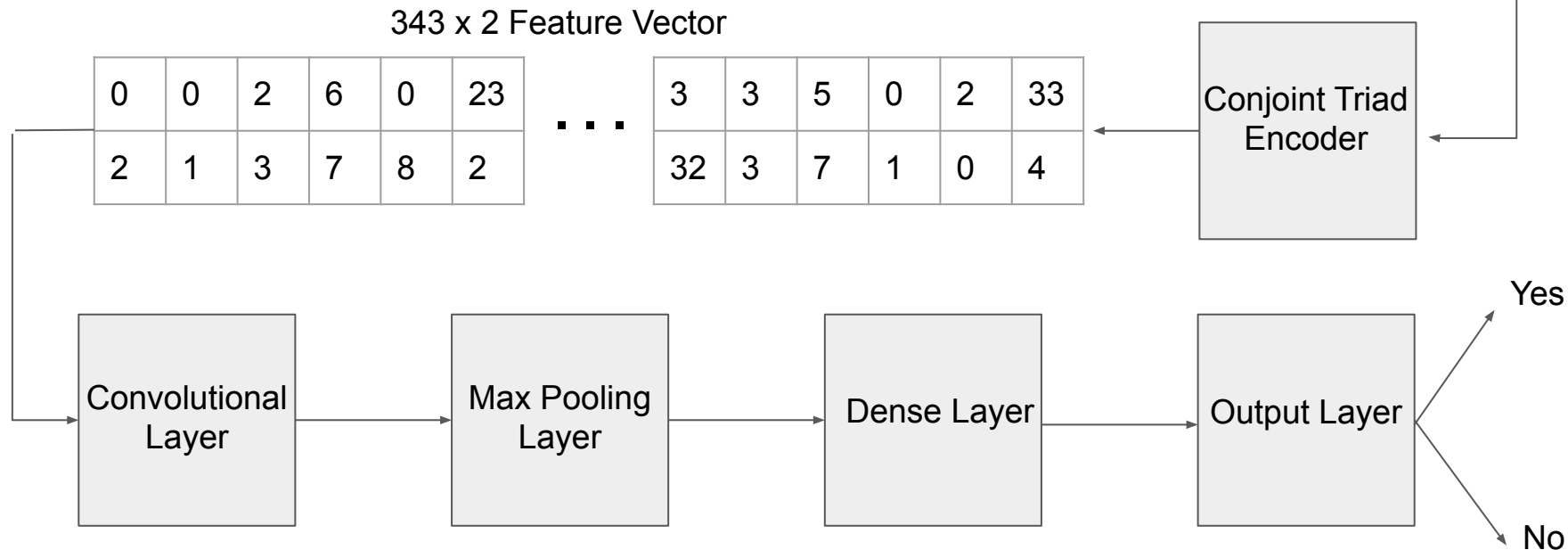
Step Two: replace amino acid labels with cluster numbers.

Step Three: count triads with sliding scale.



# CNN Architecture:

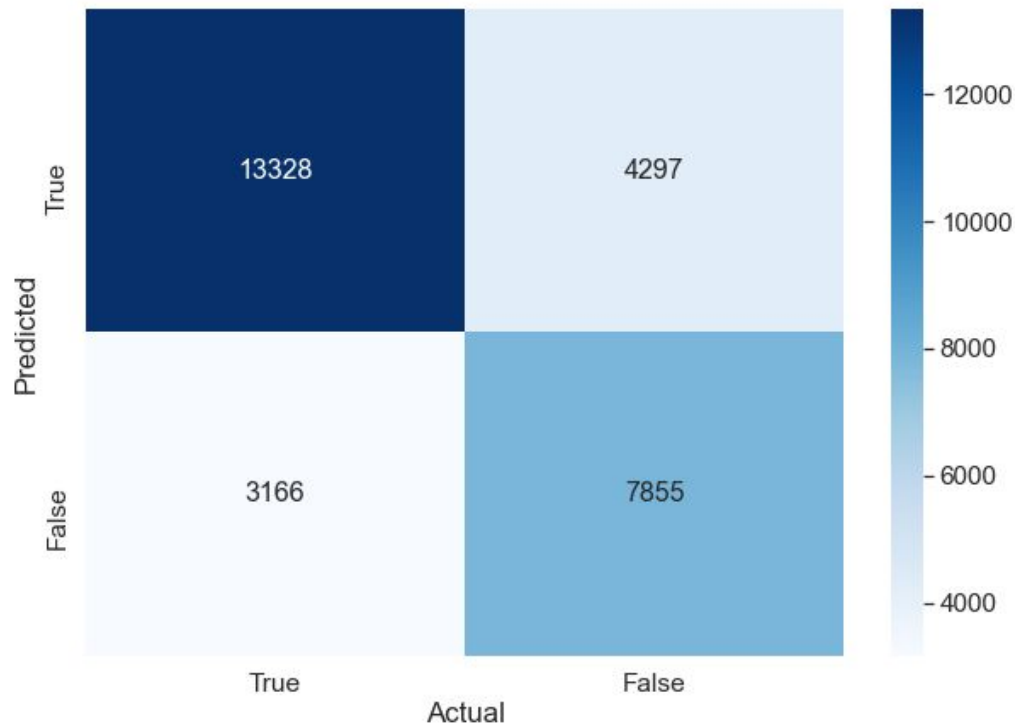
**Protein 1 Sequence: AGKEDPS...WYIKCCOI**  
**Protein 2 Sequence: VTSW1RK...DEYMPHF**





# Model Evaluation

Confusion Matrix of Protein-Protein Interactions



**Baseline Accuracy: 57%**

**Model Accuracy: 73.9%**

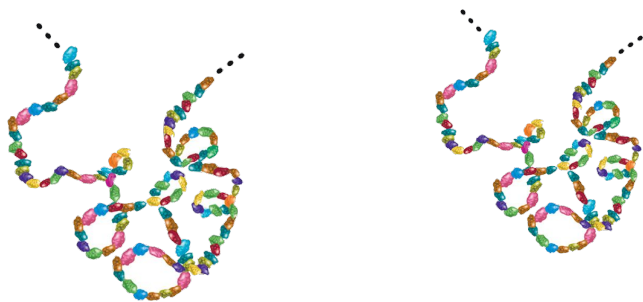
Specificity: 64.6%

Sensitivity: 80.8%

Precision: 75.6%

# Conclusions

- 17% improvement on baseline accuracy is impressive but still scratching the surface of model's potential
- Pursue more feature extraction: autocovariance, feature embedding, etc.
- Other deep learning techniques: Long-Short Term Memory
- Applications: Determine SARS-COV-2 cell receptor targets, generate "Interactome"
- More PPI modeling to come, stay tuned



thank you !!

