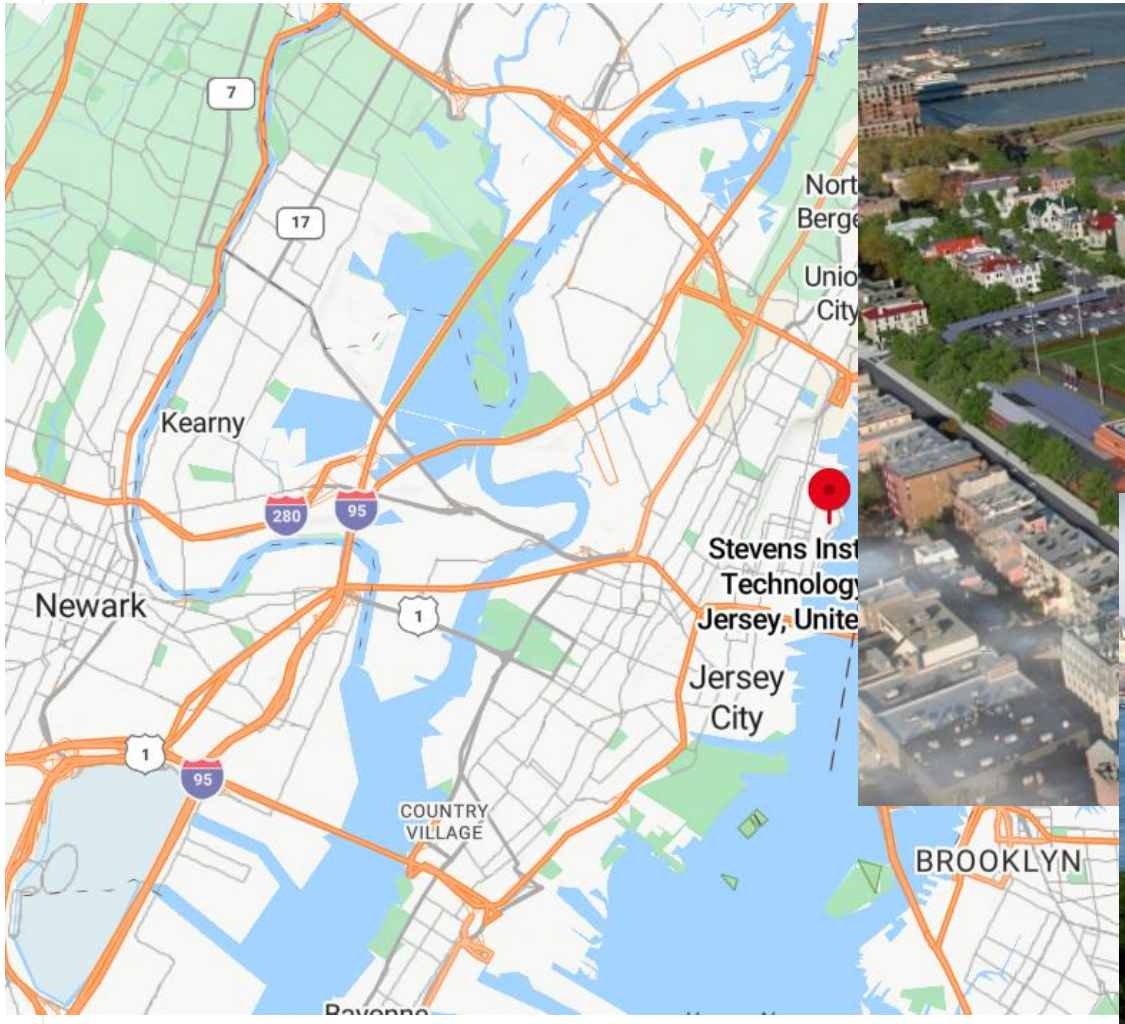
Predict the Viscosity of Concentrated Antibody Solutions Using Integrative Experimental and Computational Screening

CASSS HOS 2024

Pin-Kuang Lai
Department of Chemical Engineering and Materials Science
Stevens Institute of Technology

Sept 12, 2024


7
17
Kearny
280
95
Newark
1
Stevens Inst
Technology
Jersey, Unite
Jersey
City
1
95
COUNTRY
VILLAGE
BROOKLYN

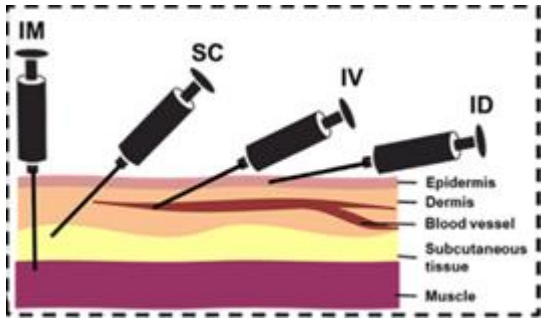
Outline

- Viscosity issue for high-concentration formulation
- Introduction of machine learning
- Machine learning models for antibody viscosity prediction
- DeepViscosity – novel deep learning model for viscosity prediction

