Using AI to Determine Crystal Structure

NIST Center for Neutron Research

Elizabeth Baggett

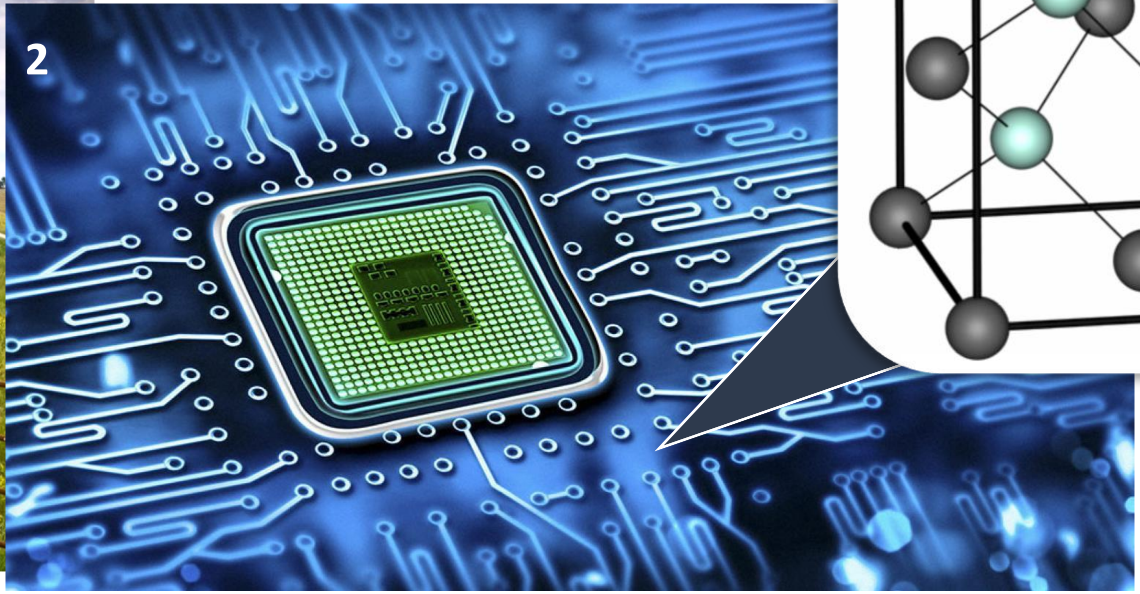
Mentor: Dr. William Ratcliff

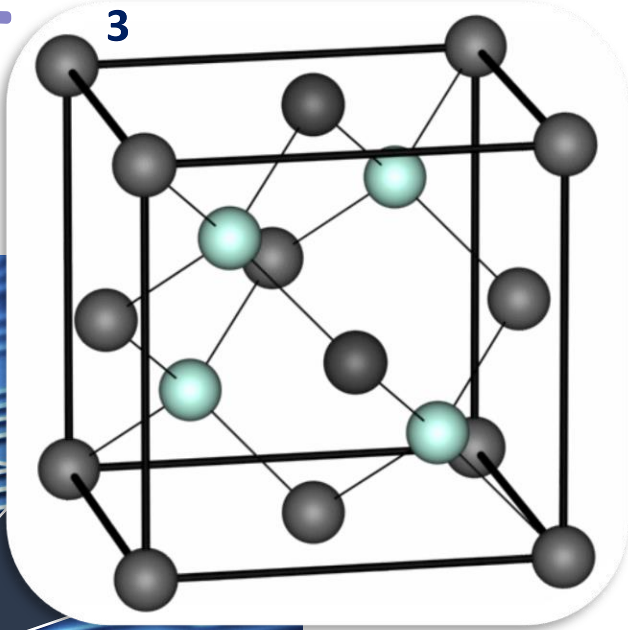
Boston College, 2026

Collaborators

- William Ratcliff (NIST Center for Neutron Research)
- Edward Freidman (Wheaton High School, NIST Center for Neutron Research)
- Paul Kienzle (NIST Center for Neutron Research)
- Nebil Ayape Katcho (Institute Laue-Langevin)

Motivation





1. Unknown. (n.d.). *"two-faced" solar cells generate a lot more power*. "Two-faced" solar cells generate a lot more power. https://greatfactsnow.blogspot.com/2019/12/two-faced-solar-cells-generate-lot-more.html
2. *Experimental morpheus CPU is "mind-bogglingly terrible" to crack*. Network World. (2021, June 4). https://www.networkworld.com/article/969940/experimental-morpheus-cpu-is-mind-bogglingly-terrible-to-crack.html
3. Weerasinghe, Hasitha. Electrical characterization of metal-to-insulator transition in iron silicide thin films on sillicone substrates

Translational Symmetry

Image from Burlew, R. (2016, October 10). *Uncapped honey fermenting in the Comb*. Honey Bee Suite. https://www.honeybeesuite.com/uncapped-honey-fermenting-in-the-comb/

Translation Symmetry in Crystals

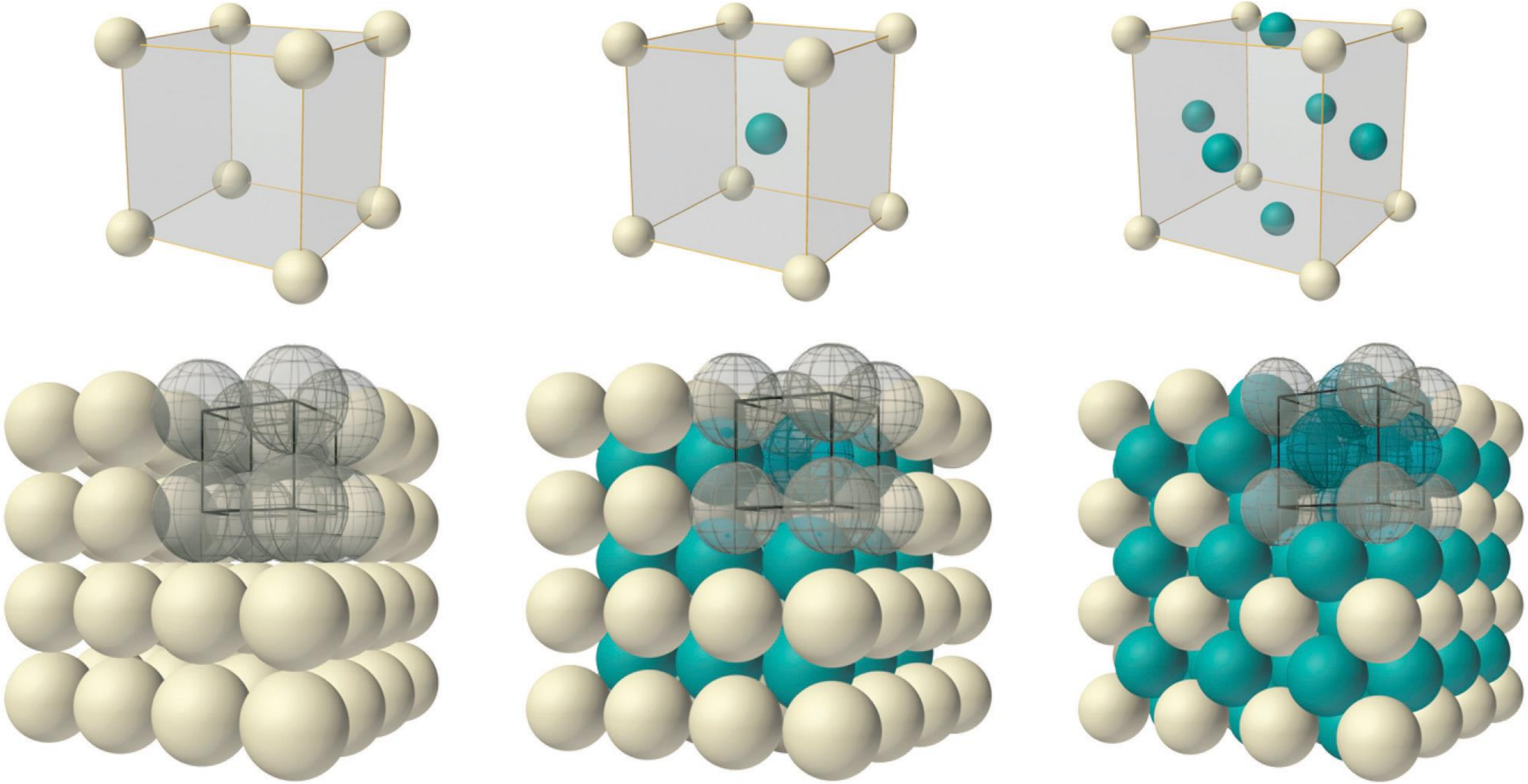


Image from Libretexts. (2021, August 27). *Chapter 12.2: Arrangement Of Atoms in Crystals*. Chemistry LibreTexts.
https://chem.libretexts.org/Courses/Howard_University/General_Chemistry:_An_Atoms_First_Approach/Unit_5:_States_of_Matter/Chapter_12:_Solids/Chapter_12.02:_Arrangement_of_Atoms_in_Crystals

14 Bravais lattices

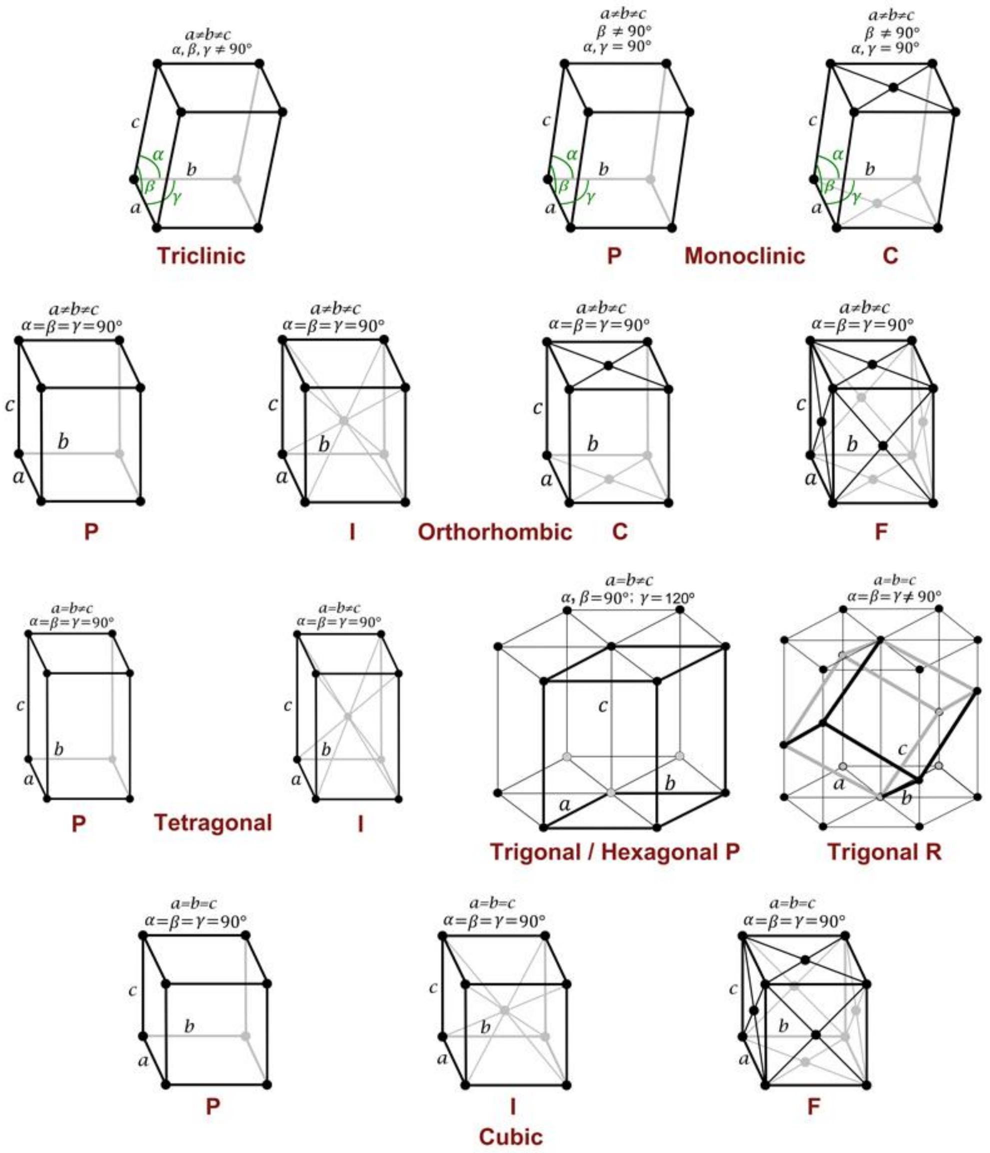


Image from Ripoll, M. M. (n.d.). Cristalografía. Simetría de los Cristales. Representación de las redes de bravais. https://www.xtal.iqfr.csic.es/Cristalografia/parte_03_4.html

Internal Symmetries

Image from Adobe Stock

230 Space Groups

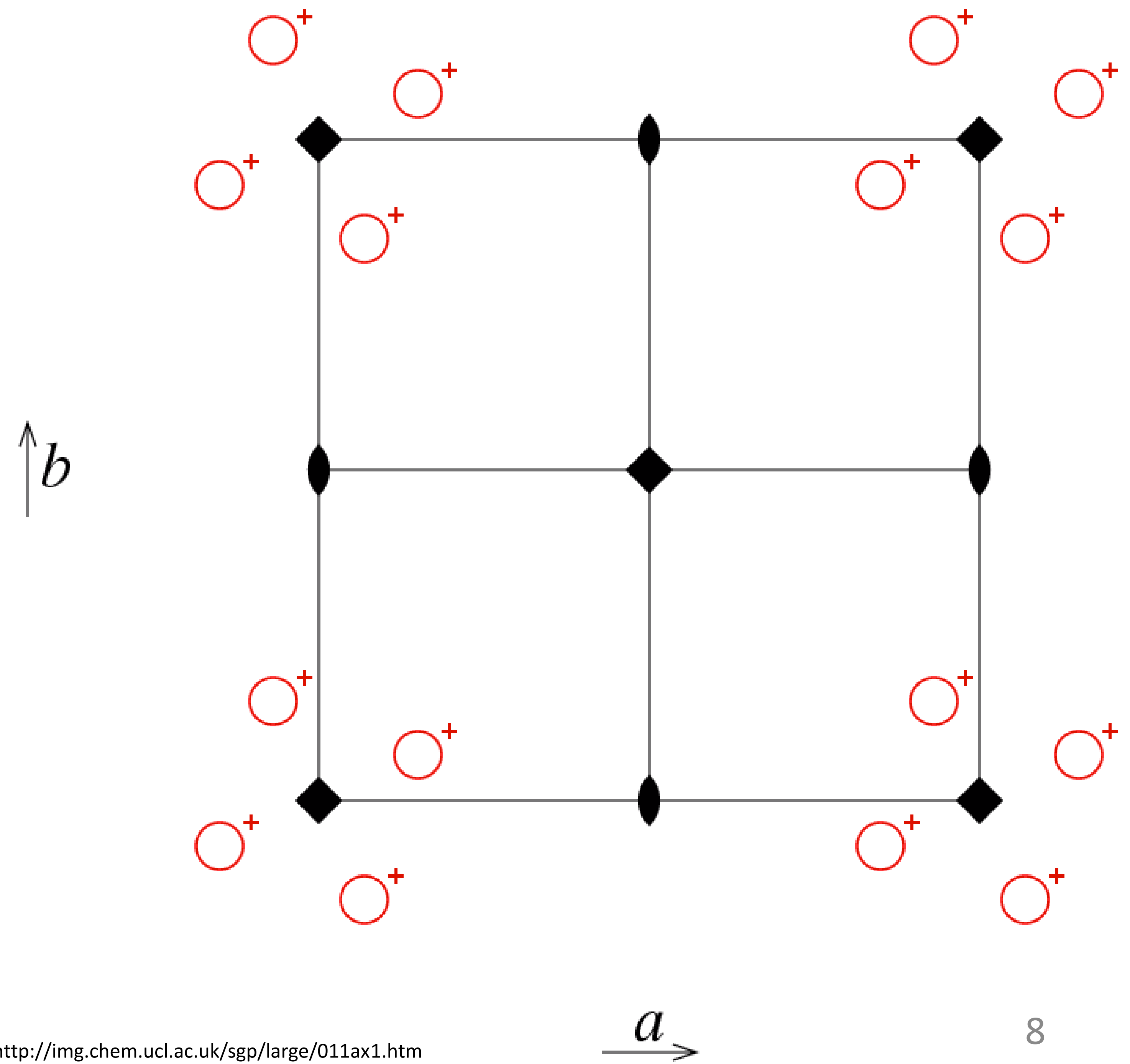
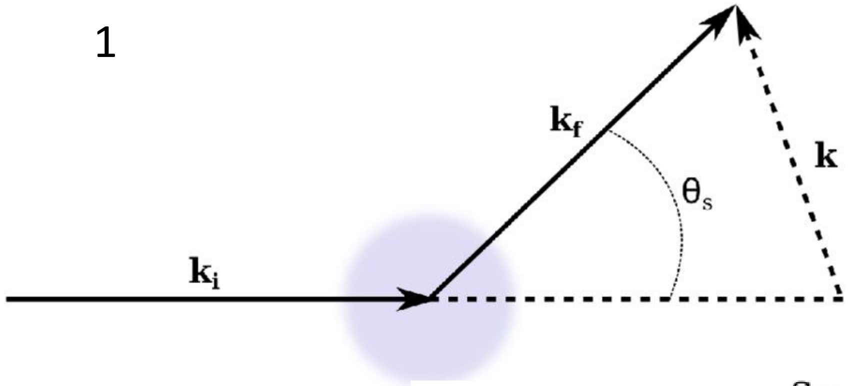


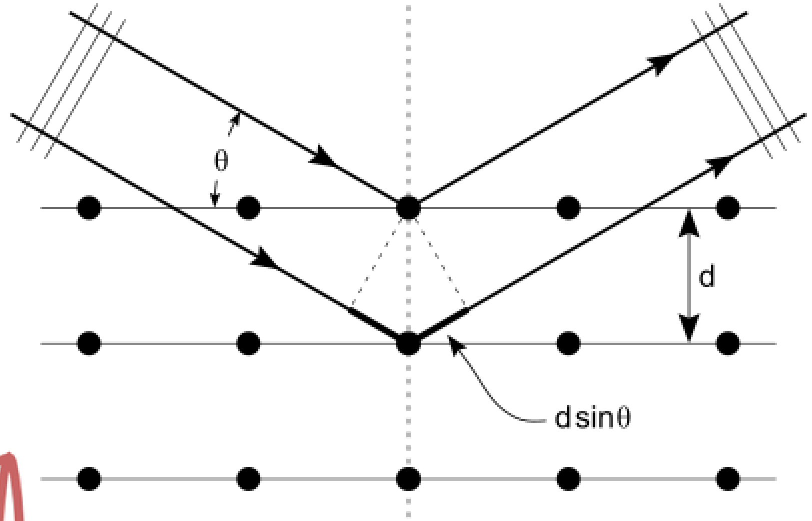
Image from Cockcroft, J. K. (n.d.). Space group 11: P2(1)/M11; p21/M11; P 21/m 1 1. http://img.chem.ucl.ac.uk/sgp/large/011ax1.htm