Antibody drug delivery

Intravenous injection

Subcutaneous injection



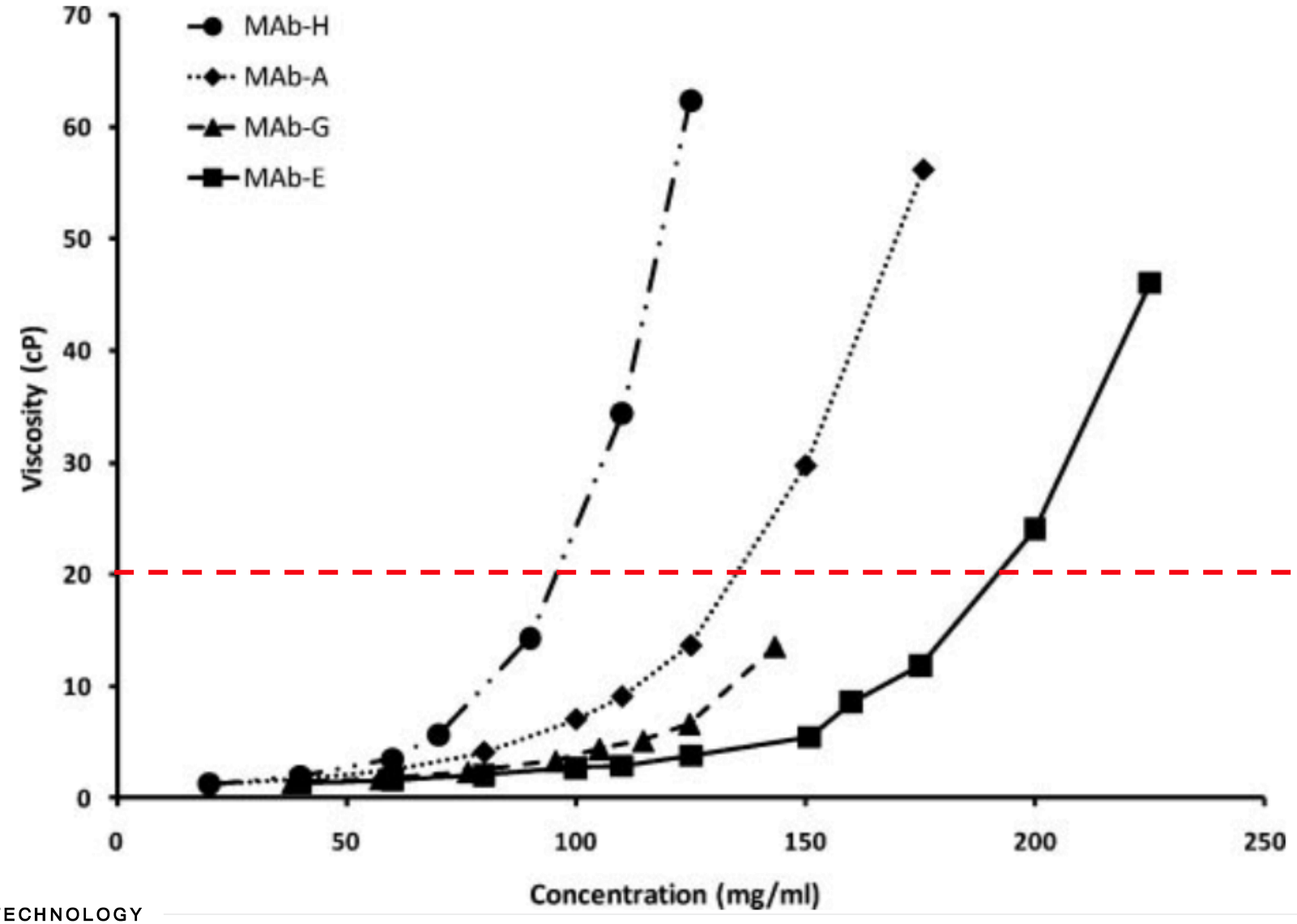
The volume of a syringe < 2 mL

SubQ needs high concentrations

High Conc. may lead to High Viscosity

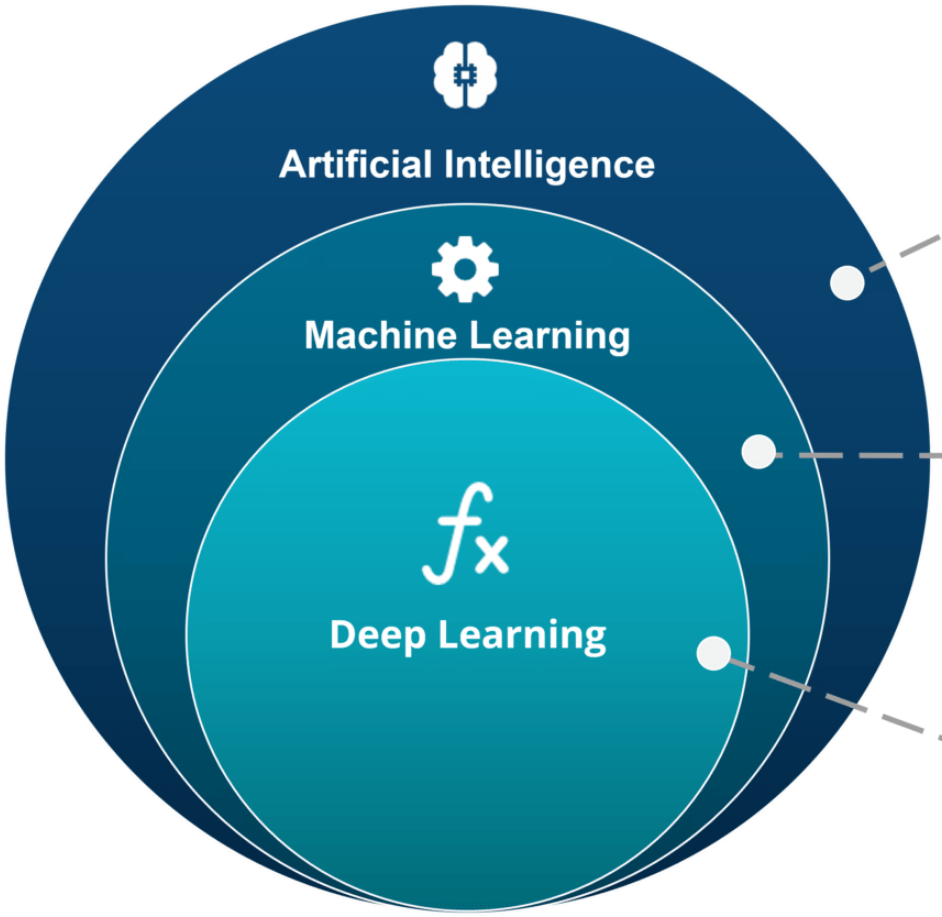
https://www.gograph.com/photo/nurse-preparing-dropper-for-intravenous-injection-gg64187859.html

Viscosity of mAbs



J Pharm Sci. 2010, 99:4812-29

AI / Machine Learning / Deep Learning



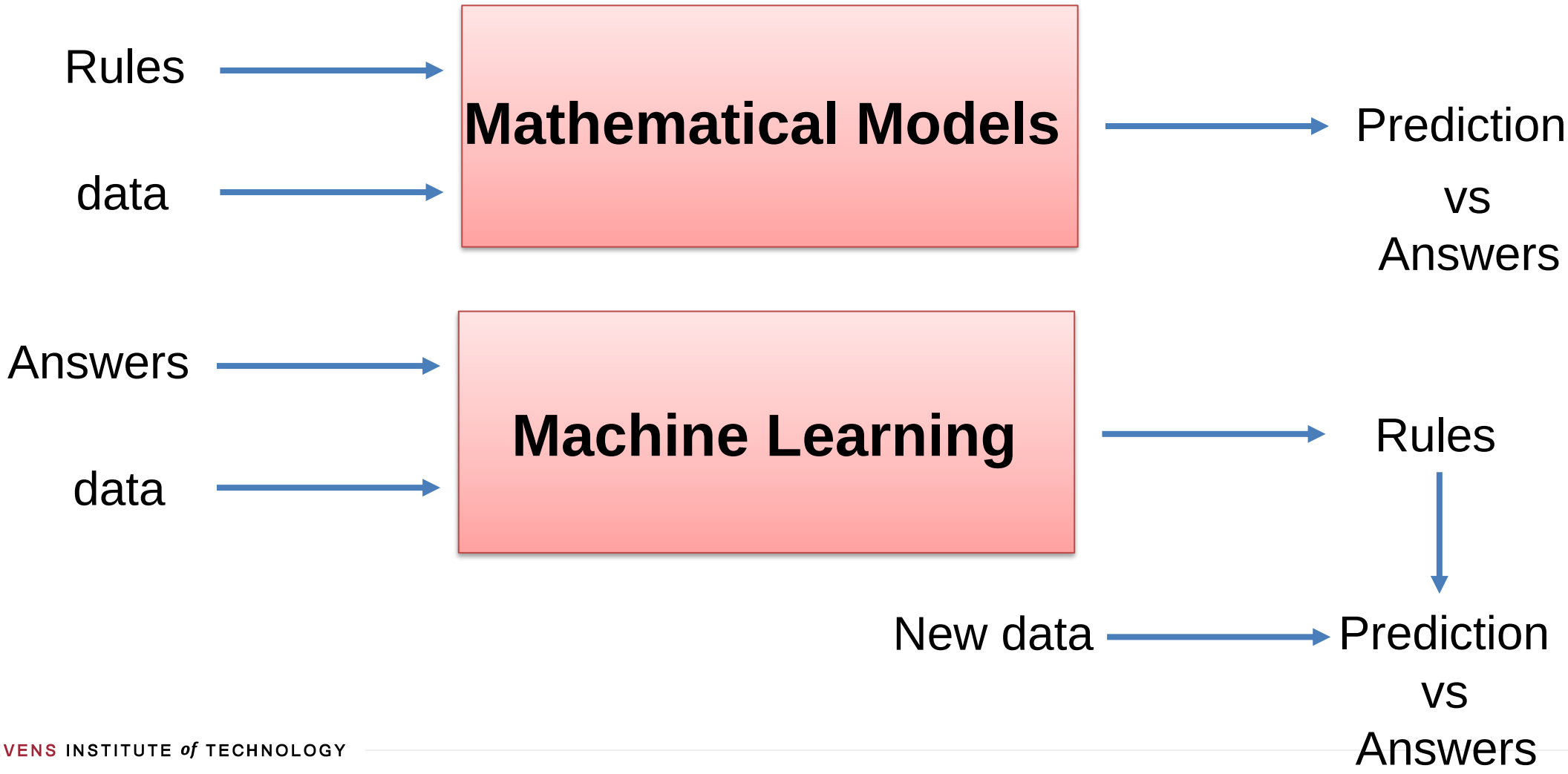
ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

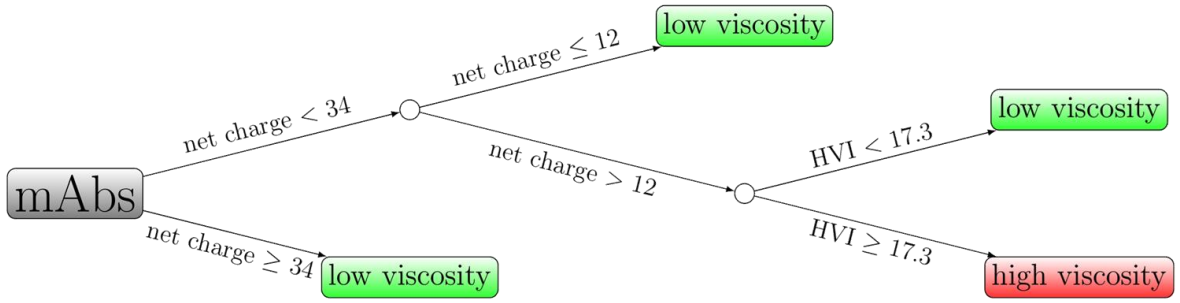
DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

https://www.edureka.co/blog/ai-vs-machine-learning-vs-deep-learning/

General Concepts of Machine Learning (ML)



ML for viscosity prediction



P. K. Lai and B. L. Trout et al., *Mol. Pharm.*,18 (3), 1167-1175, 2021

N = 27, commercial mAbs

$$\begin{cases} \text{Low viscosity: net charges\_VL} > -0.50 \\ \begin{cases} \text{Low viscosity: net charges\_VL} \leq -0.50 \text{ and N\_phobic\_VL} > 38.0 \\ \text{High viscosity: net charges\_VL} \leq -0.50 \text{ and N\_phobic\_VL} \leq 38.0 \end{cases} \end{cases}$$

P. K. Lai and B. L. Trout et al., *mAbs*,14(1), 2026208, 2022

N = 20, clinical mAbs from AZ

M. Mock and I. D. G. Campuzano et al., Development of in silico models to predict viscosity and mouse clearance using a comprehensive analytical data set collected on 83 scaffold-consistent monoclonal antibodies. *MAbs* **15**, 2256745 (2023).

E. K. Makowski and P. M. Tessier et al., Reduction of monoclonal antibody viscosity using interpretable machine learning. *MAbs* **16**, 2303781 (2024).

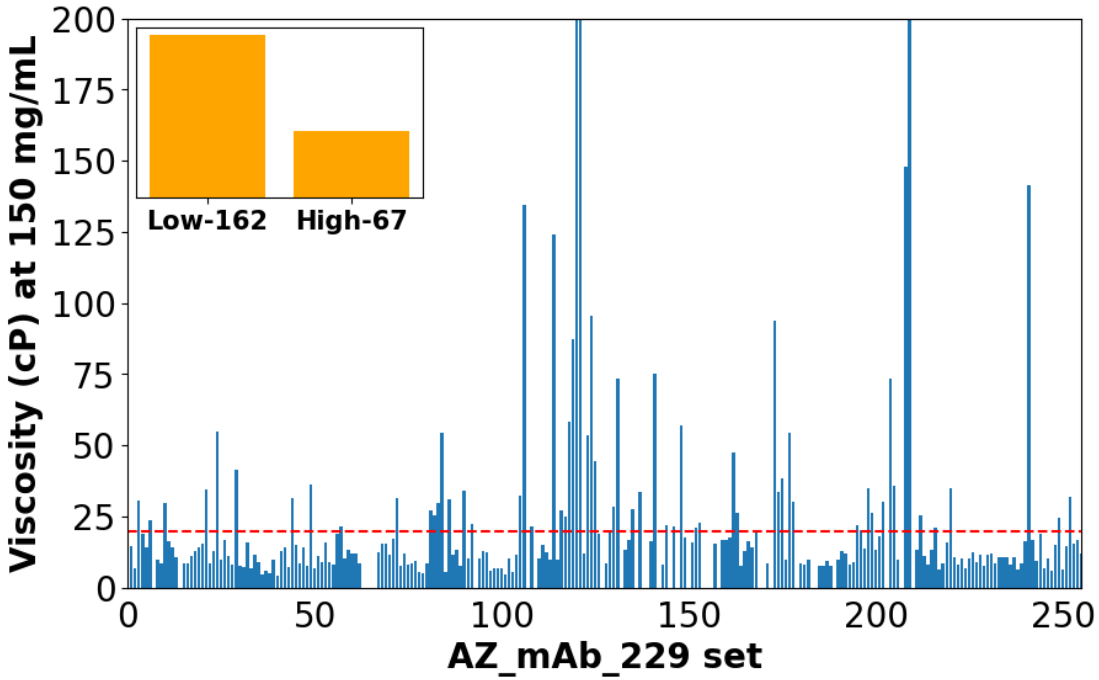
B. K. Rai and E. M. Bennett et al., Low-data interpretable deep learning prediction of antibody viscosity using a biophysically meaningful representation. *Sci Rep* **13**, 2917 (2023).

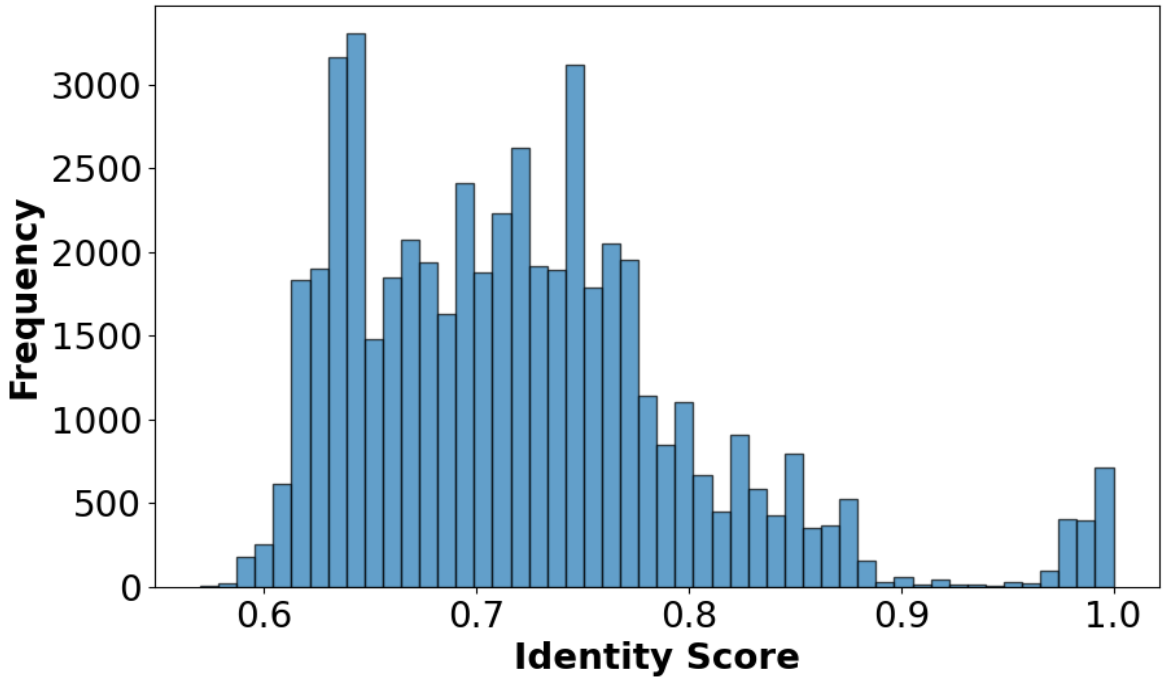
J. Schmitt and C. Grapentin et al., Predictive modeling of concentration-dependent viscosity behavior of monoclonal antibody solutions using artificial neural networks. *MAbs* **15**, 2169440 (2023).