Experimentally Determining Crystal Structure

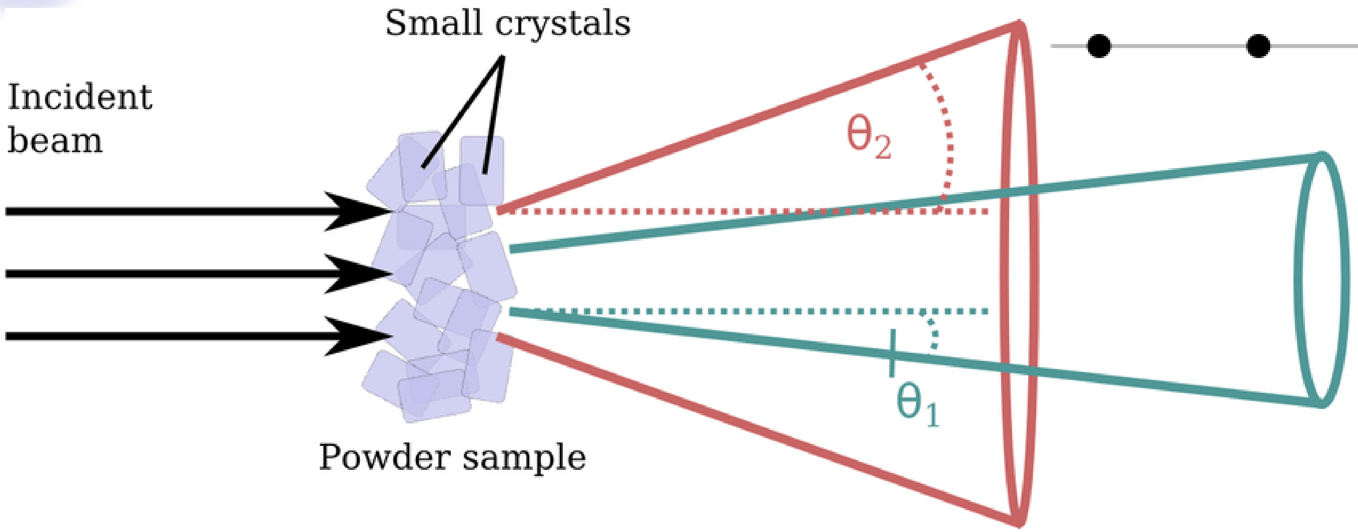
1



2



3



1, 3 Glenzer, S. H., & Redmer, R. (2009, December 1). *X-ray thomson scattering in high energy density plasmas*. Reviews of Modern Physics. https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.81.1625
2 Wikimedia Foundation. (2024, July 15). *Bragg's law*. Wikipedia. https://en.wikipedia.org/wiki/Bragg%27s_law

Powder Diffraction Pattern

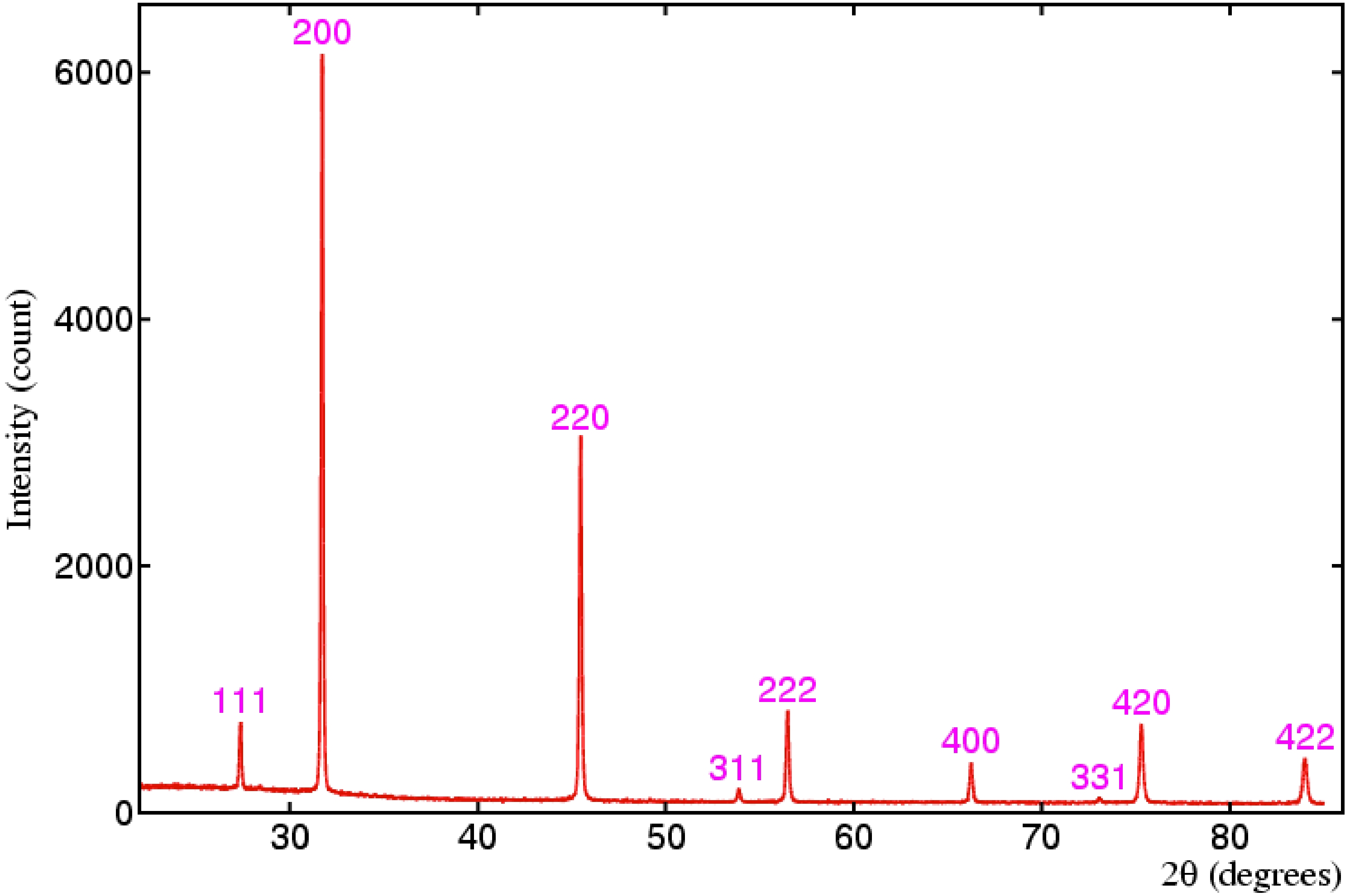
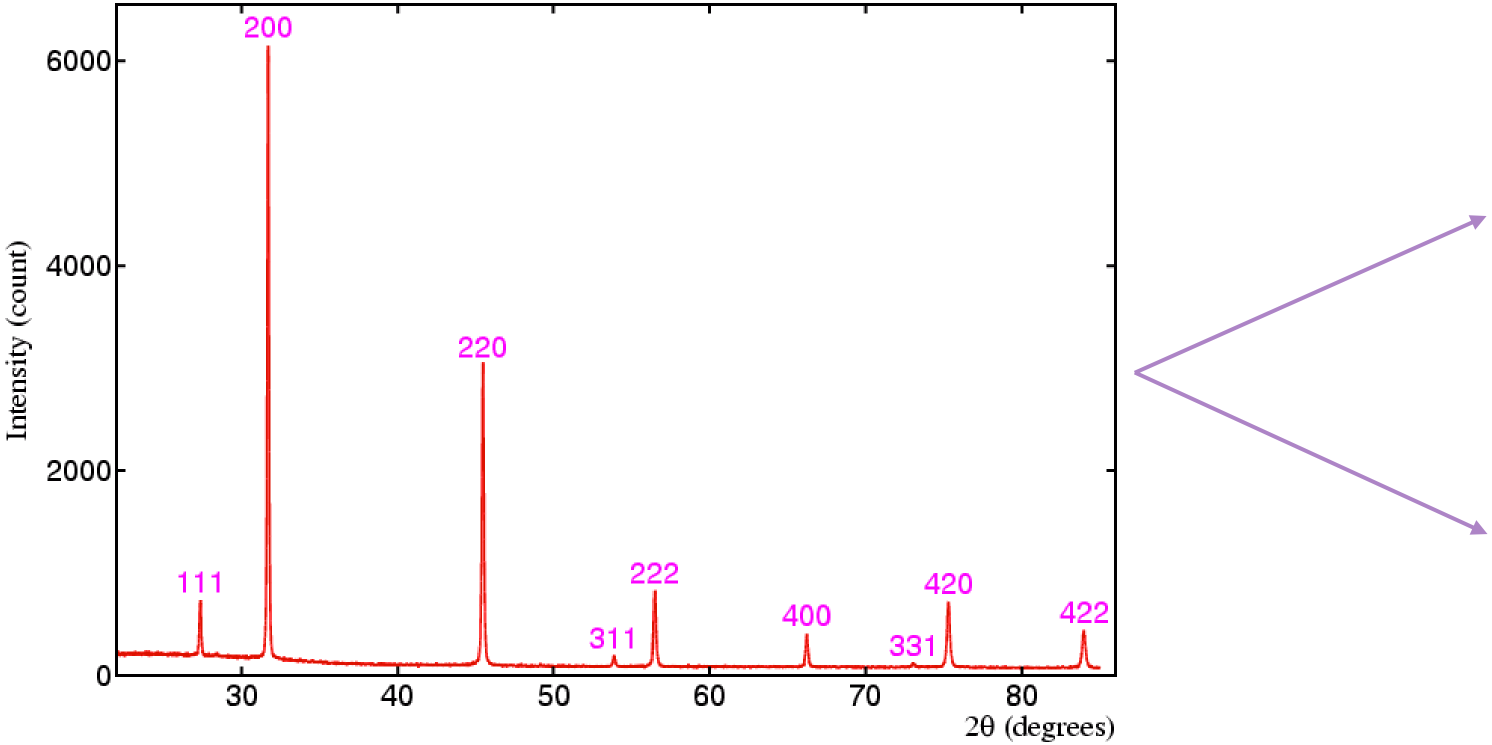


Image from NaCl data for Diffraction II. (n.d.). http://pd.chem.ucl.ac.uk/pdnn/diff2/nacl.htm

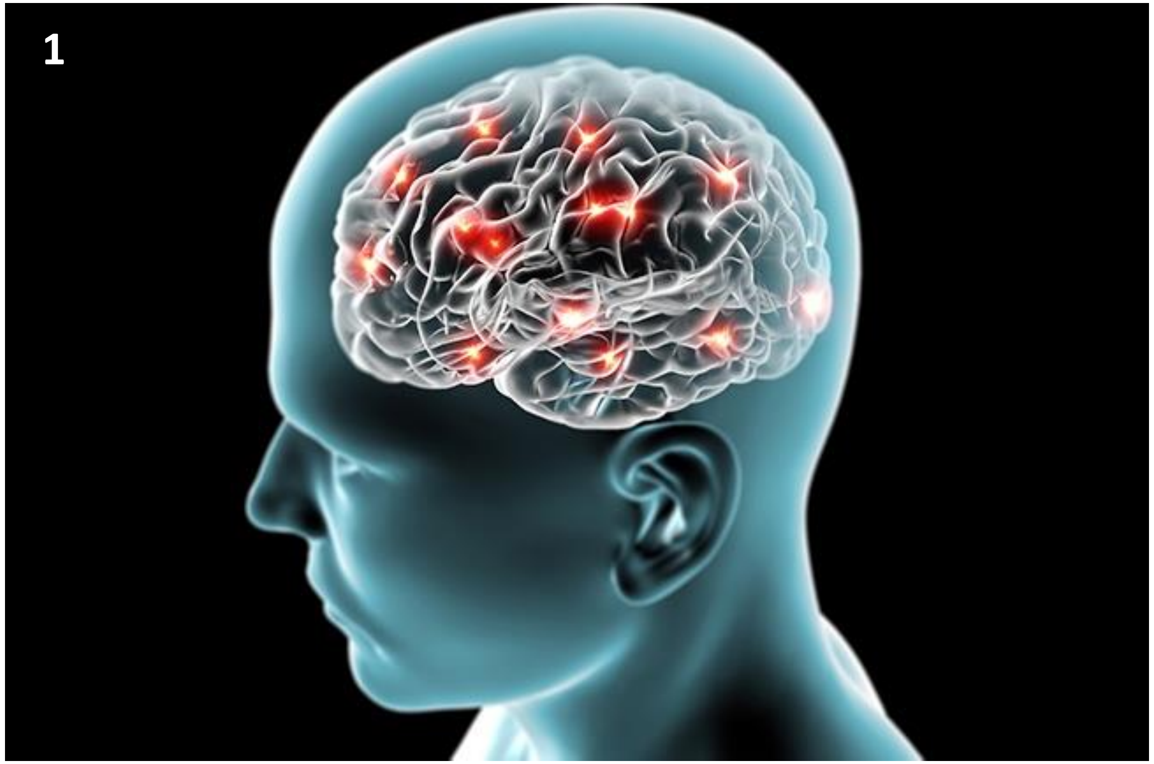
AI Approach

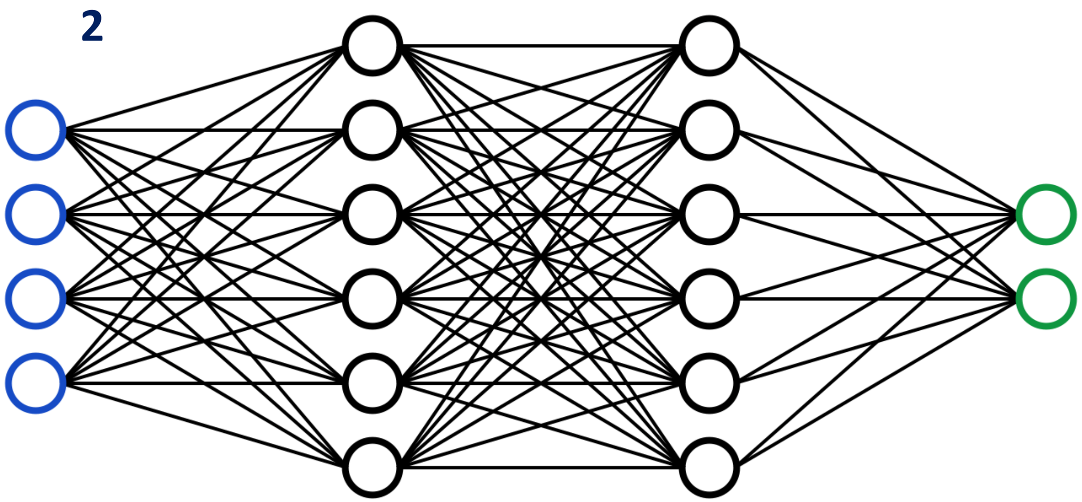


Bravais Lattice

Space group

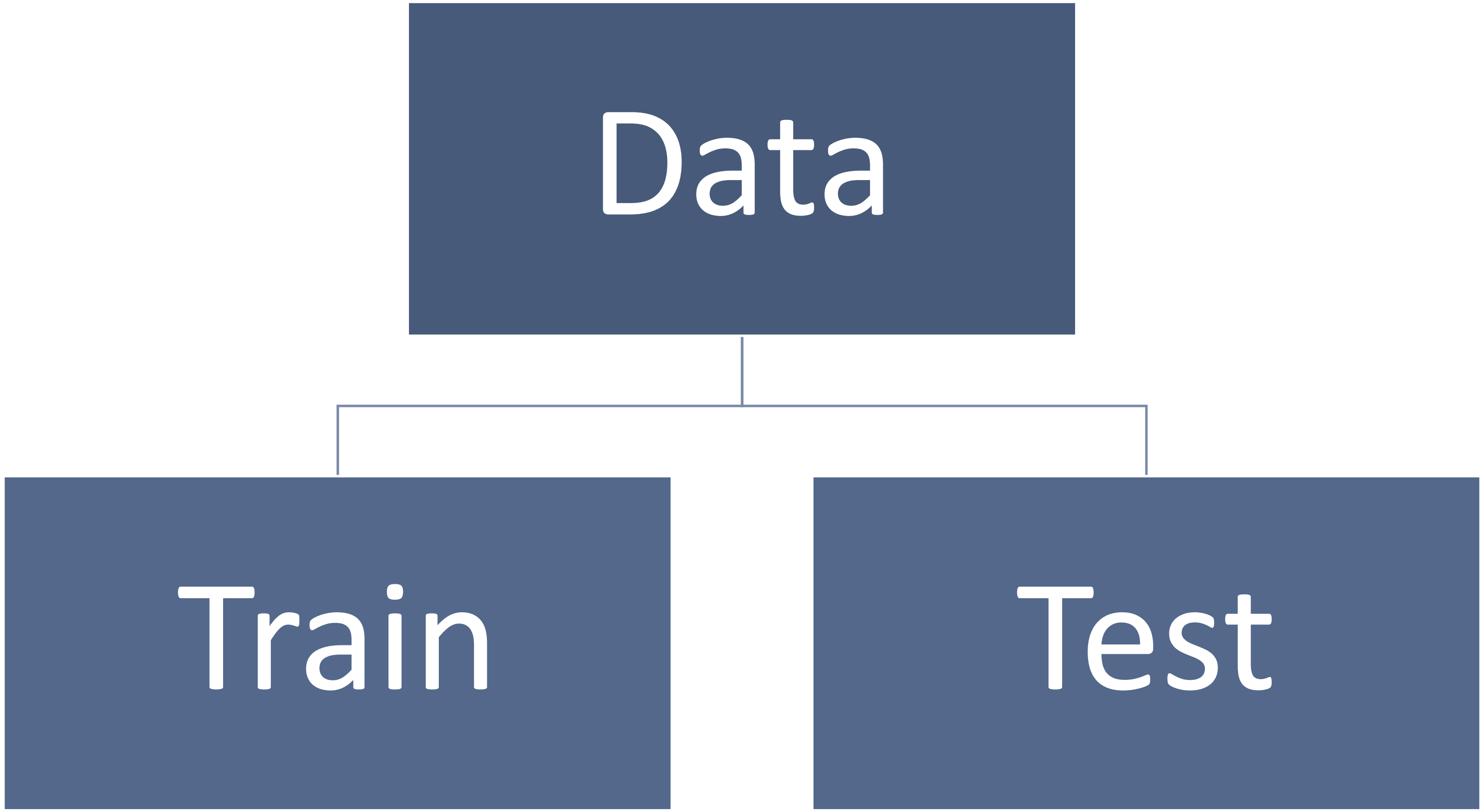
Neural Network



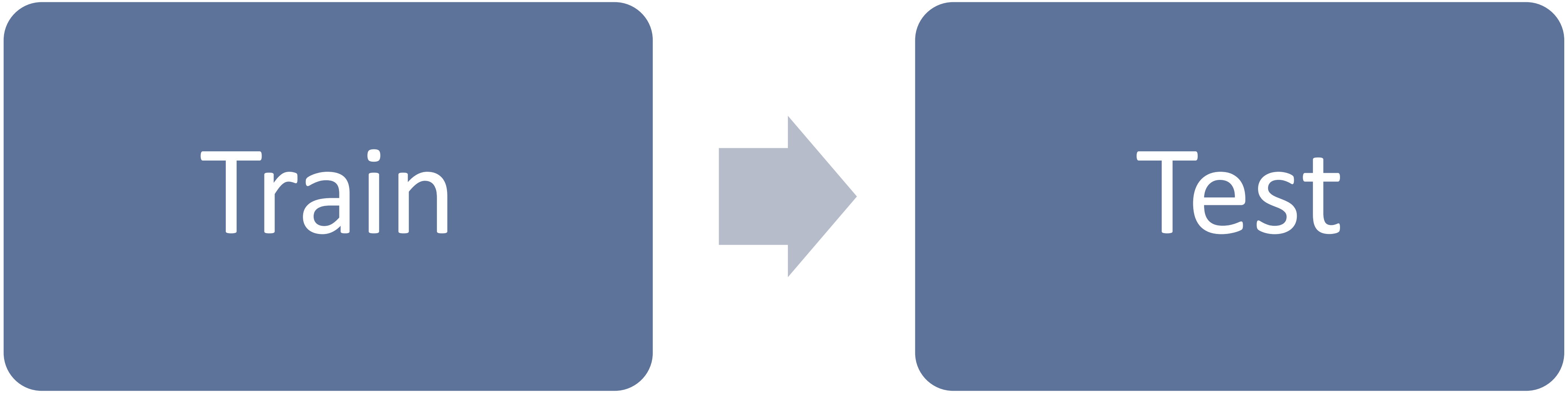


1. Peshek, S. (2019, April 30). *How do synapses work?*. Texas A&M Today. https://today.tamu.edu/2018/01/05/how-do-synapses-work/
2. Victor Zhou. (n.d.). *Neural Networks from scratch*. https://victorzhou.com/series/neural-networks-from-scratch/