Challenges: ML for viscosity prediction

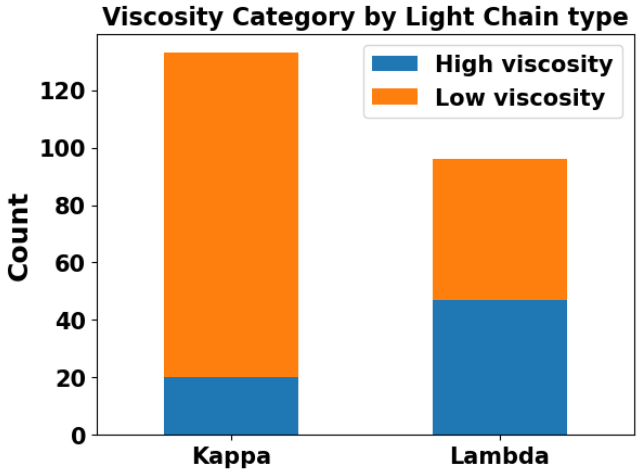
- Limited datasets
  -- might not capture the complexity of antibody interactions
  -- issues with generalizability
  -- not shared ➔ difficult to reproduce and compare
- Features for antibody
  -- sequence (antibody vs other proteins)
  -- structure (static vs dynamic)
- ML models
  -- Classical algorithms vs neural networks
  -- not shared ➔ difficult to reproduce and compare

Large-scale viscosity data generation

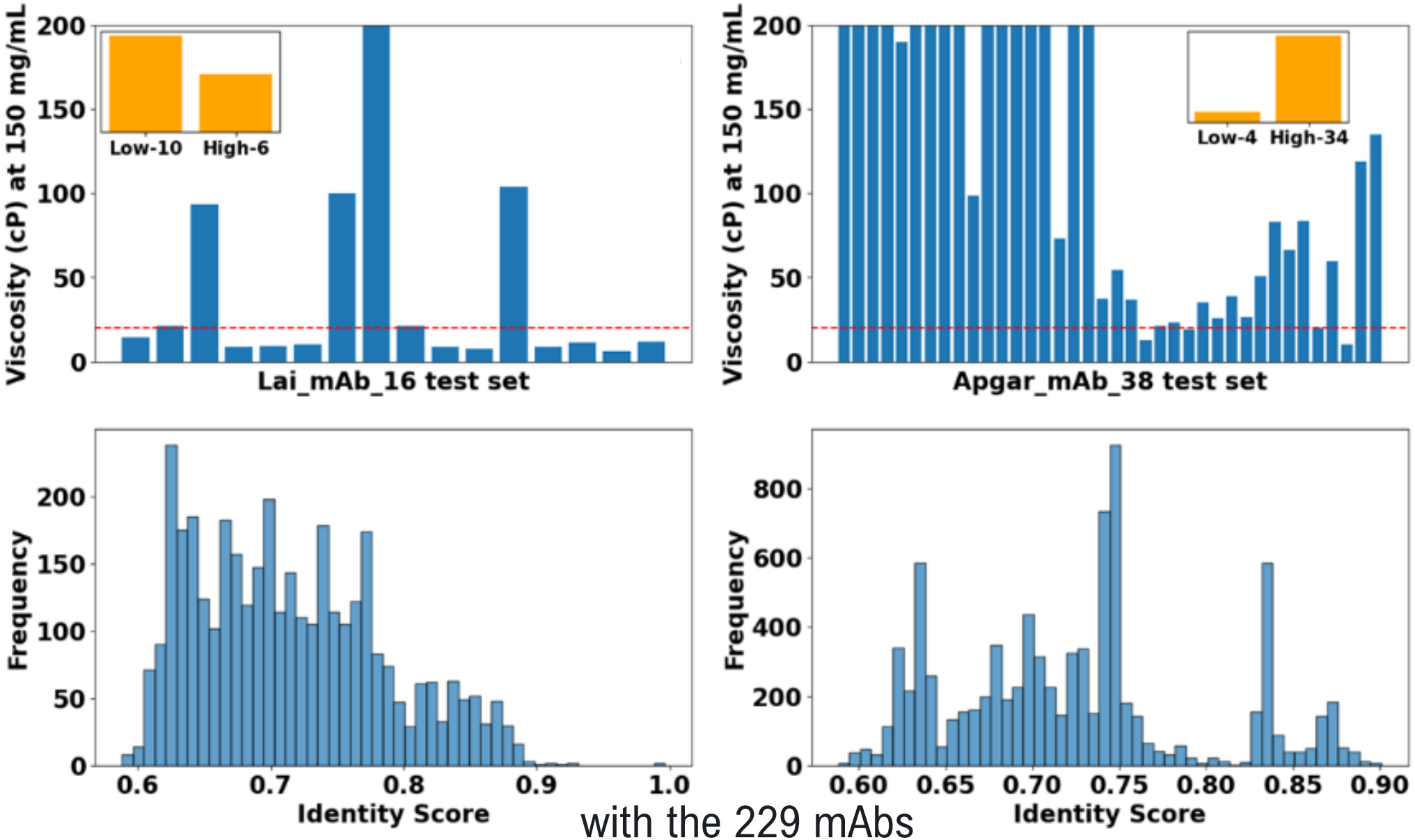




- 229 mAbs with experimental viscosity at high concentration (150 mg/mL) formulated at histidine buffer at pH 6.0 **obtained from AstraZeneca**.
- These mAbs cover a wide range of clinical to commercial mAbs.



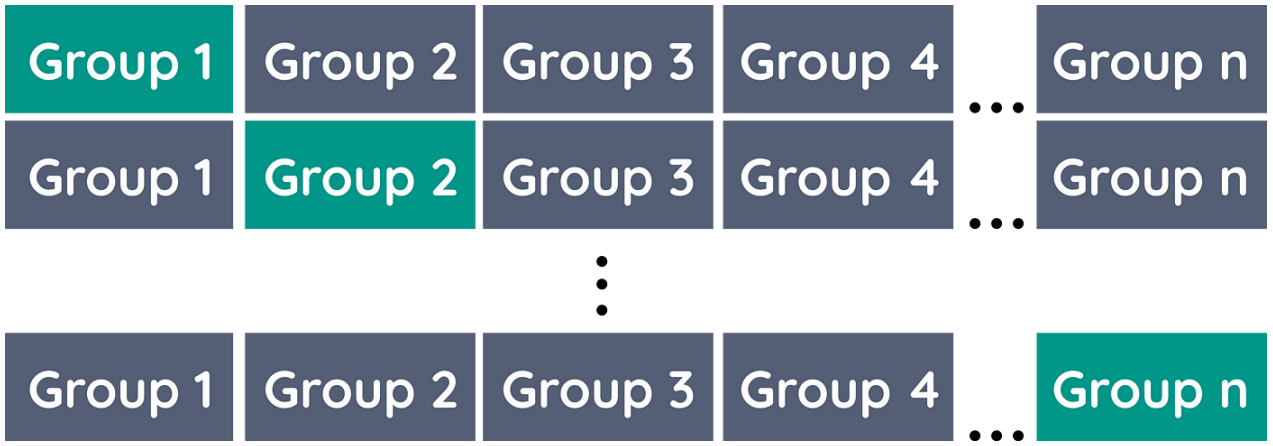
Independent datasets for model validation



with the 229 mAbs

Data processing

- The 229 mAbs were split into training and validation by the **Leave-One-Group-Out (LOGO)** method.



- A group (cluster) was defined by the Levenshtein distance of less than 10.
- In total, there are **102** groups.

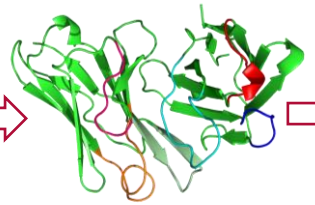
https://miro.medium.com/max/1400/1*wC3rtgM0DTec-ePTf0vAAg.png

DeepSP -- Deep learning-based spatial properties

Fv_sequence

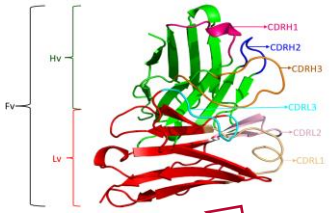
QVQLQE....TTAMDYWGQGSLVT
VSSDIQMTQSPSSLSASVGDRVTI
TG.....CQQGLPYTFGQGTKVEIK

Structure

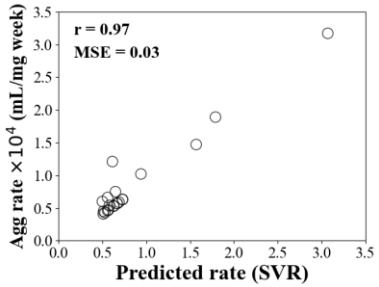


MD Simulations

Structural descriptors SAP_pos, SCM_pos, SCM_neg in 10 domains



Stability Properties like Aggregation rate etc

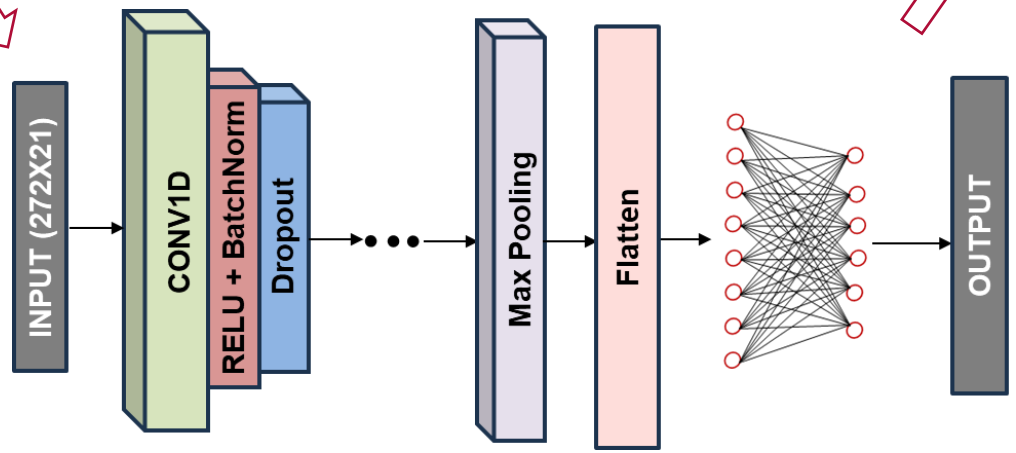


additional step of generating structures

takes hours to days to run on supercomputers

Computationally Expensive

One Hot Encoding



Comput. Struct. Biotechnol. J. 2024 (23), 2220-2229.

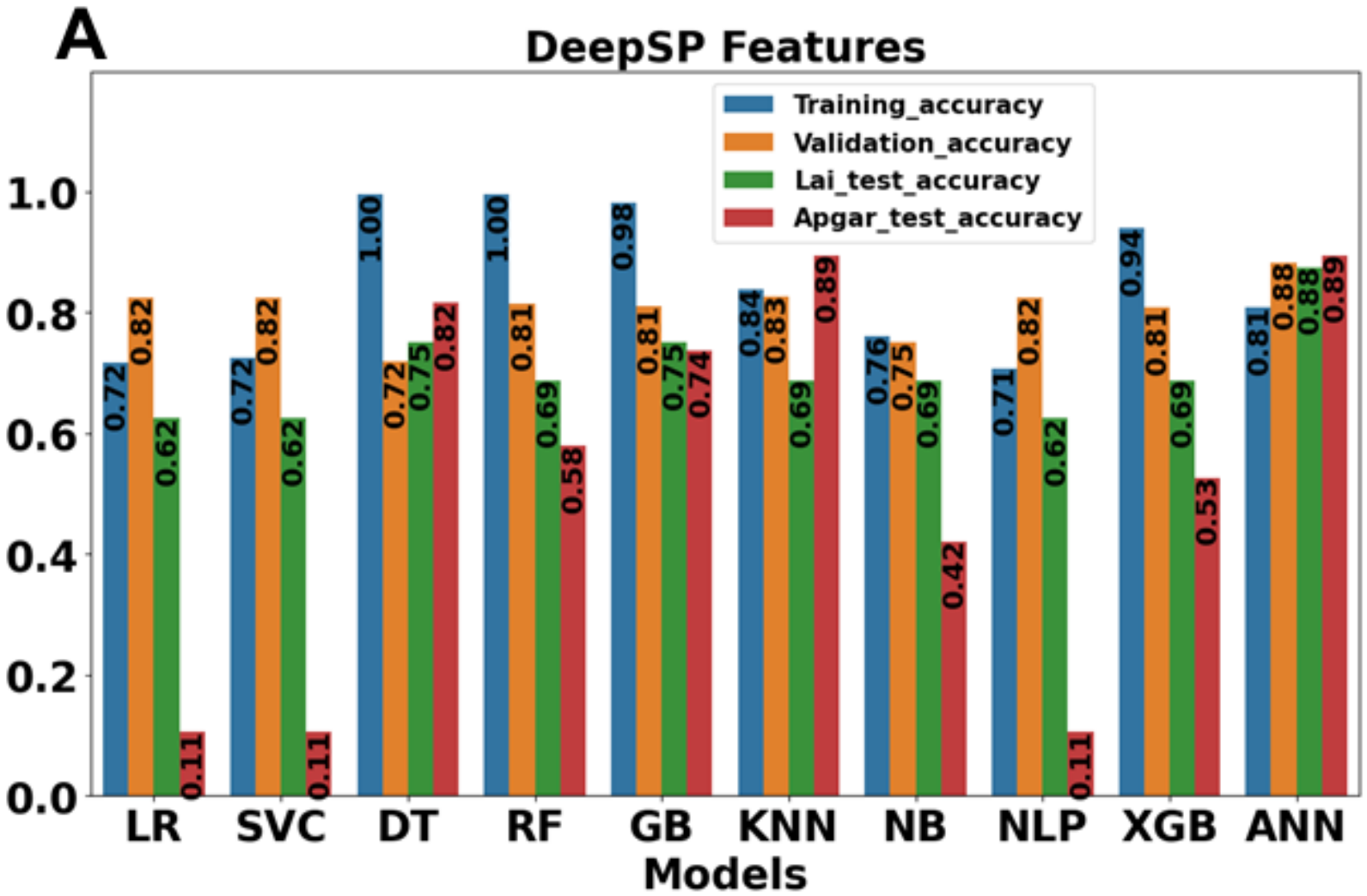
Machine learning models

Nine traditional machine learning models
- -- logistic regression (LR)
- -- support vector classifier (SVC)
- -- decision tree (DT)
- -- random forest (RF)
- -- gradient boosting (GB)
- -- k-nearest neighbors (KNN)
- -- naive Bayes (NB)
- -- multinomial naive Bayes (MNB)
- -- extreme gradient boosting (XGB)