Random Split



One Epoch (Training Cycle)



Changes the influence of nodes

Benchmark for generalization

Supervised Method

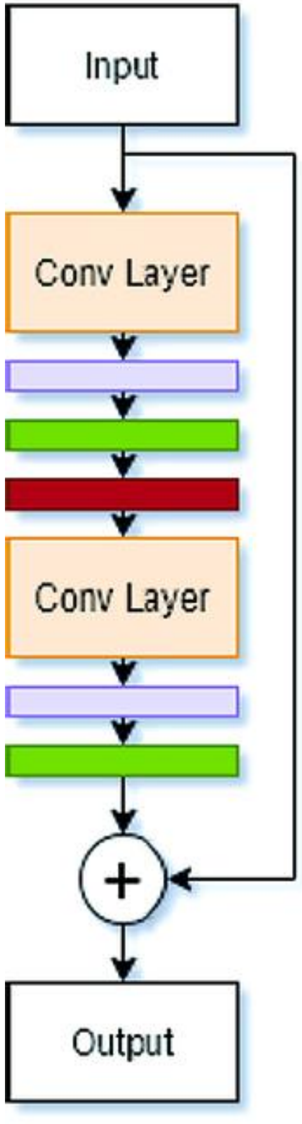
1. Selina. (2024, June 2). *Pufferfish facts*. Facts.net. https://facts.net/pufferfish-facts/
2. THOUGHTCO. (n.d.). https://www.thoughtco.com/thmb/soY69iSmFeIFV4S4yGWcr-vCkcw=/398x271/filters:fill(auto,1)/venn2-56a4b8b03df78cf77283f15f.JPG

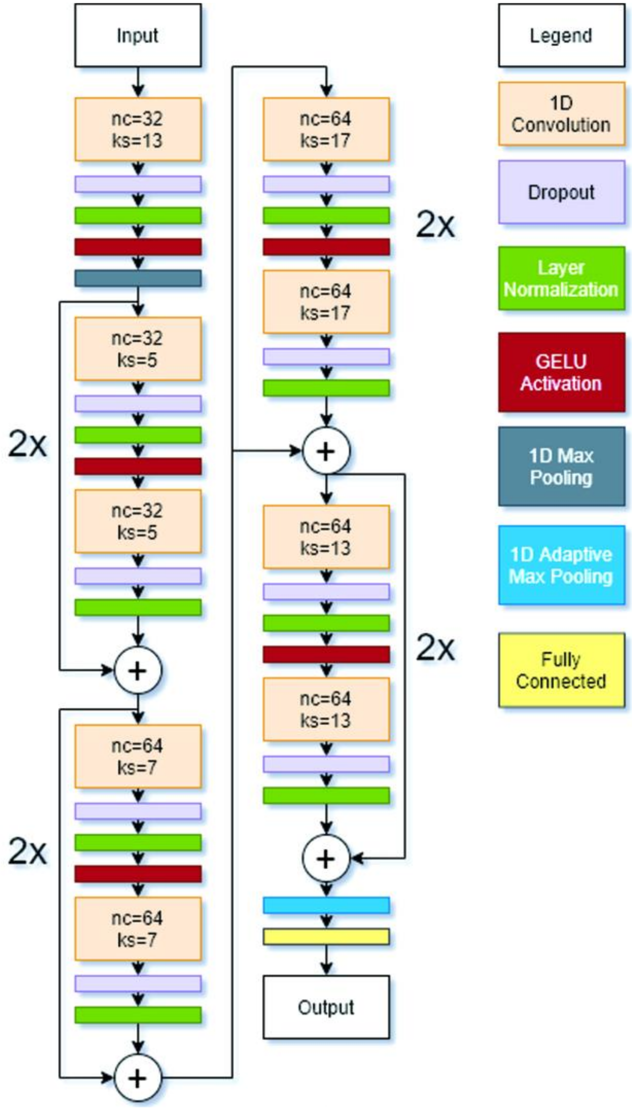
Supervised Method

1. *Whole catfish, fresh*. Goldfish Seafood Market. (n.d.-b). https://goldfishseafoodmarket.com/products/whole-catfish-fresh
2. Sharpe, S. (2024, June 20). *Learn everything you need to know about Betta Fish*. The Spruce Pets. https://www.thesprucepets.com/siamese-fighting-fish-bettas-1378308

Model

- Convolution neural network
- ResNet architecture
    - Haotong Liang





Images from Lolla, S., Liang, H., Kusne, A. G., Takeuchi, I., & Ratcliff, W. (2022). A semi-supervised deep-learning approach for automatic crystal structure classification. *Journal of Applied Crystallography*, *55*(4), 882–889. https://doi.org/10.1107/s1600576722006069

Previous Work

Satvik Lolla used Inorganic Crystal Structure Database data to simulate X-ray diffraction patterns

	Space Group	Bravais Lattice
Semi-supervised	78	85
Supervised	74	**88**

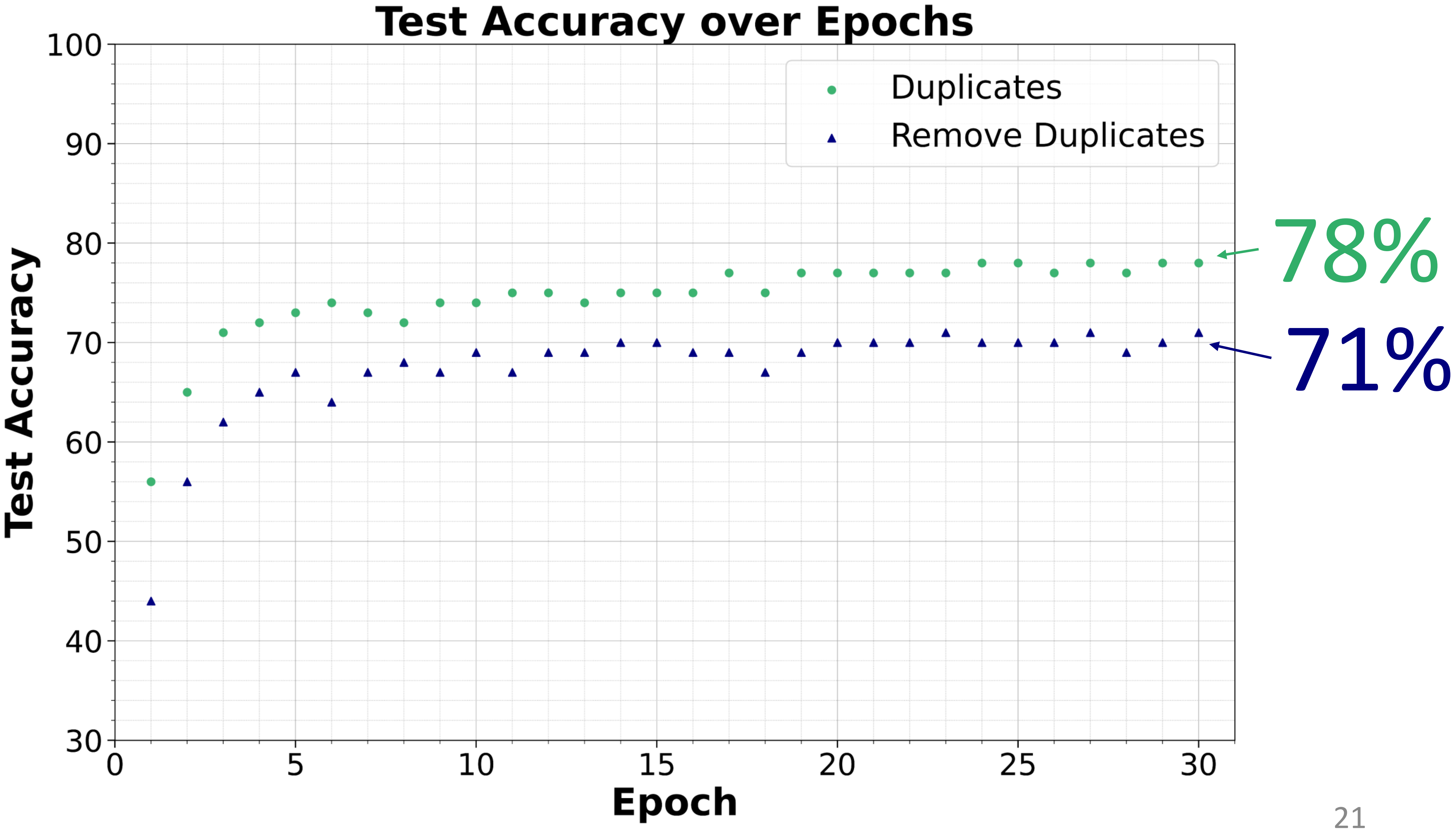
Lolla, S., Liang, H., Kusne, A. G., Takeuchi, I., & Ratcliff, W. (2022). A semi-supervised deep-learning approach for automatic crystal structure classification. *Journal of Applied Crystallography*, *55*(4), 882–889. https://doi.org/10.1107/s1600576722006069

Data Cleaning

- ICSD contains duplicates
- Neutron diffraction patterns
  - Lower resolution

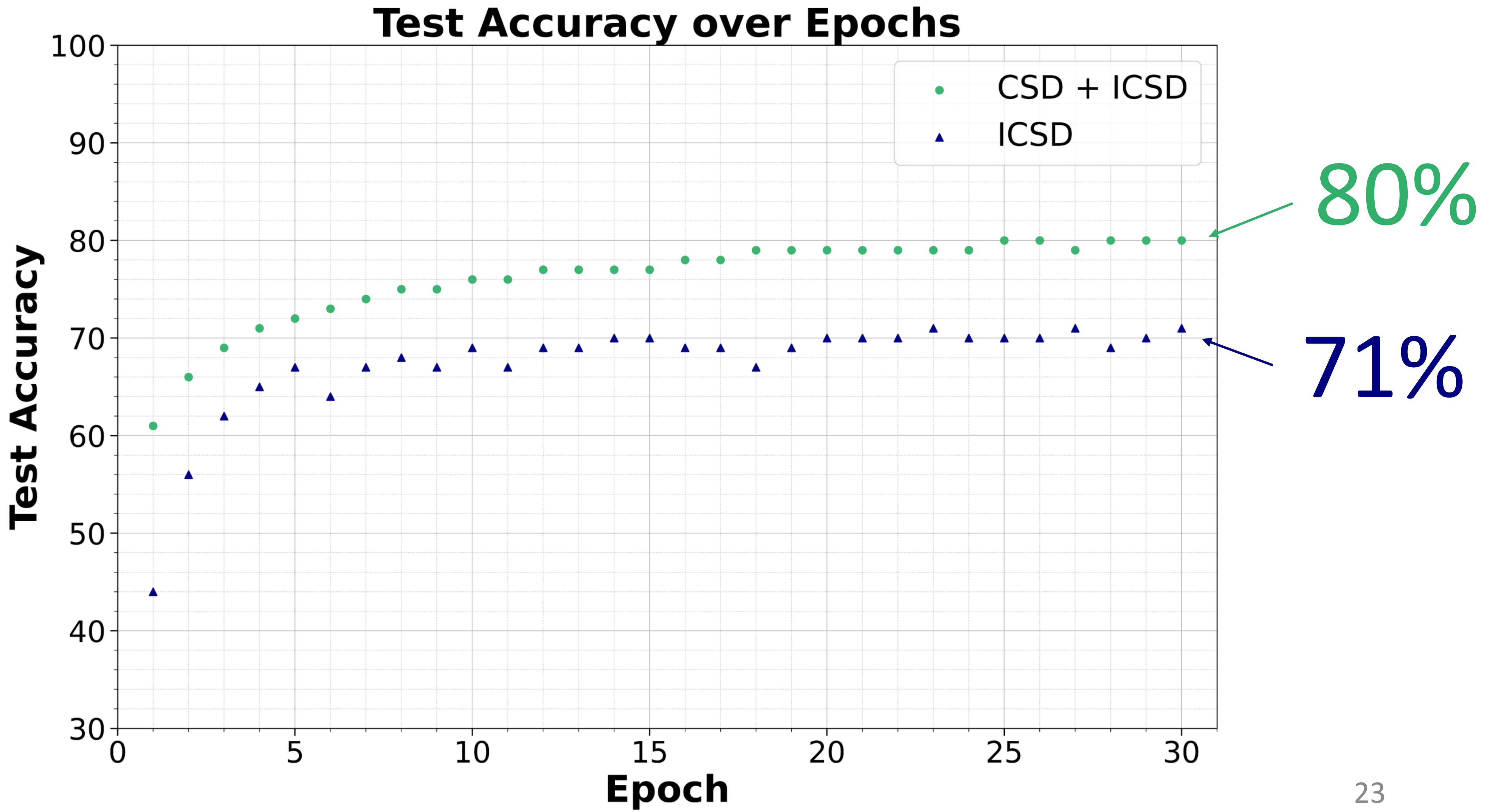
Remove Duplicates

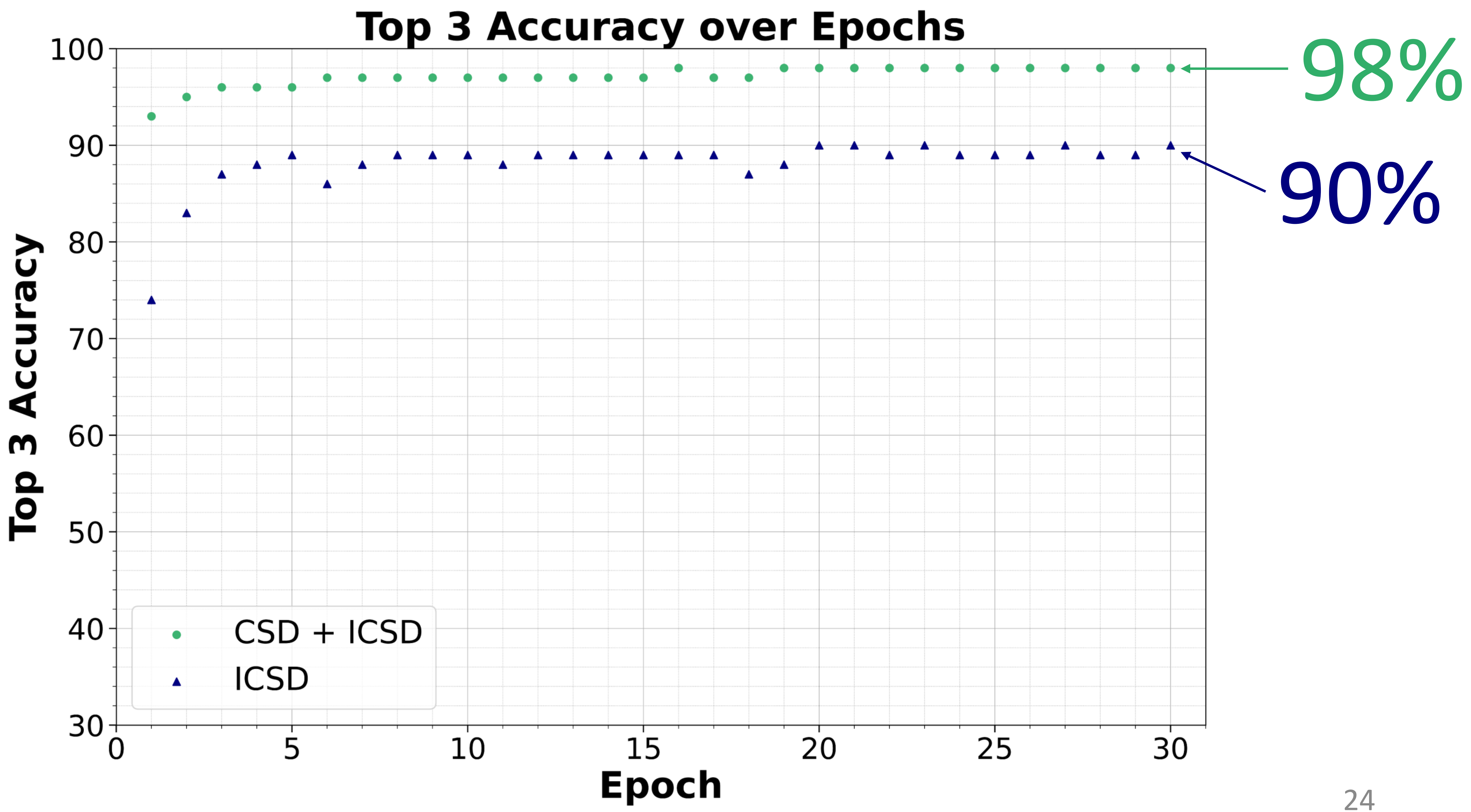
181,362 → 123,039


Test Accuracy over Epochs
Duplicates
Remove Duplicates
Test Accuracy
Epoch
100
90
80
70
60
50
40
30
0
5
10
15
20
25
30
78%
71%