One deep learning model
- -- Artificial neural network (ANN)

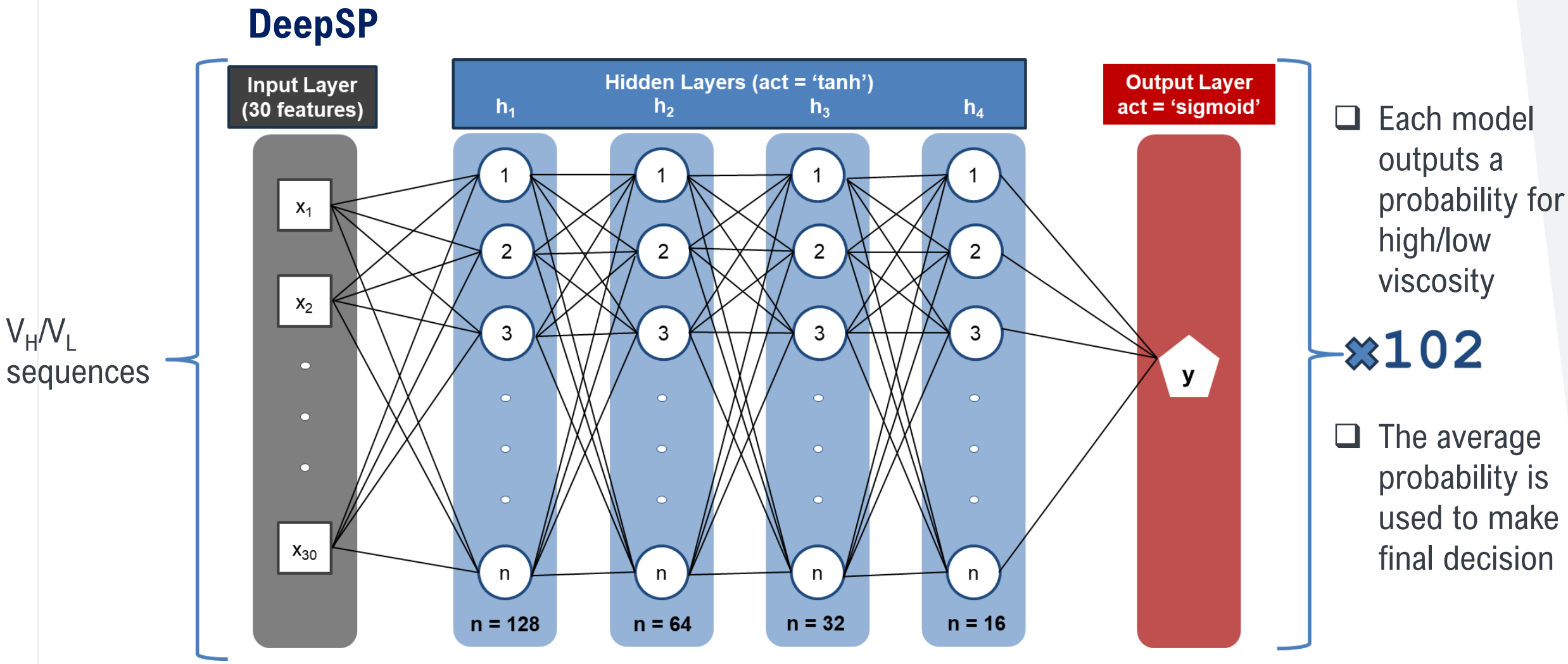
- ❑ The **DeepSP** features were used for all the models.
- ❑ The hyperparameters of each model were determined by the best **LOGO** score.

Machine learning results



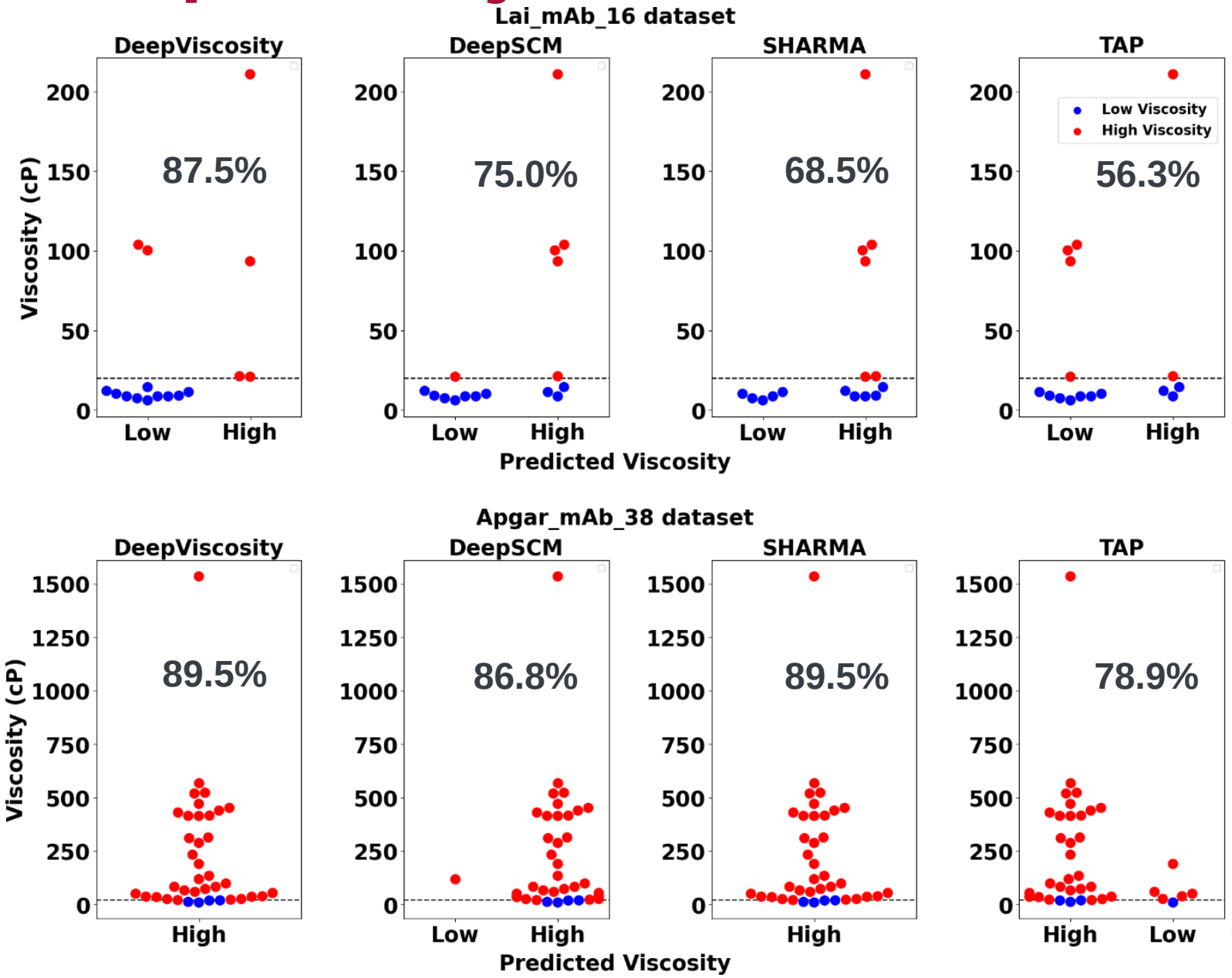
- The ANN model outperforms other models on the validation and two independent test sets

DeepViscosity (DeepSP + ensemble deep learning)



- Each model outputs a probability for high/low viscosity
- The average probability is used to make final decision

DeepViscosity vs other models



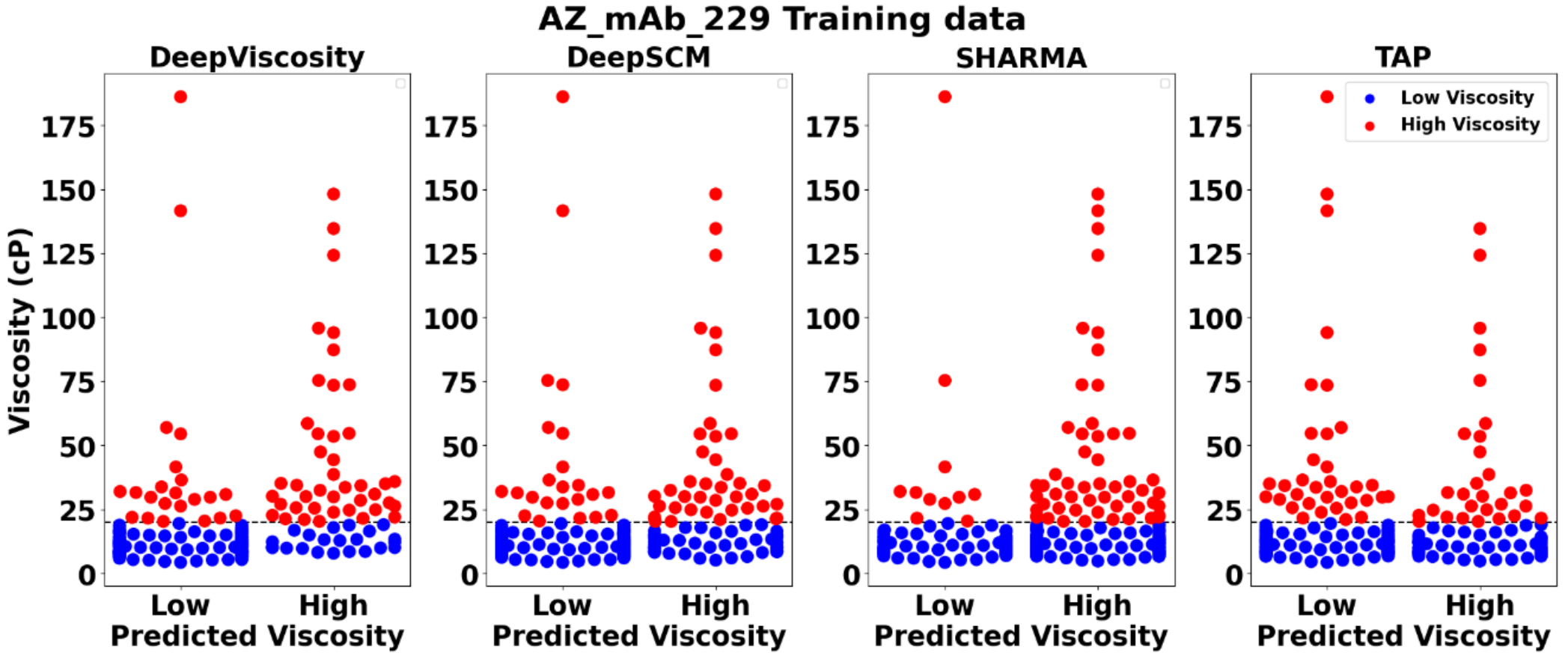
- The Apgar datasets have different formulation conditions
- Many recent ML models used either dataset for training, not suitable for comparison

Comput. Struct. Biotechnol. J. 2022 (20), 2113-2152.

PNAS 2014 (111), 18601–18606.

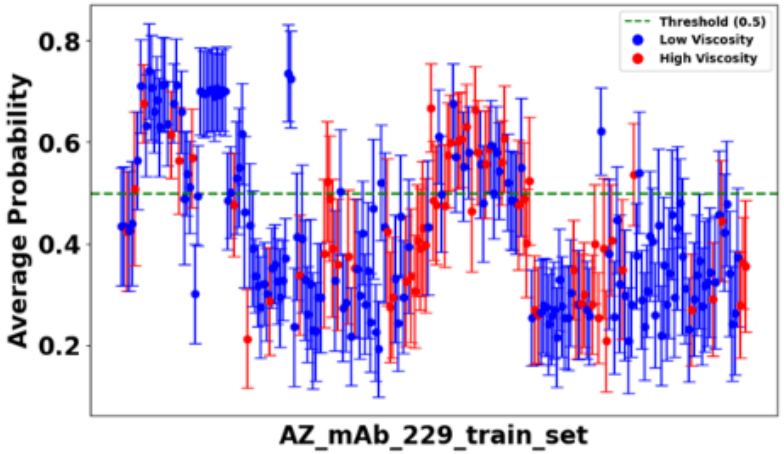
PNAS 2019 (116), 4025–4030.