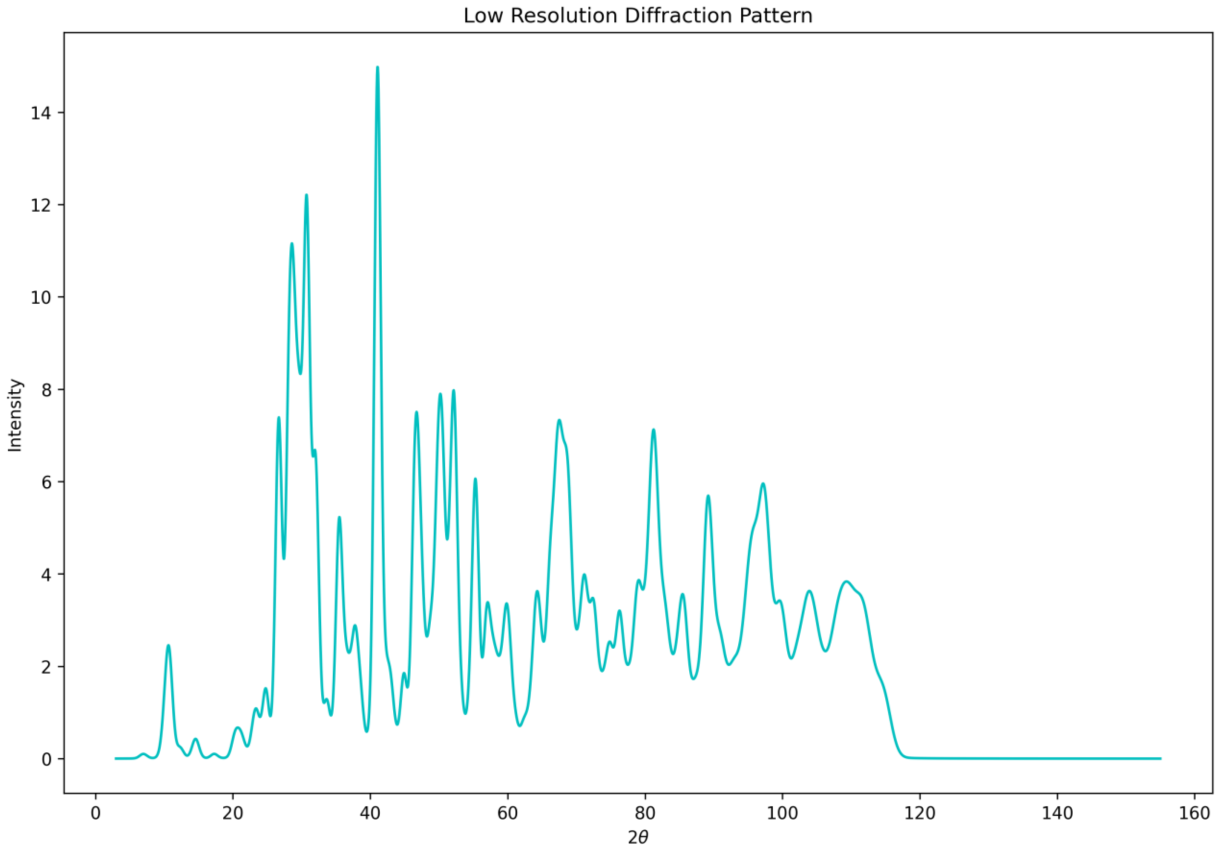
Adding More Data

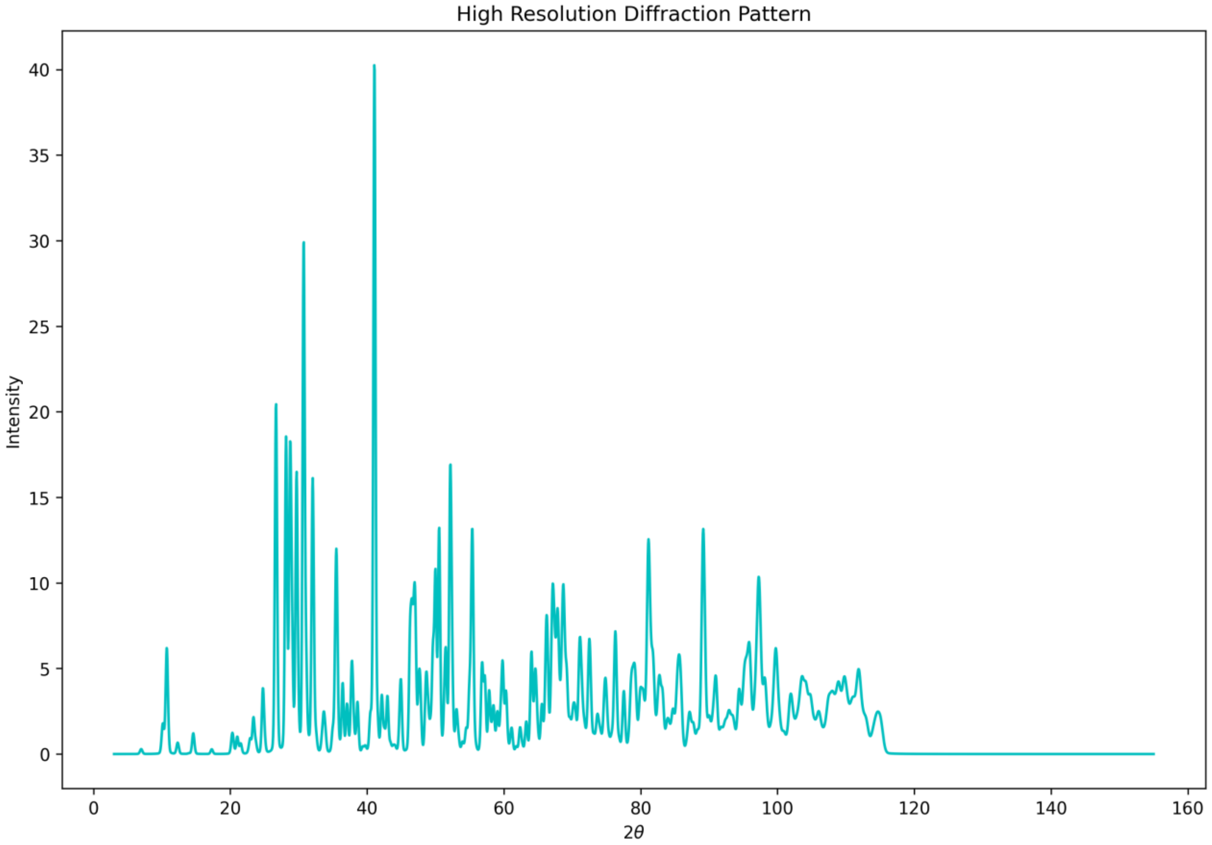
Adding data from Cambridge Structural Database (~ 1 million)

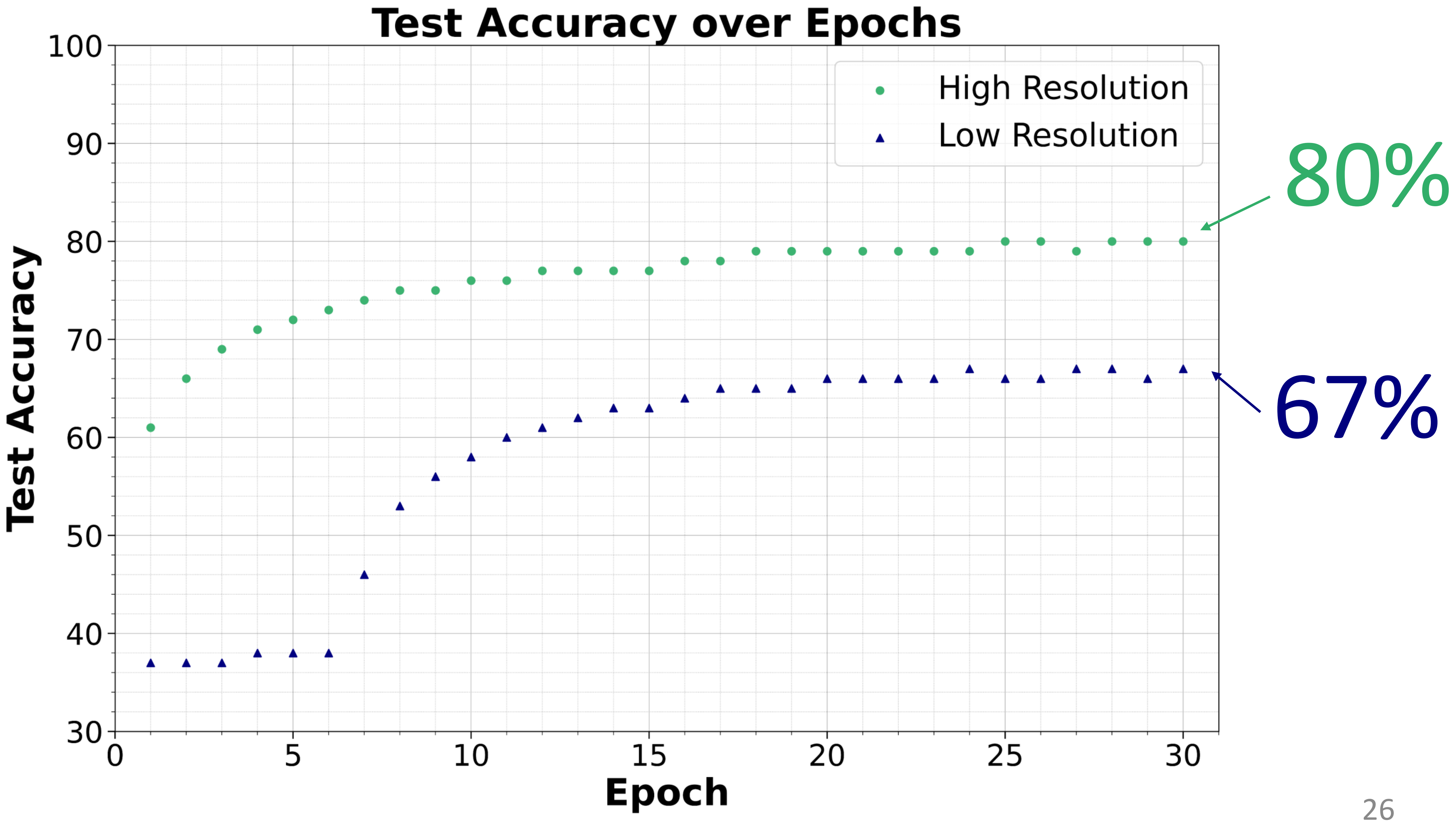

Test Accuracy over Epochs
CSD + ICSD
ICSD
Test Accuracy
Epoch
100
90
80
70
60
50
40
30
0
5
10
15
20
25
30
80%
71%


Top 3 Accuracy over Epochs
Top 3 Accuracy
Epoch
CSD + ICSD
ICSD
98%
90%
100
90
80
70
60
50
40
30
0
5
10
15
20
25
30

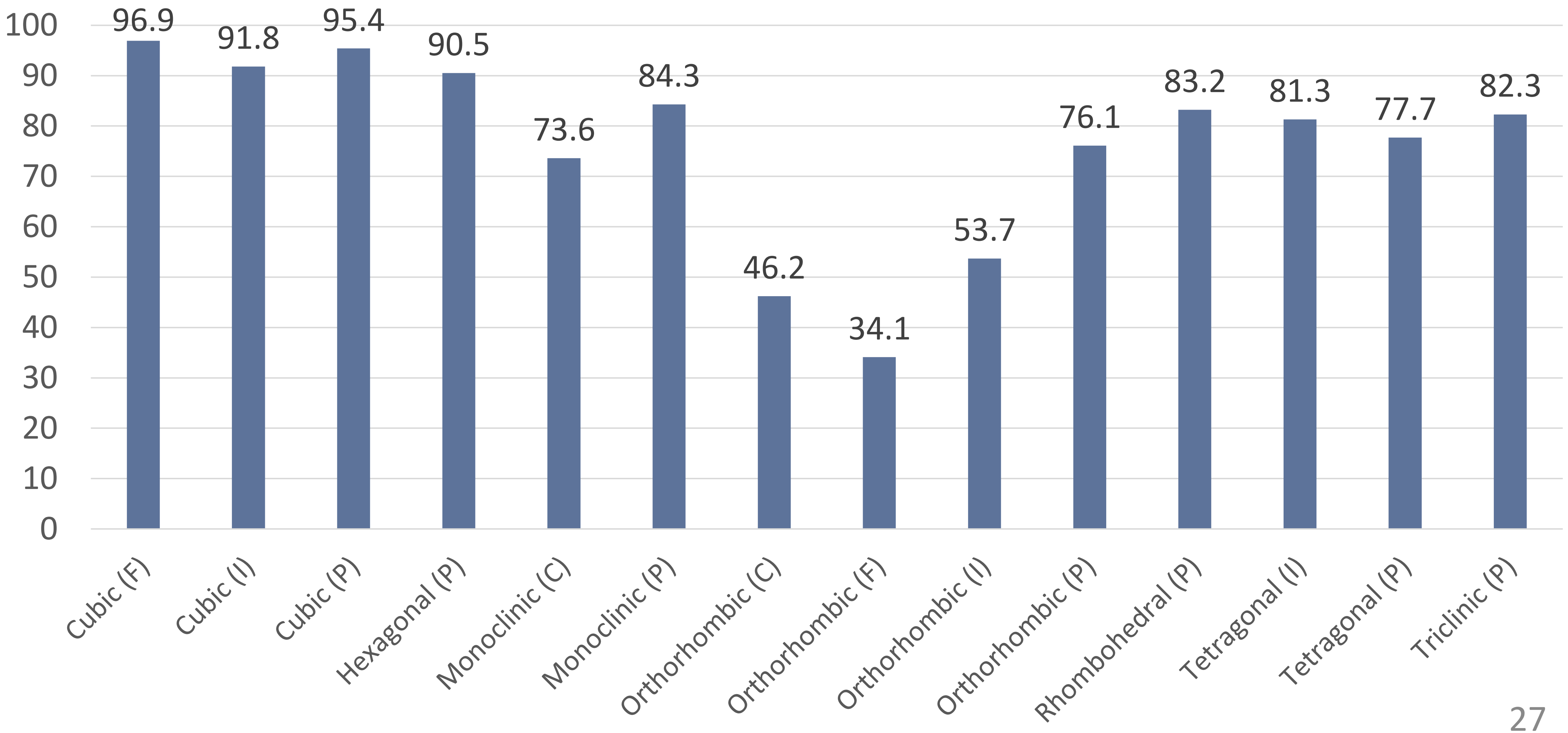
Resolution


Low Resolution Diffraction Pattern
Intensity
2θ
0
2
4
6
8
10
12
14
0
20
40
60
80
100
120
140
160

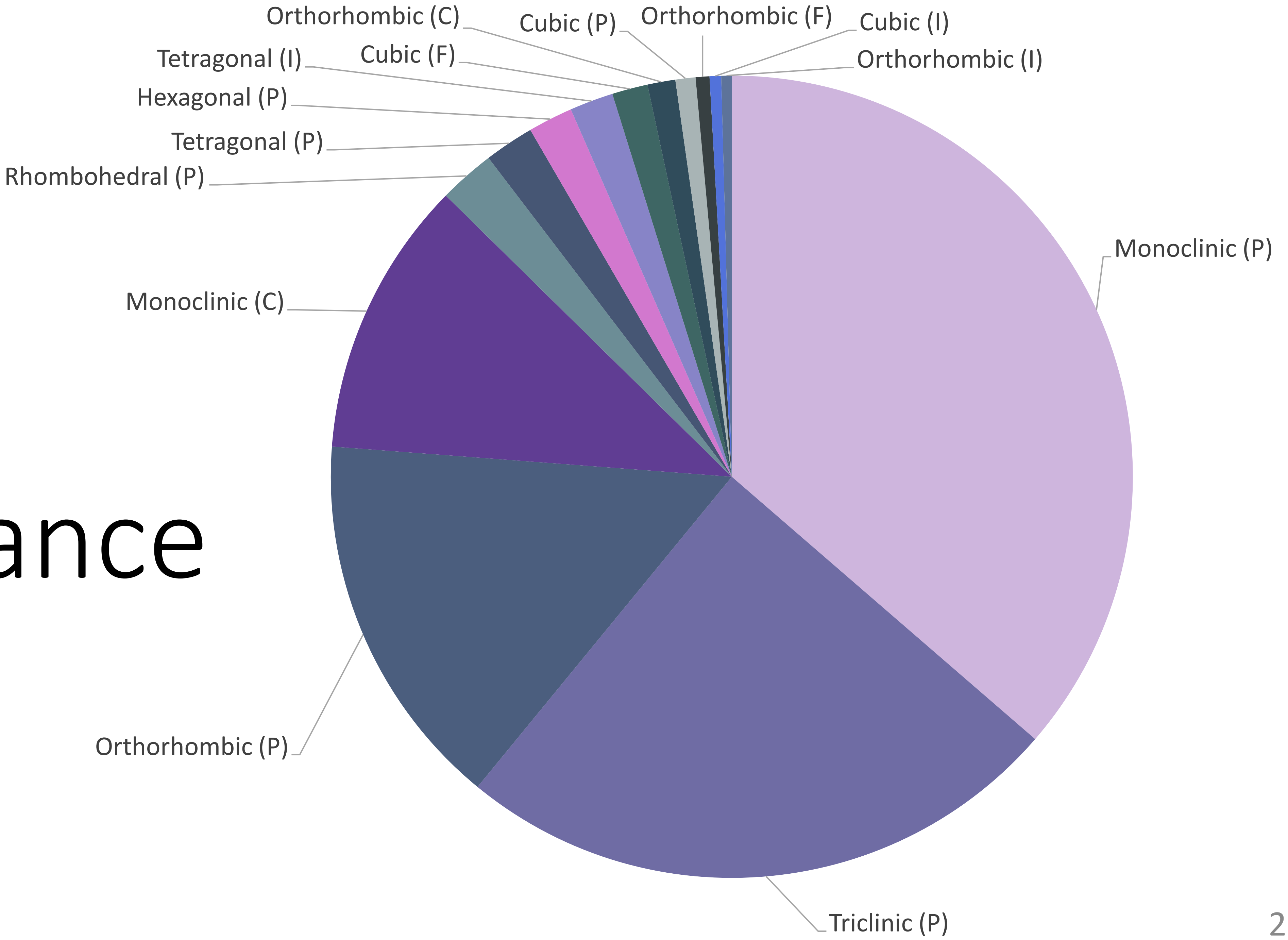

High Resolution Diffraction Pattern
Intensity
2θ
0
5
10
15
20
25
30
35
40
0
20
40
60
80
100
120
140
160


Test Accuracy over Epochs
High Resolution
Low Resolution
Test Accuracy
Epoch
100
90
80
70
60
50
40
30
0
5
10
15
20
25
30
80%
67%

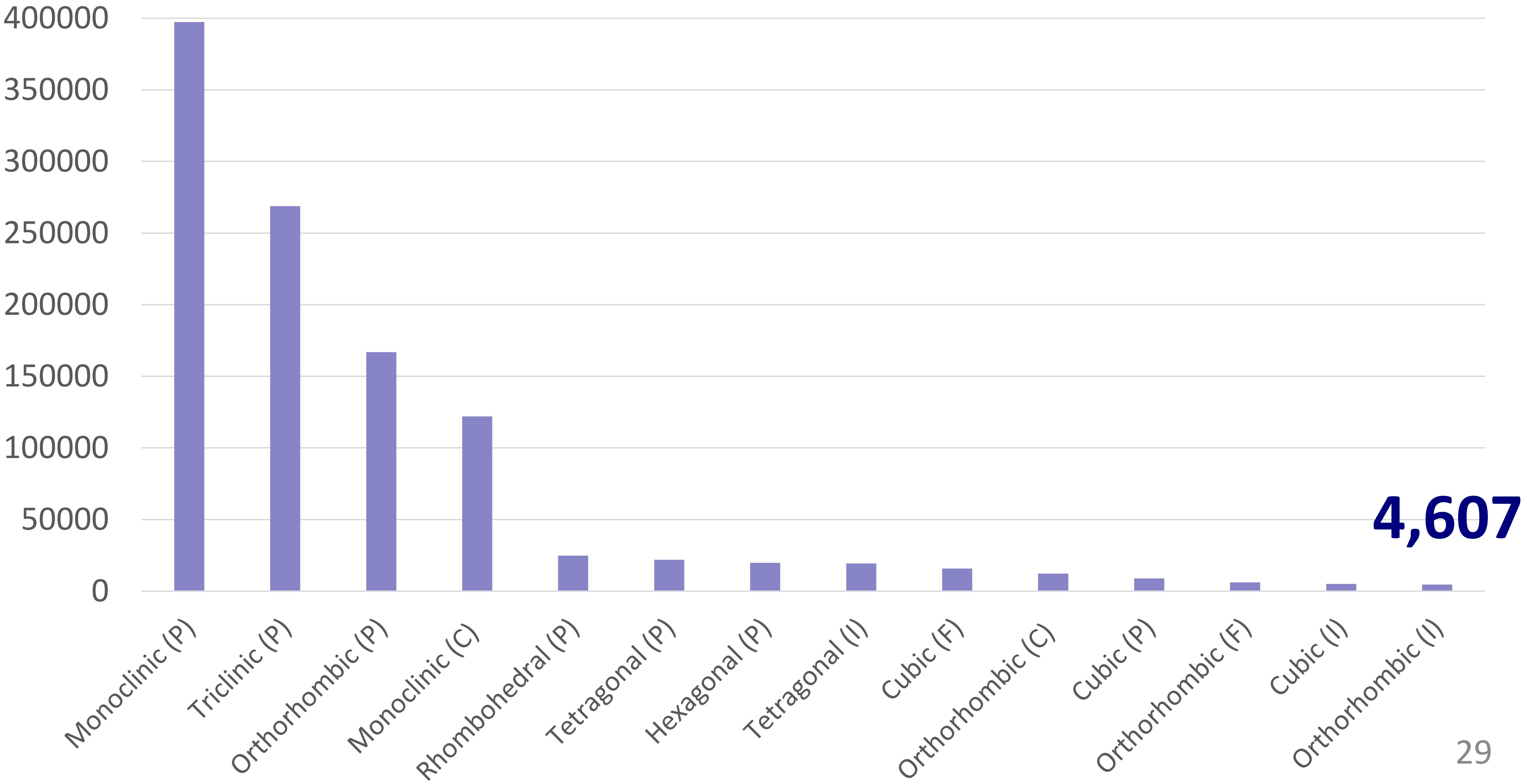
Results on Excluded Data



Data Imbalance

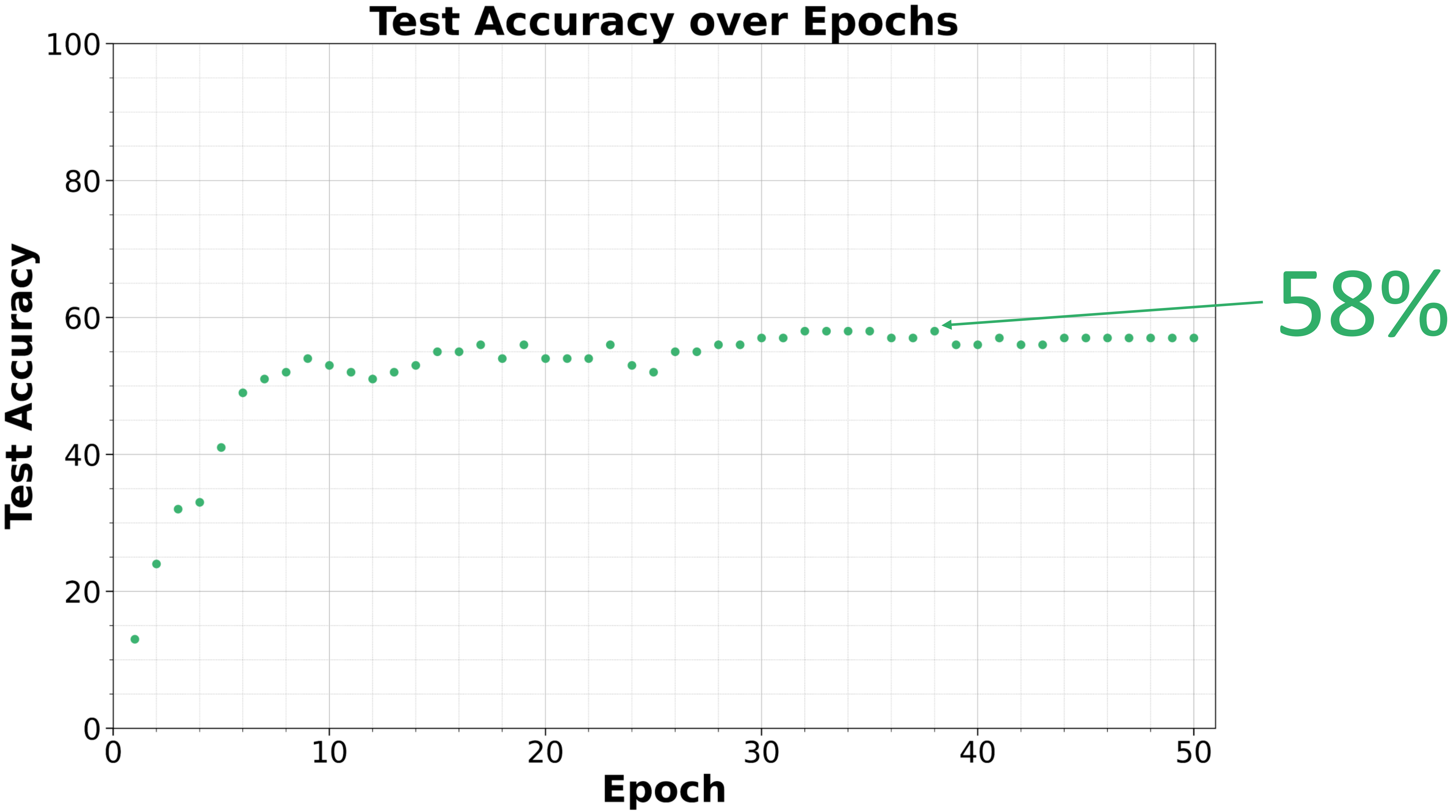


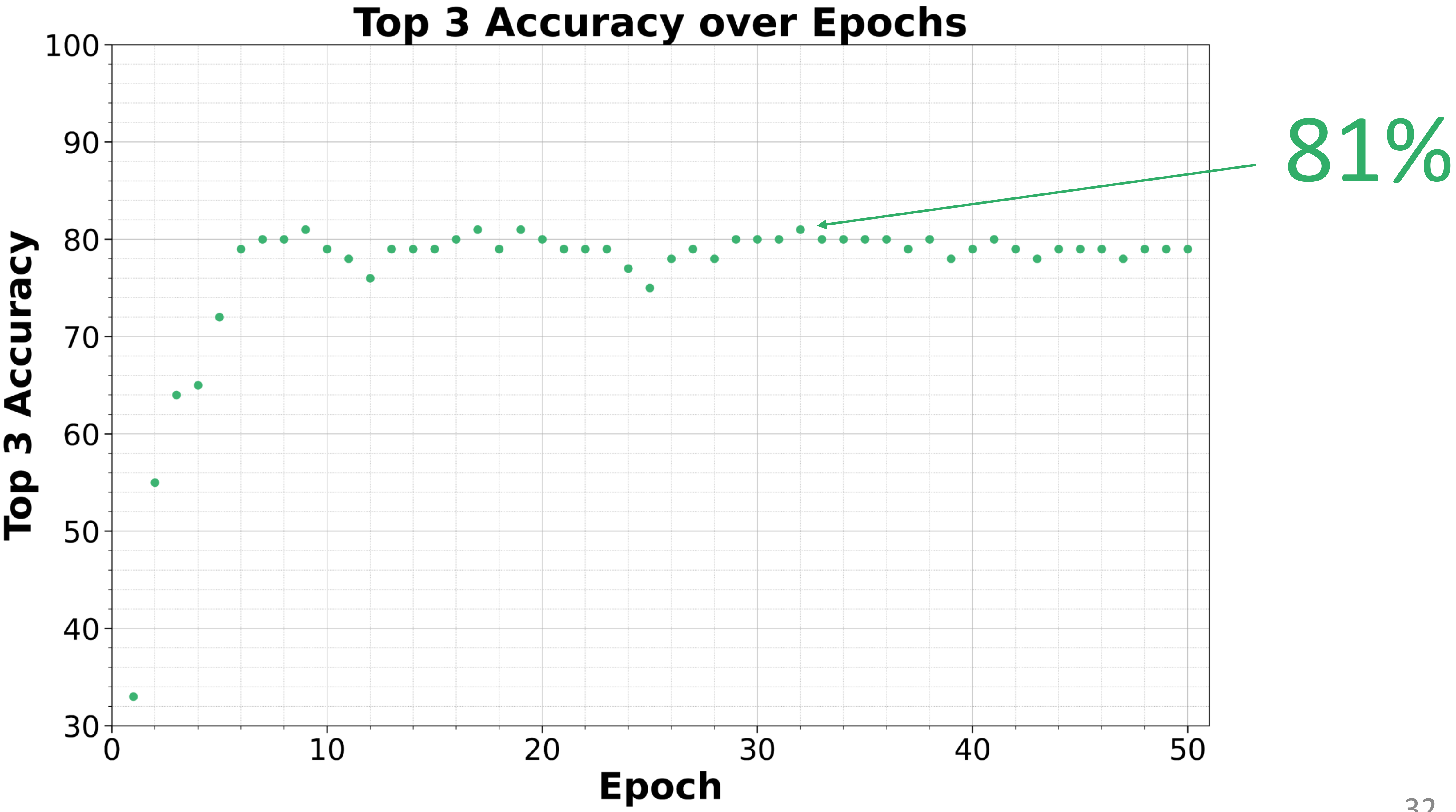
Data Imbalance



Balance Existing Data

Only use 4,607 patterns for each Bravais lattice

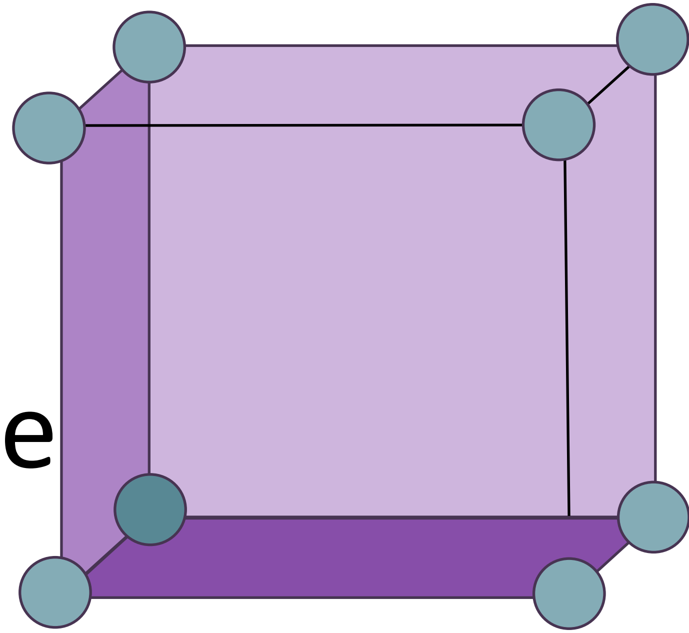

Test Accuracy over Epochs
Test Accuracy
Epoch
0
20
40
60
80
100
0
10
20
30
40
50
58%



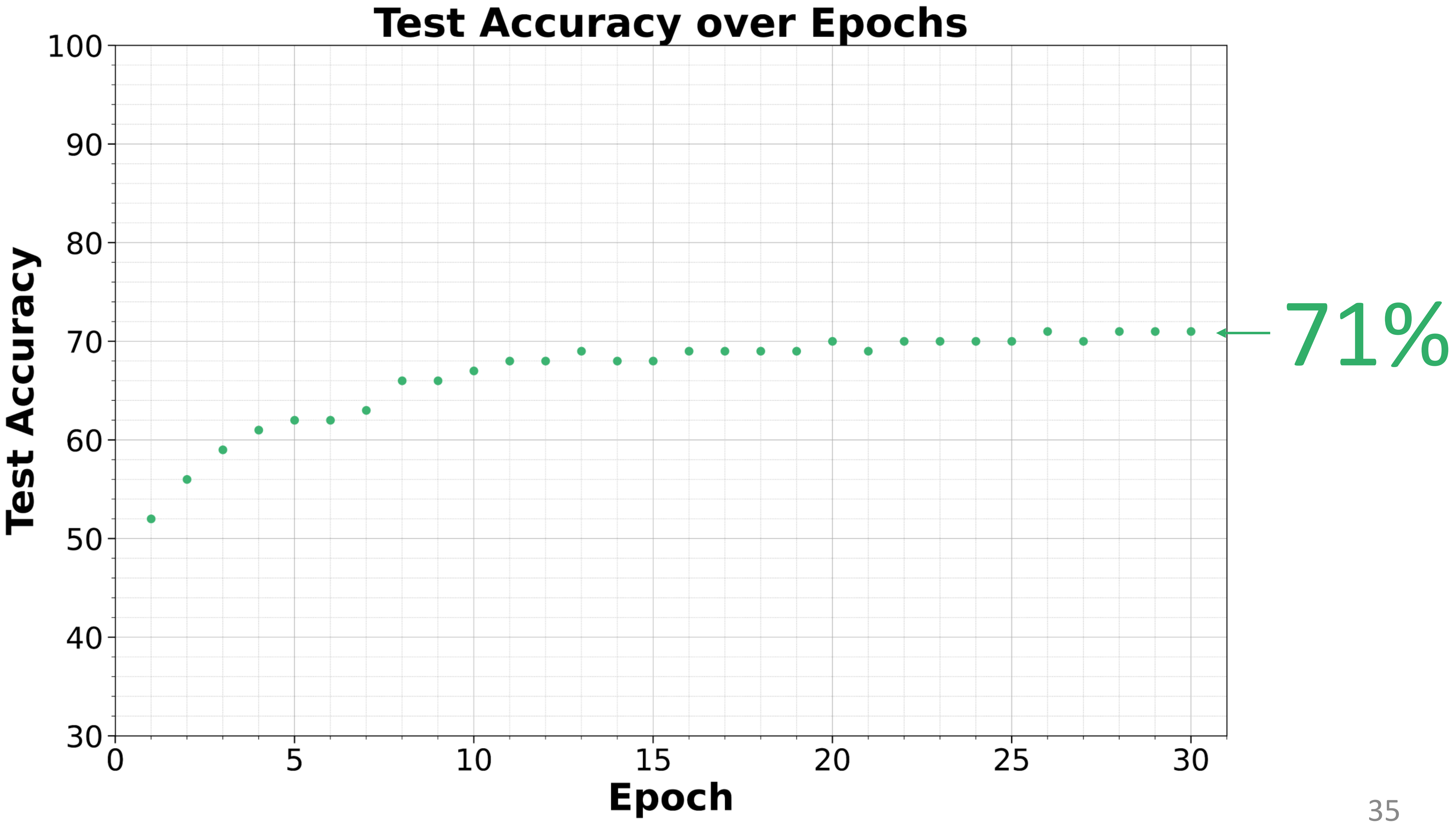
Top 3 Accuracy over Epochs
Top 3 Accuracy
Epoch
100
90
80
70
60
50
40
30
0
10
20
30
40
50
81%