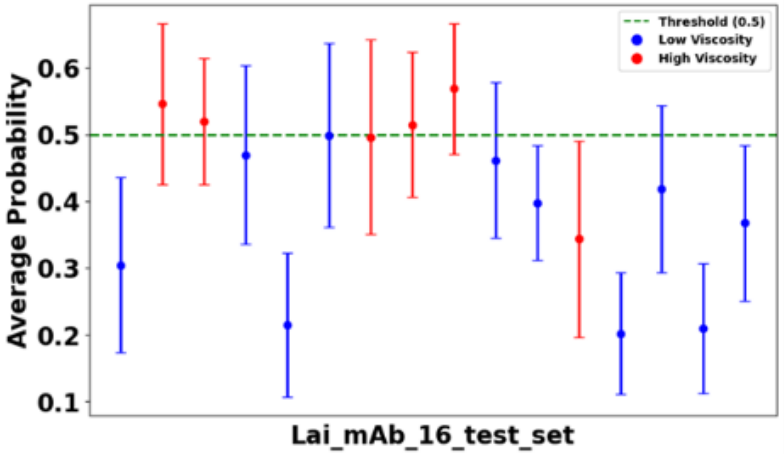
DeepViscosity on the 229 datasets

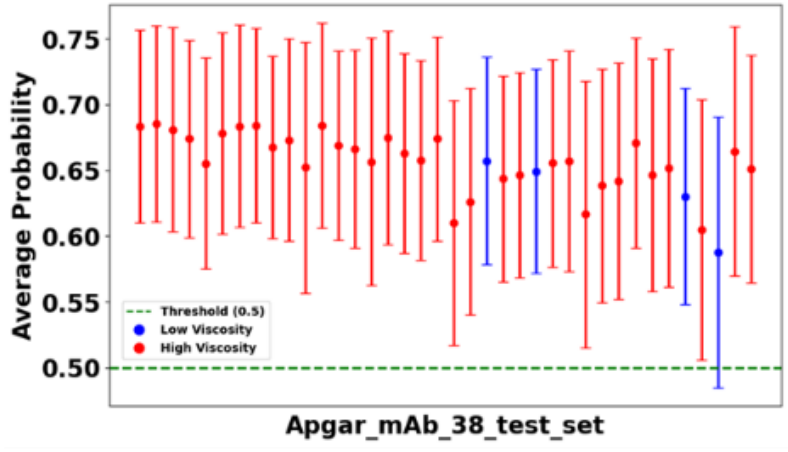


	DeepViscosity	DeepSCM	SHARMA	TAP
Accuracy	80.7%	68.1%	52.8%	62.9%

DeepViscosity ensemble averages

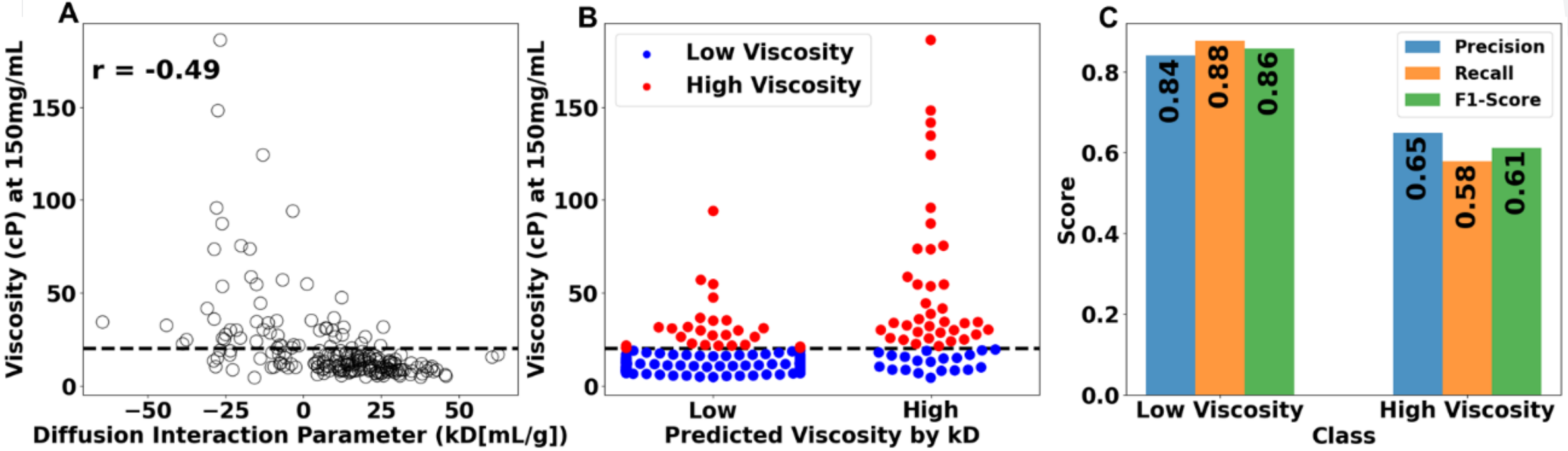






- Average and standard deviation of the probability from 102 ensemble models
- A probability of 0.5 is the threshold to make a prediction

kD as a predictor for viscosity

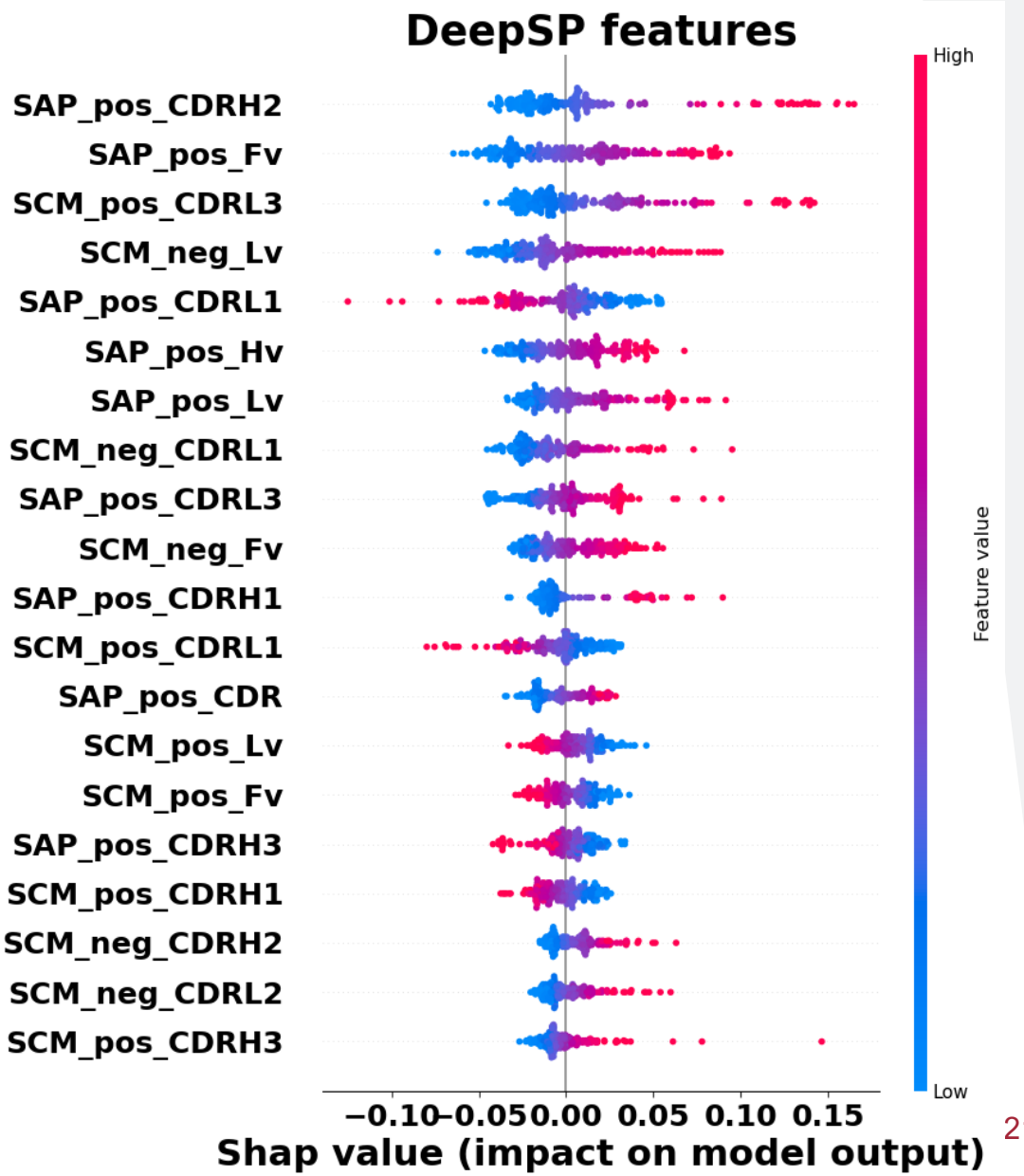


kD ≤ -10 mL/g ➔ high viscosity (> 20 cP)
kD > -10 mL/g ➔ low viscosity (≤ 20 cP)

Accuracy 79%

Shapley Additive exPlanations (SHAP) analysis for feature importance

- All the SCM_neg features have a **positive** correlation with high viscosity
- Most SAP features have a **positive** correlation with high viscosity except CDRL1 and CDRH3
- Most SCM_pos features have a **negative** correlation with high viscosity except CDRL3 and CDRH3



Conclusion

- A large collection of 229 high-concentration mAb viscosity was obtained.
- DeepViscosity (DeepSP + ensemble ANN) was developed with consistent accuracy on the validation and test sets (~88%).
- DeepViscosity depends only on the Fv sequences. The model will be released after publication.

Acknowledgement

Lai Lab

- **Pin-Kuang Lai**
- **Lateefat Kalejaye**
- **I-En Wu**
- Taylor Terry
- Malcolm Harrison
- Manavi Panjnani
- Atharva Shaligram
- Raj Rana
- Lauren Espineli
- Ibrahim Wichka
- Miles Cabreza
- Kyle Savino

AstraZeneca

- **Neil Mody**
- **Jenna Caldwell**
- **MLab**

Computing resources

STEVENS
INSTITUTE OF TECHNOLOGY
1870

THANK YOU

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030