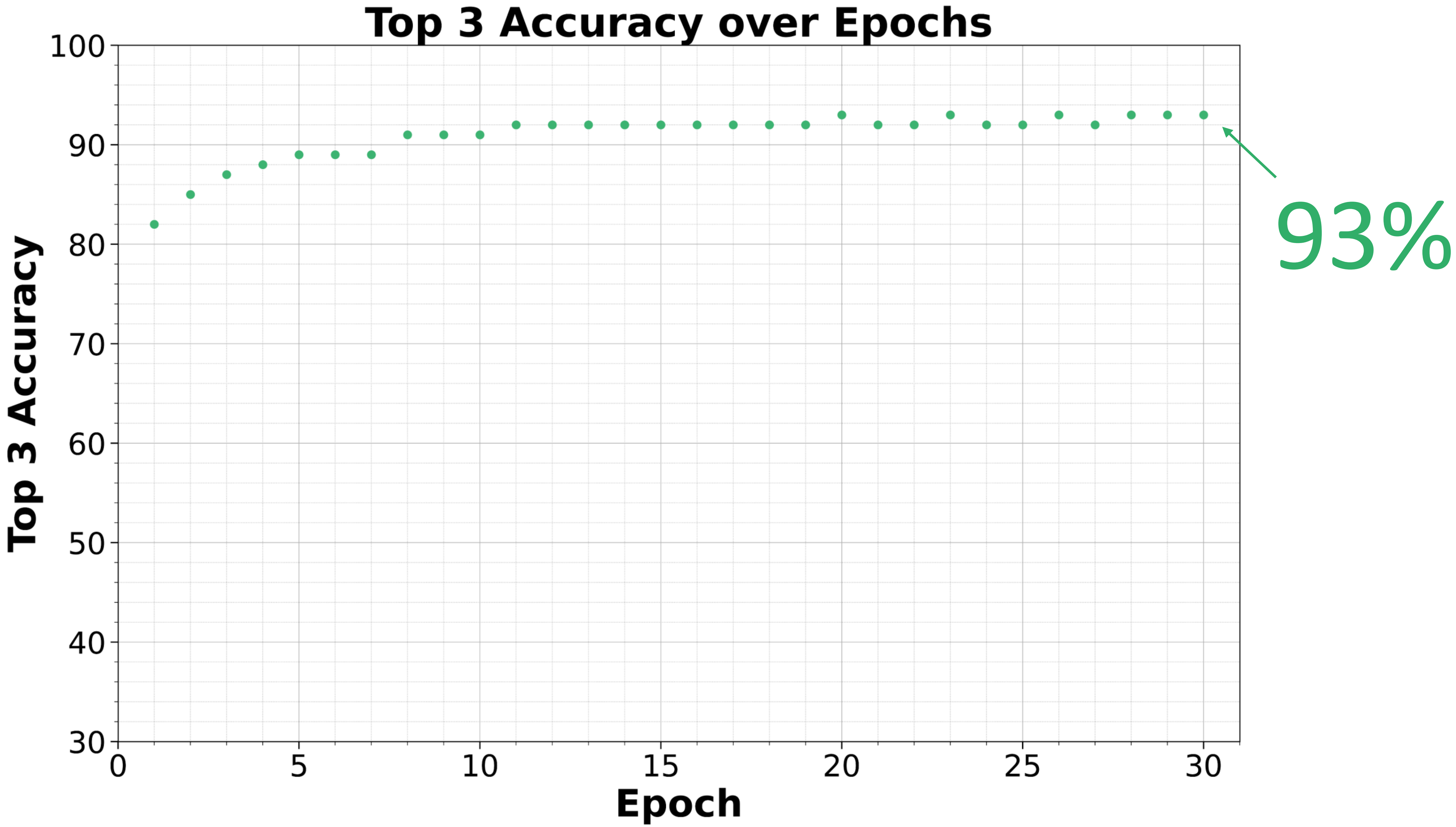
Data Augmentation

- Adjust unit cell parameters
- Maintain Bravais lattice
- 50,000 patterns per Bravais lattice
- 200,000 patterns per Bravais lattice

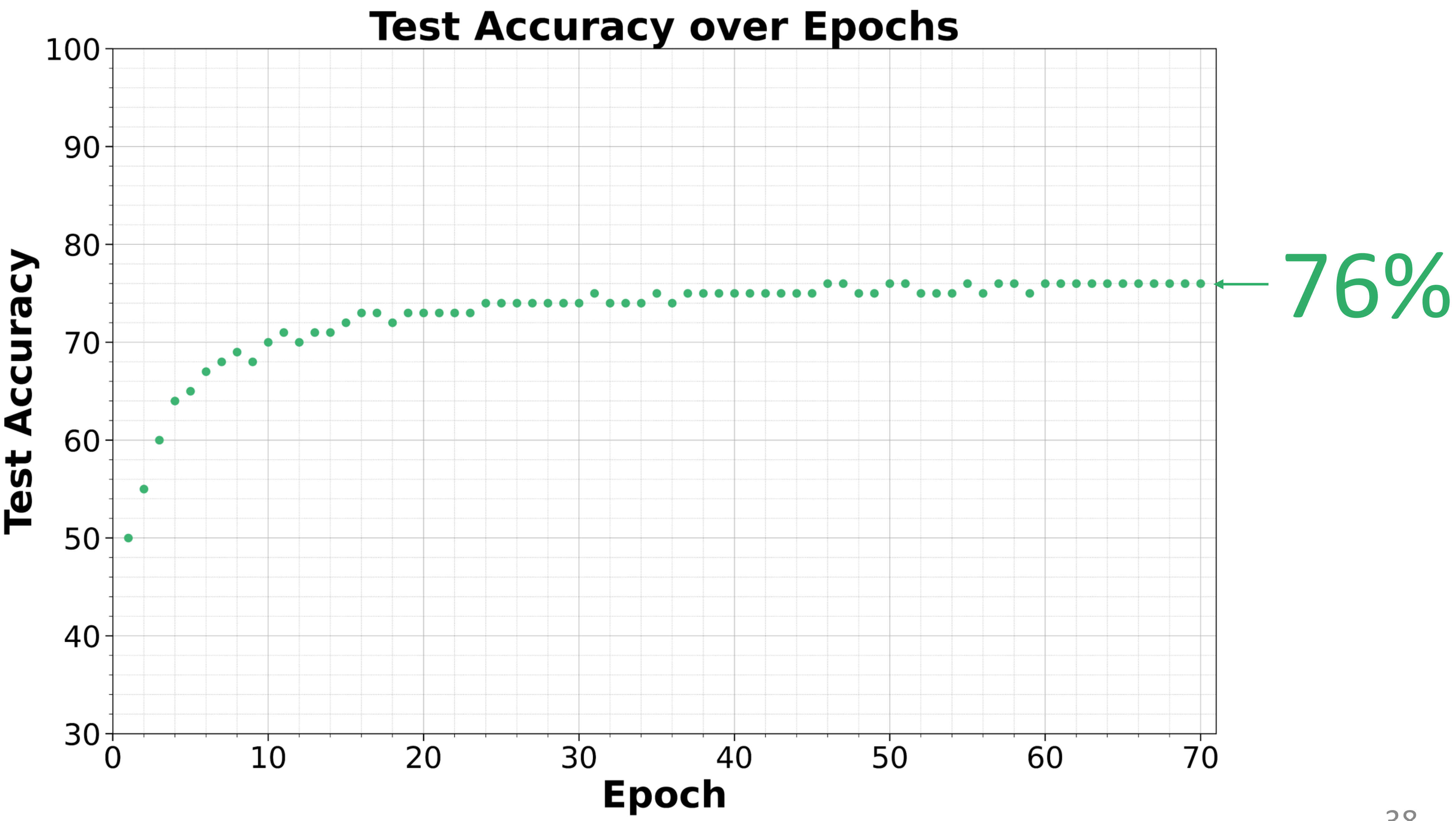


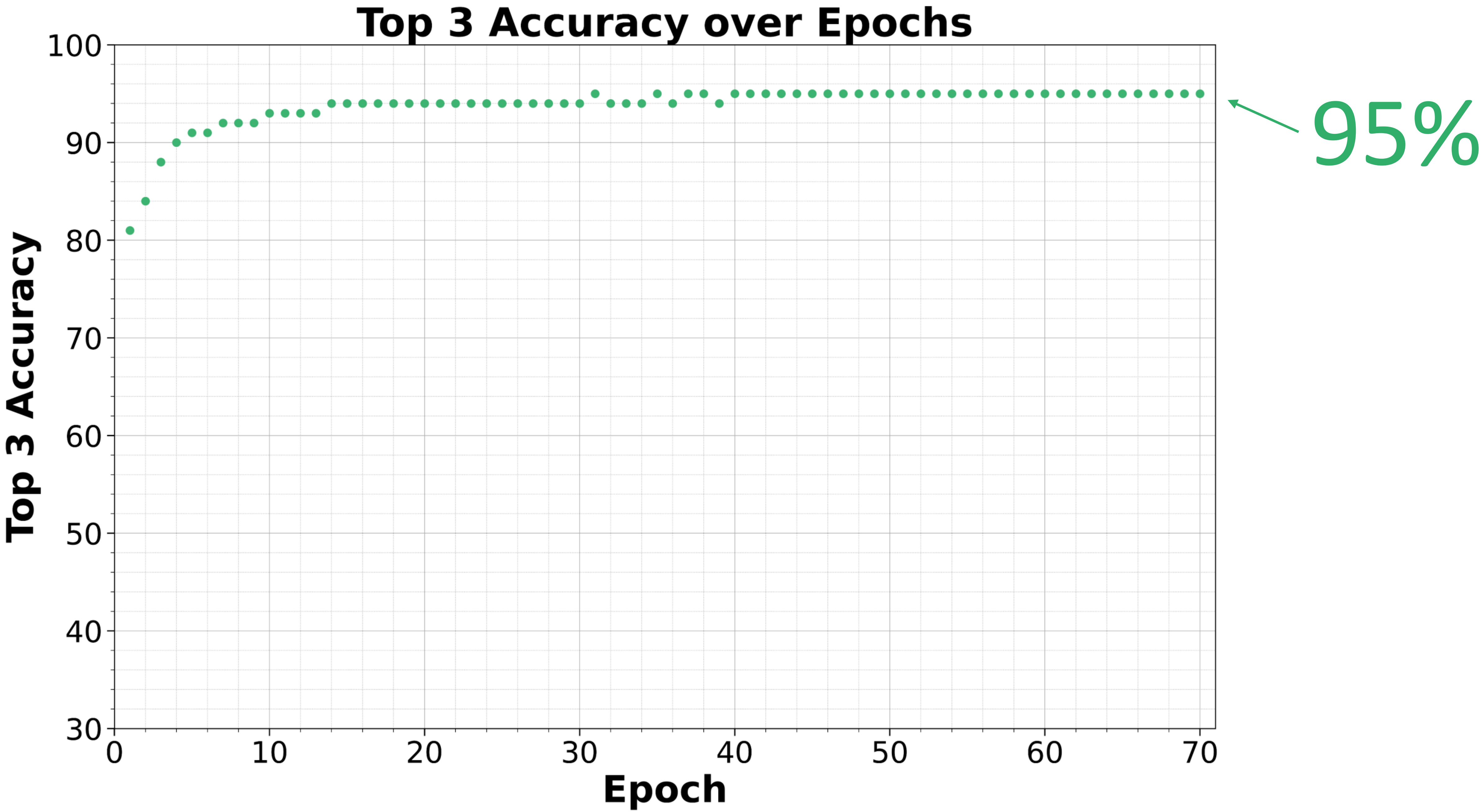
50,000 Patterns per Bravais lattice


Test Accuracy over Epochs
Test Accuracy
Epoch
100
90
80
70
60
50
40
30
0
5
10
15
20
25
30
71%


Top 3 Accuracy over Epochs
Top 3 Accuracy
Epoch
100
90
80
70
60
50
40
30
0
5
10
15
20
25
30
93%

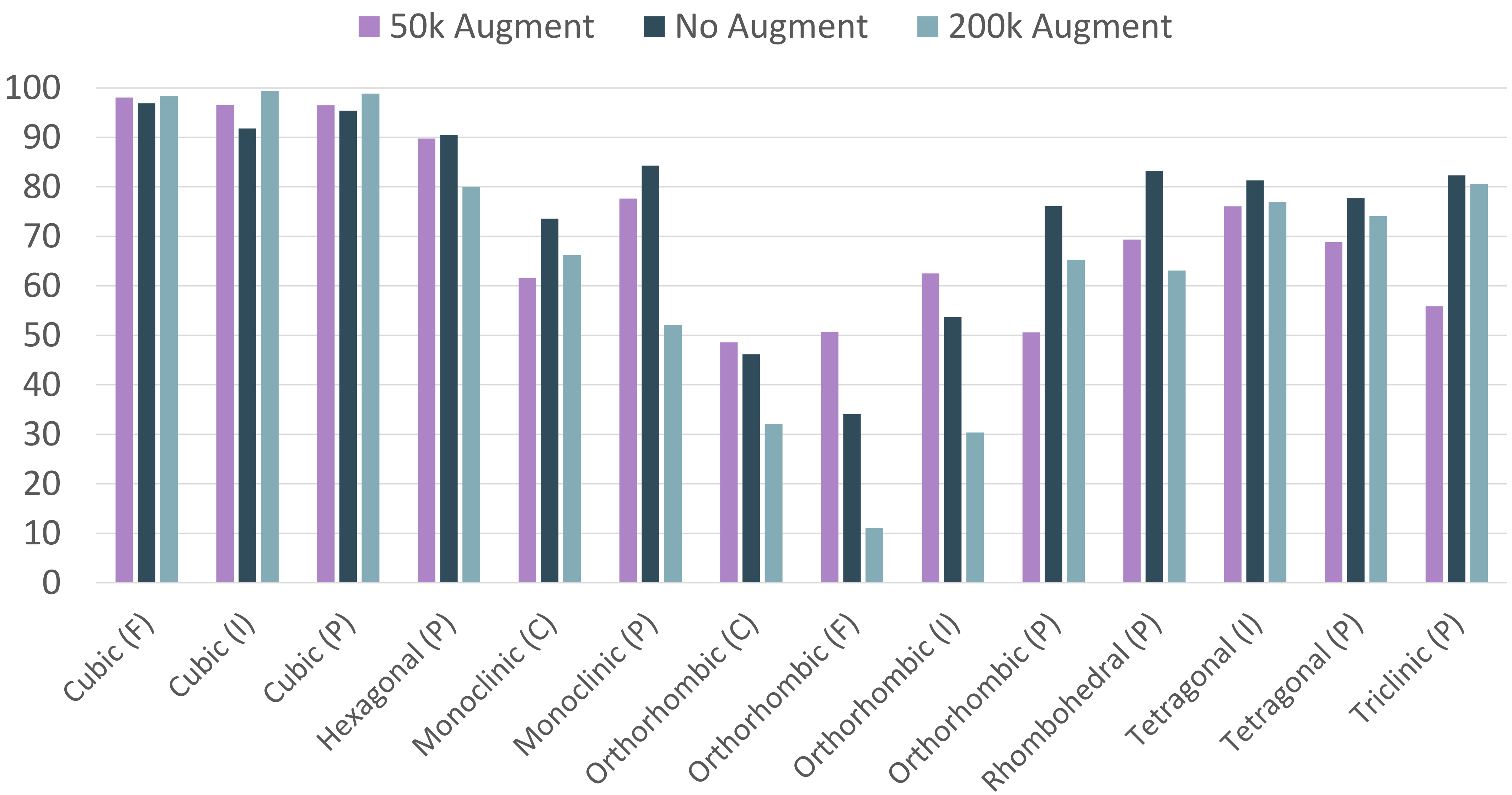
200,000 Patterns per Bravais lattice


Test Accuracy over Epochs
Test Accuracy
Epoch
100
90
80
70
60
50
40
30
0
10
20
30
40
50
60
70
76%



Top 3 Accuracy over Epochs
Top 3 Accuracy
Epoch
100
90
80
70
60
50
40
30
0
10
20
30
40
50
60
70
95%

Comparison



50k Augment
No Augment
200k Augment
100
90
80
70
60
50
40
30
20
10
0
Cubic (F)
Cubic (I)
Cubic (P)
Hexagonal (P)
Monoclinic (C)
Monoclinic (P)
Orthorhombic (C)
Orthorhombic (F)
Orthorhombic (I)
Orthorhombic (P)
Rhombohedral (P)
Tetragonal (I)
Tetragonal (P)
Triclinic (P)

Future Steps

- More augmented data
- New data generation
- Publish results ☺
- Include impurity phases
- Work with real data

Acknowledgements

Mentor

William Ratcliff

SURF Directors

Julie Borchers

Leland Harriger

Susana Marujo Teixeira

Academic Program Manager

Cara O'Malley

[1]

[2]

1. *Center for high resolution neutron scattering*. NIST. (2023, June 8). https://www.nist.gov/ncnr/chrns

2. *Utakeit. education made easy.* UTakeIt. (n.d.). https://utakeit.tacc.utexas.edu/

References

Lolla, S., Liang, H., Kusne, A. G., Takeuchi, I., & Ratcliff, W. (2022). A semi-supervised deep-learning approach for automatic crystal structure classification. *Journal of Applied Crystallography*, *55*(4), 882–889. https://doi.org/10.1107/s1600576722006069

Radaelli, P. G. (2016). *Symmetry in crystallography: Understanding the international tables*. Oxford University Press.

S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186. keywords: {Convolution;Neurons;Convolutional neural networks;Feature extraction;Image edge detection;machine learning;artificial neural networks;deep learning;convolutional neural networks;computer vision;Image recognition},

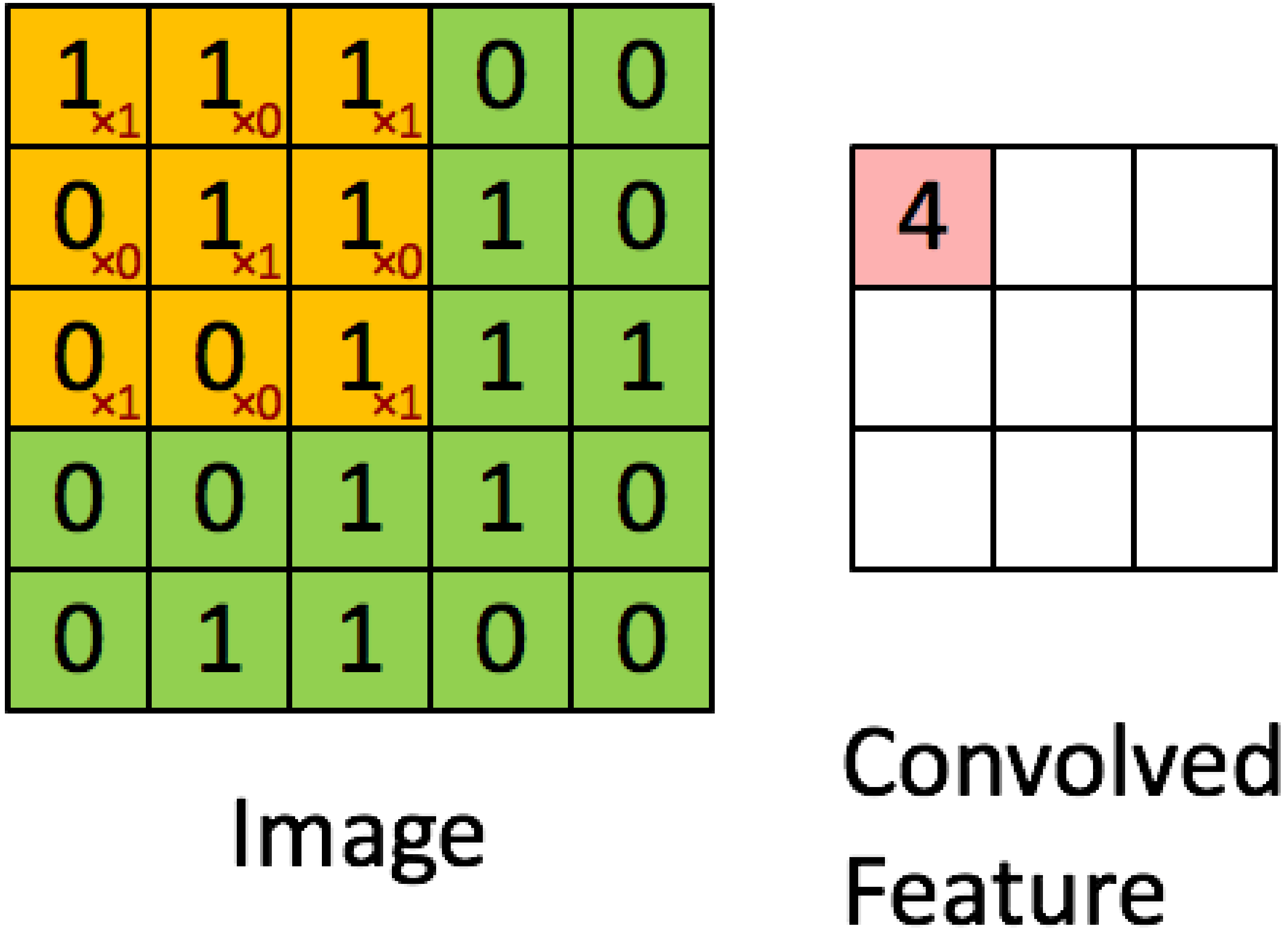
Sands, D. (2020). *Introduction to crystallography*. Dover Publications, Inc.

Sivia, D. S. (2011). *Elementary scattering theory for X-ray and neutron users*. Oxford University Press.

Young, R. A. (1995). *The rietveld method*. International Union of Crystallograhy.

Extra Slides

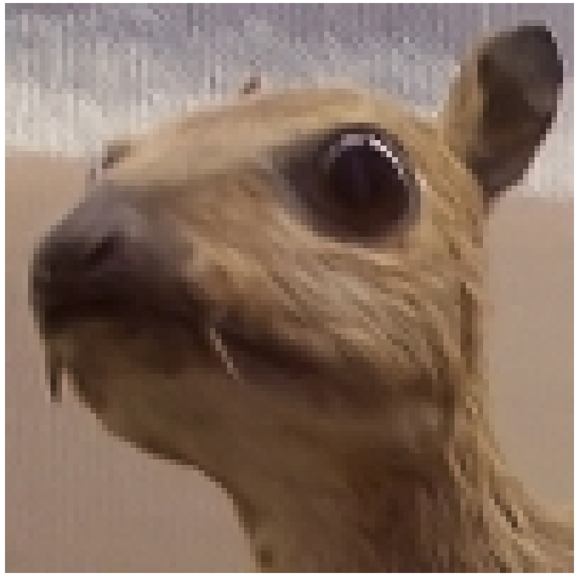
Convolutional Neural Network



Convolution. NVIDIA Developer. (n.d.). https://developer.nvidia.com/discover/convolution

Convolutional Neural Network

Input image



Convolution Kernel

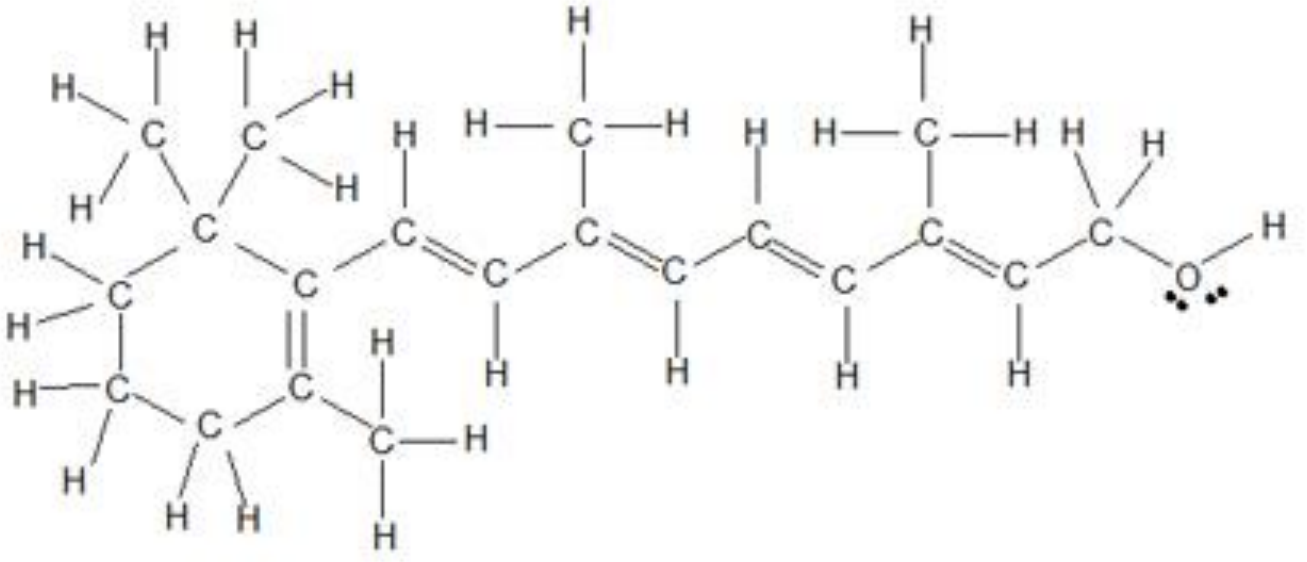
$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Feature map

Convolution. NVIDIA Developer. (n.d.). https://developer.nvidia.com/discover/convolution

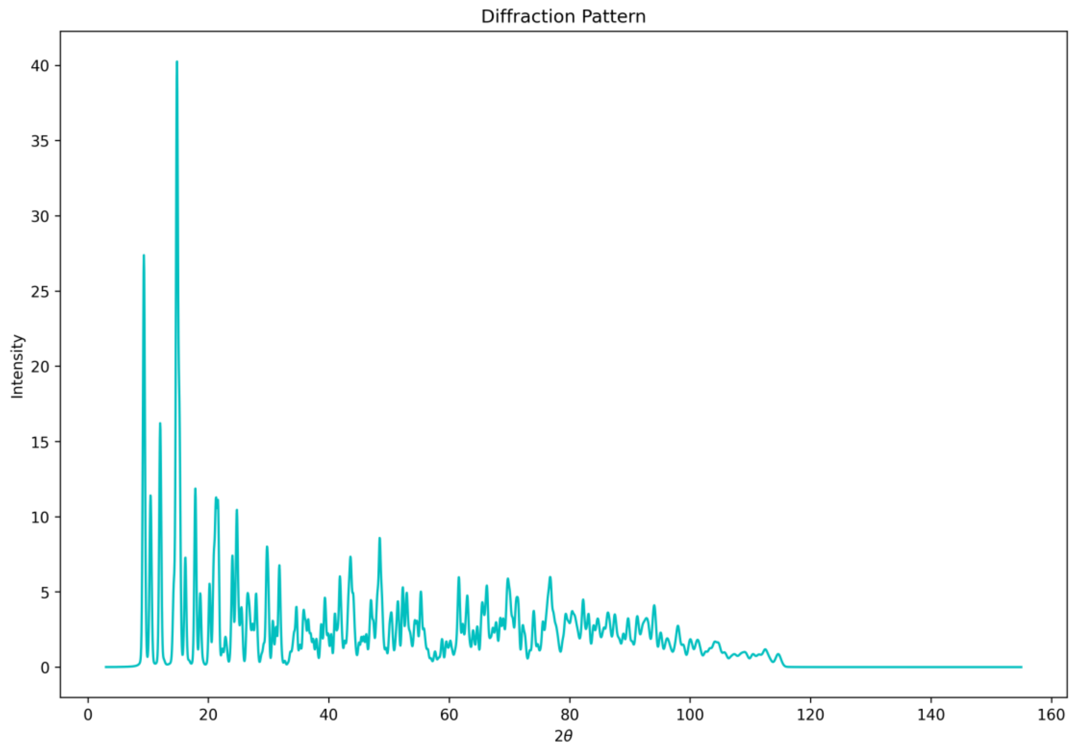
Deuterium Substitution

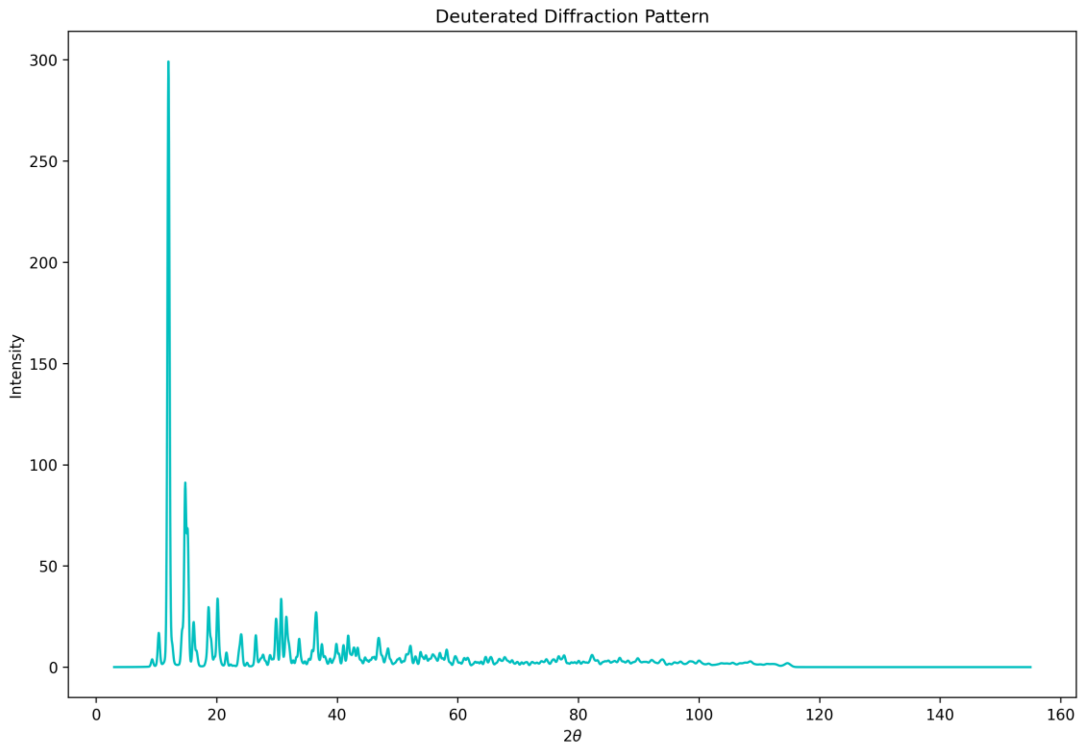
- CSD contains organic compounds
  - More C-H bonds
- Hydrogen scattering cross section: 82.03
- Deuterium scattering cross section: 7.64



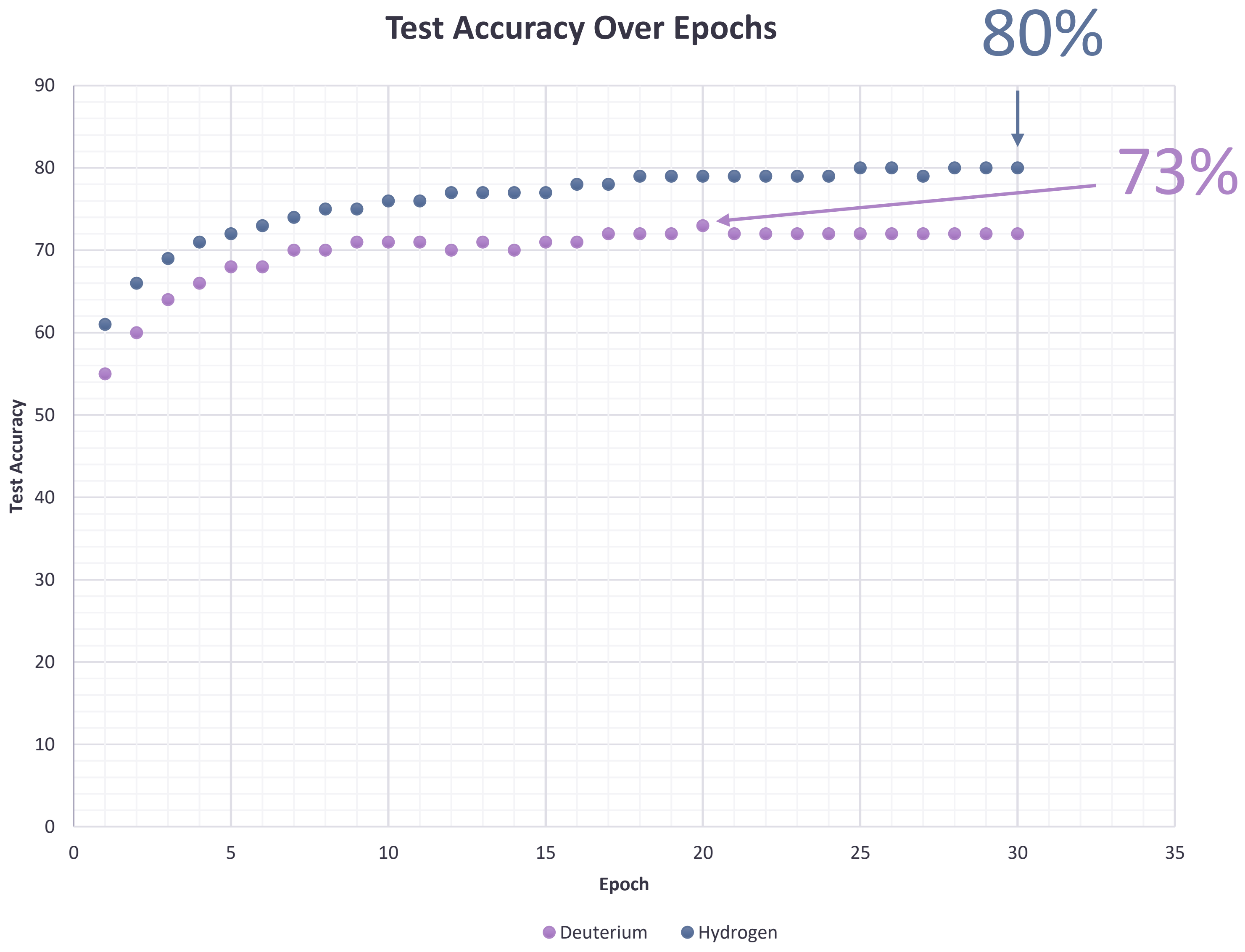
Libretexts. (2023, January 23). *Structure of organic molecules*. Chemistry LibreTexts. https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Supplemental_Modules_(Organic_Chemistry)/Fundamentals/Structure_of_Organic_Molecules

Original vs Deuterated

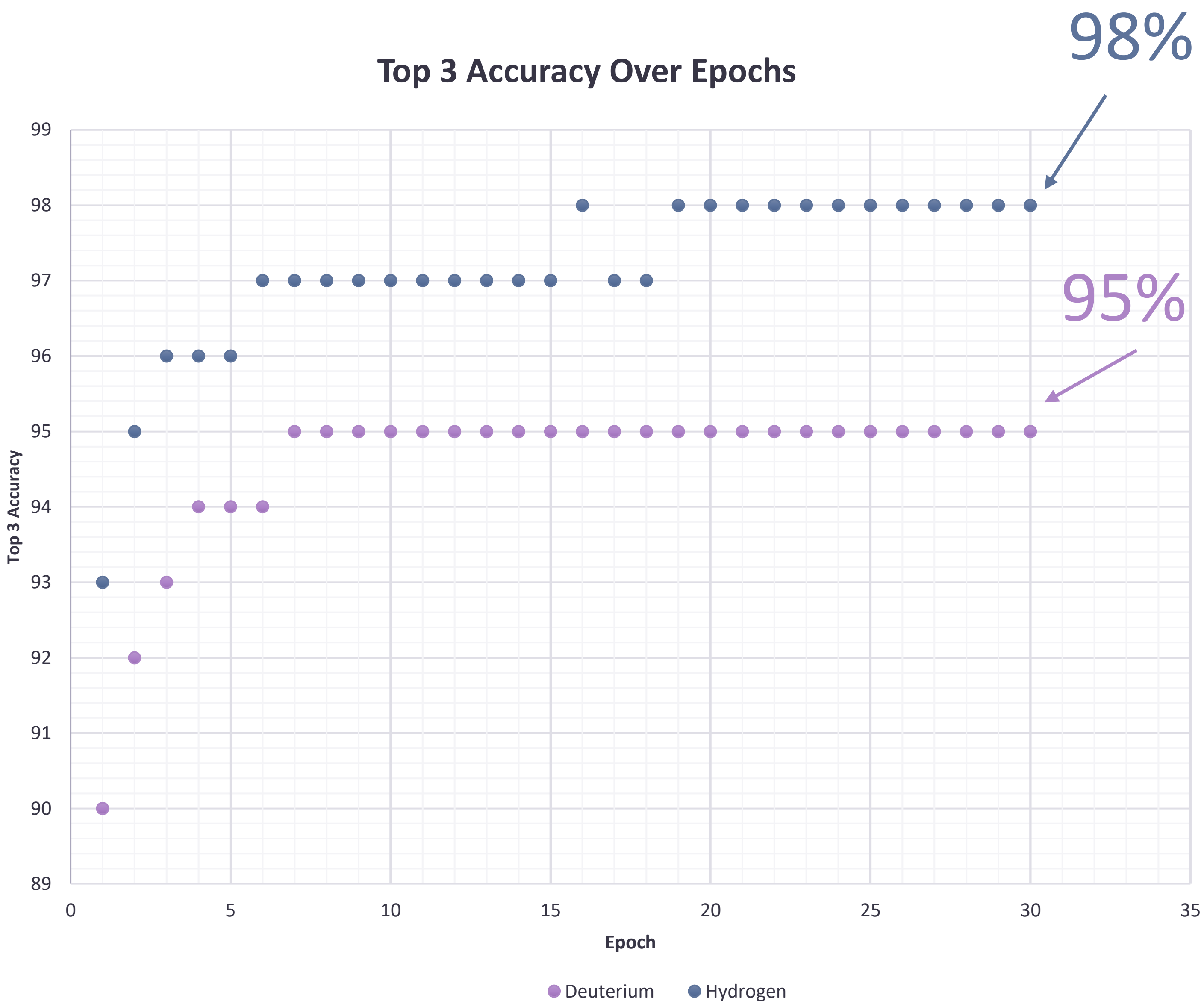




Deuterium Substitution



Deuterium Substitution



Model Information

- Using 3 GPU, takes a little under 5 hours to train

ICSD + CSD

  - Epochs: 30
  - Batch size: 50
  - Learning rate: 0.0005

  - 580 seconds per epoch
    - 8.83 minutes

Resolution

- BT-1 resolution
  - U: 3.47
  - V: -2.78
  - W: 1.66

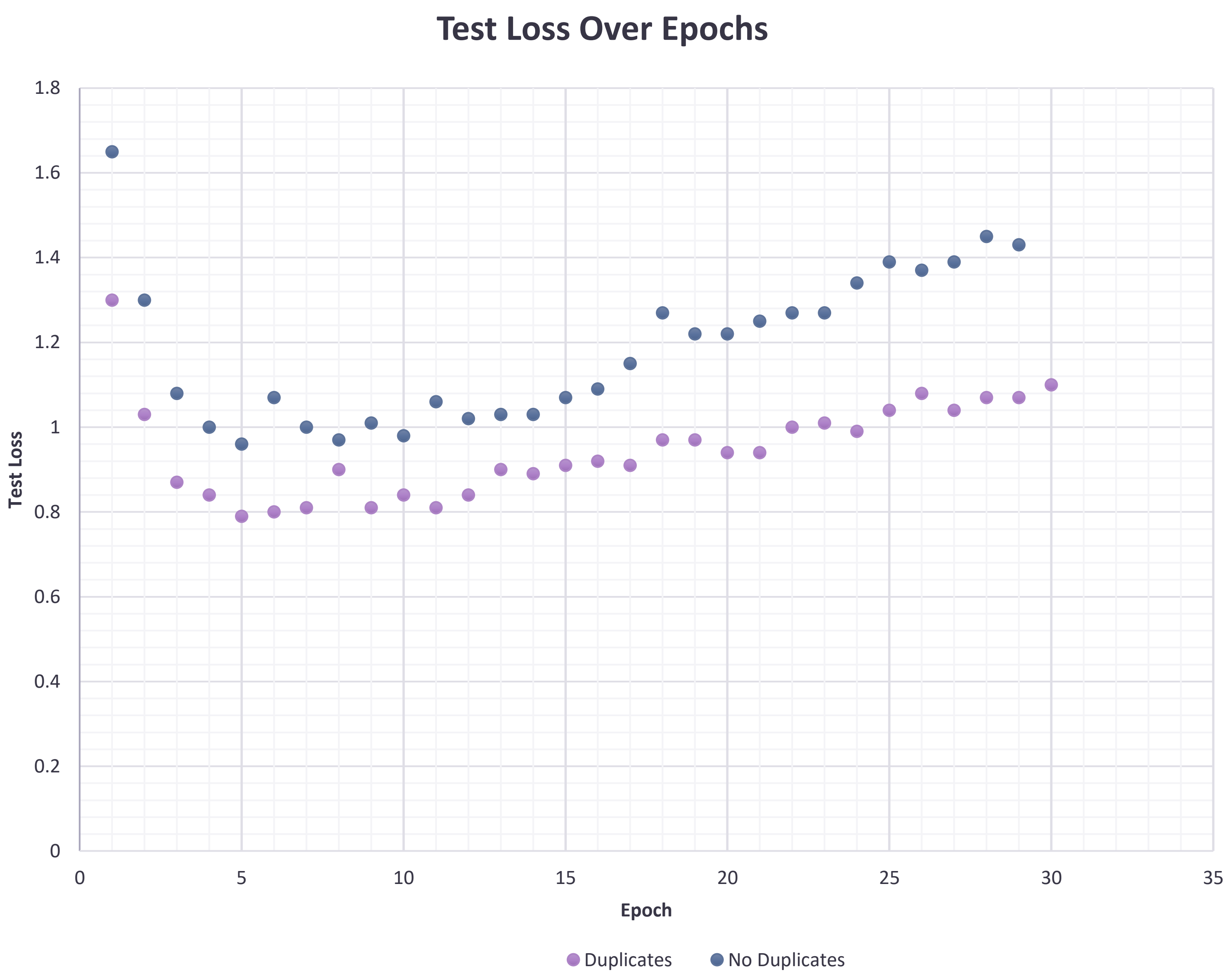
$$\frac{C_0^{1/2}}{H_K \pi^{1/2}} \exp\left(-\frac{C_0(2\theta_i - 2\theta_k)^2}{H_K^2}\right)$$

$$H^2 = U tan^2\theta + V tan\theta + \mathrm{W}$$

Test Loss

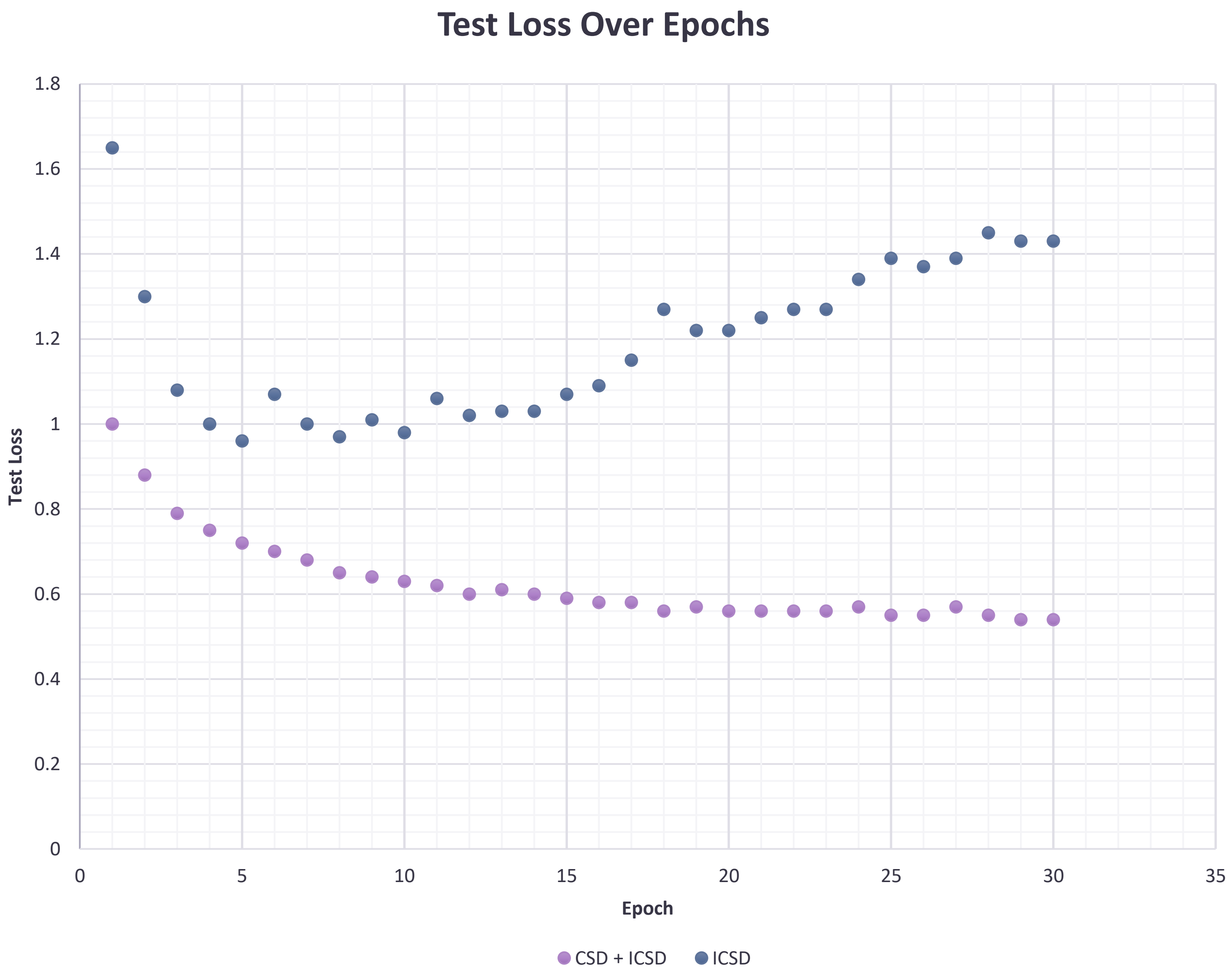
Removing Duplicates

181,362 → 123,039

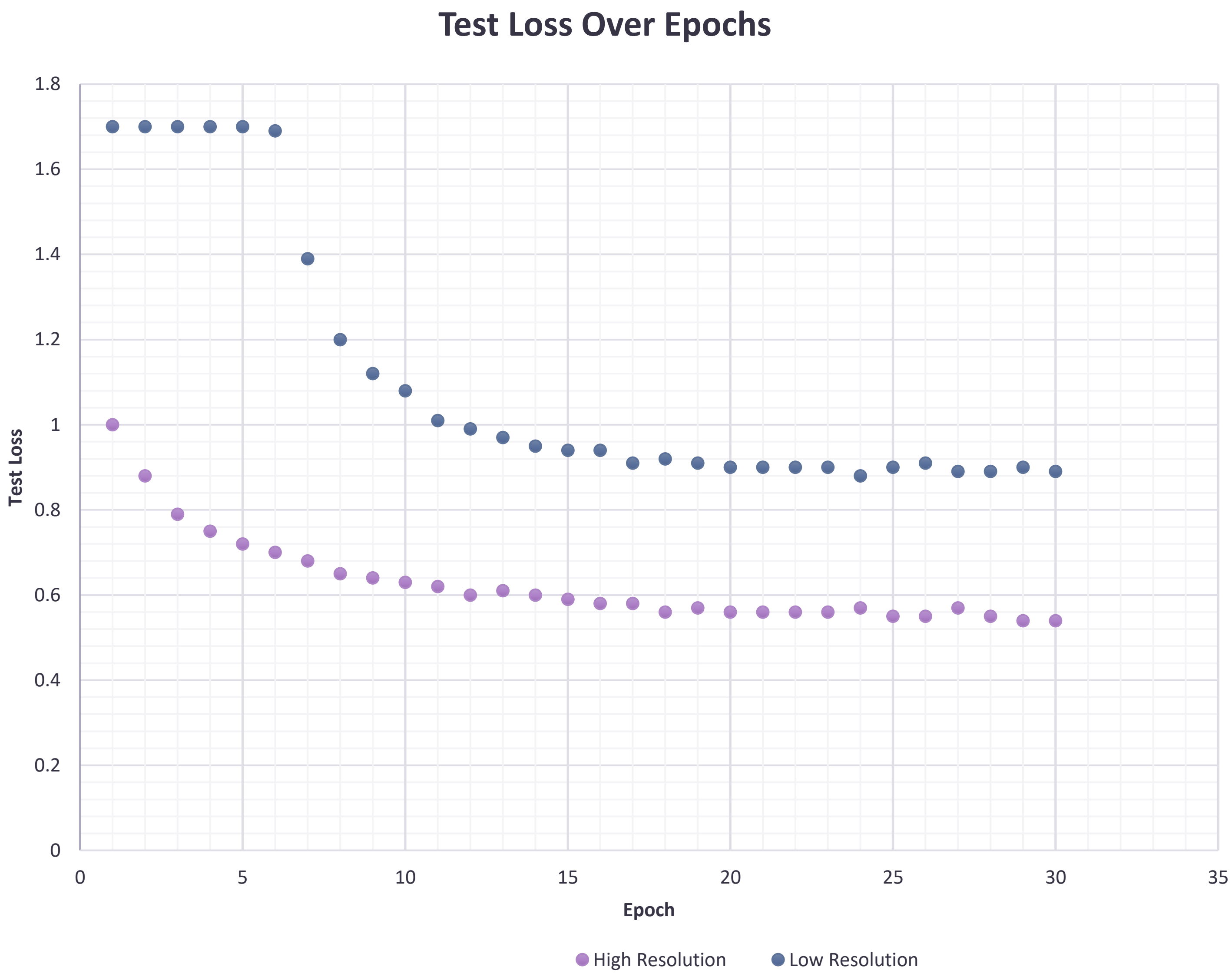


Adding More Data

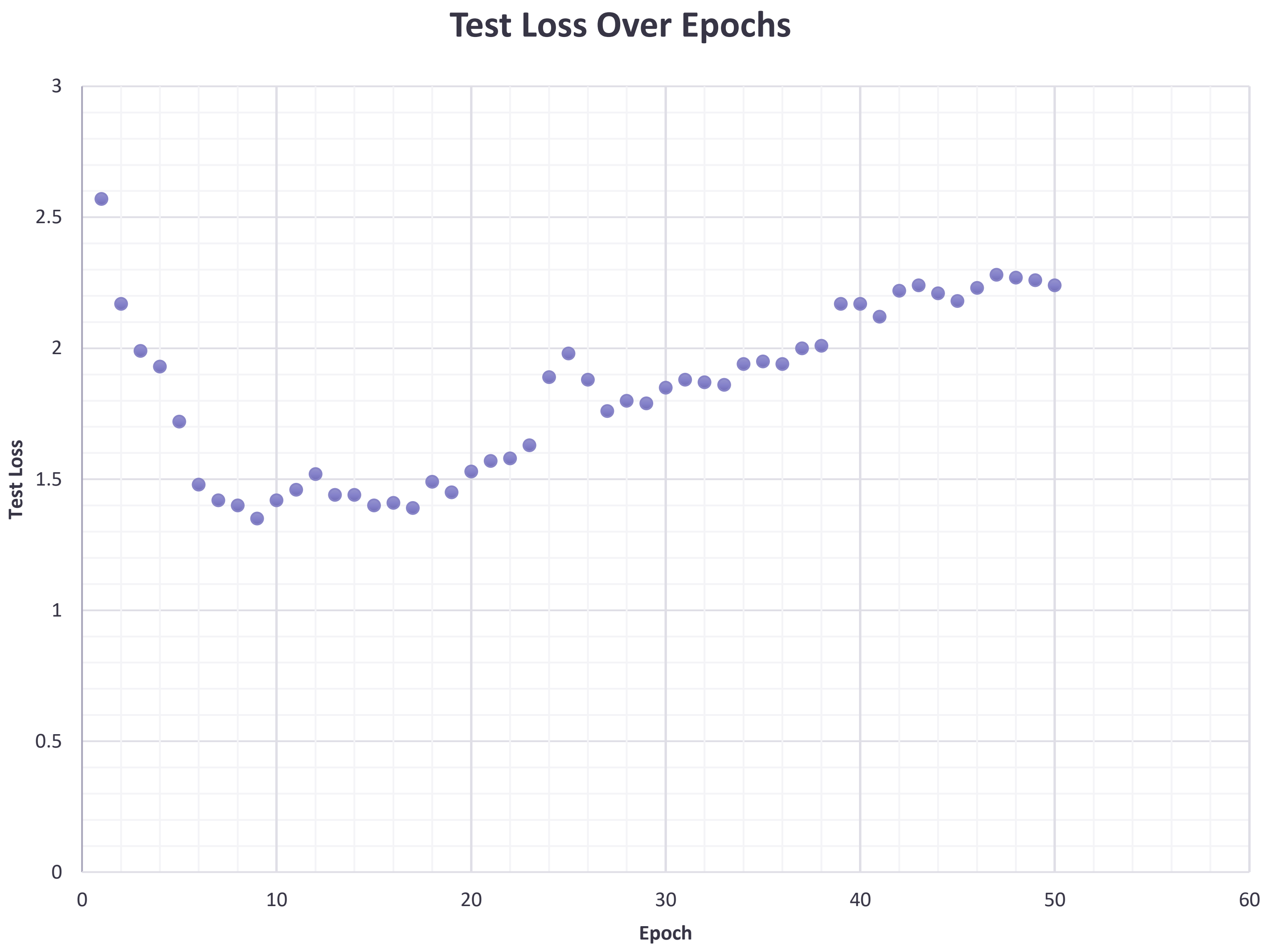
Adding data from Cambridge Structural Database (~ 1 million)



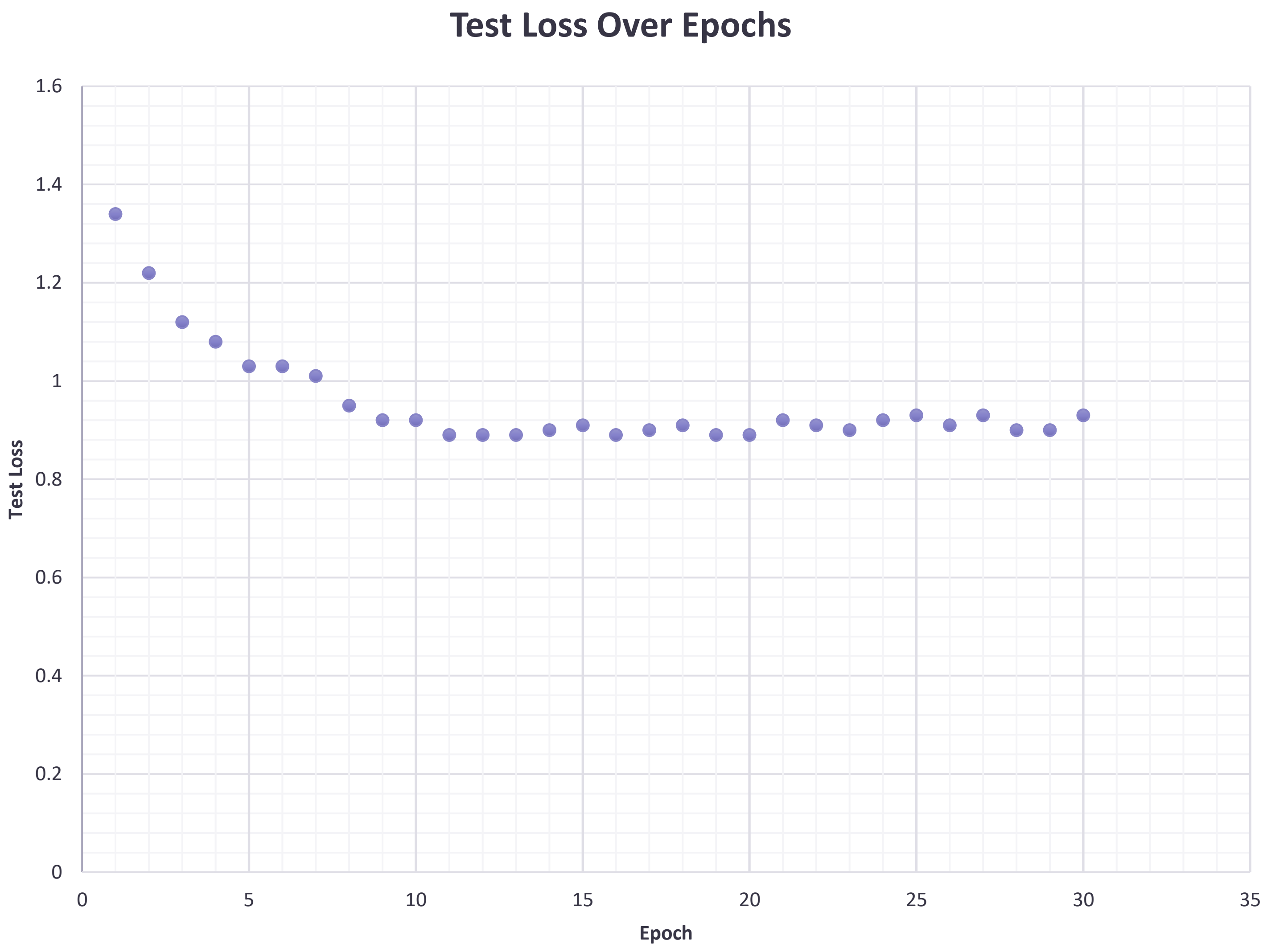
Improve Resolution



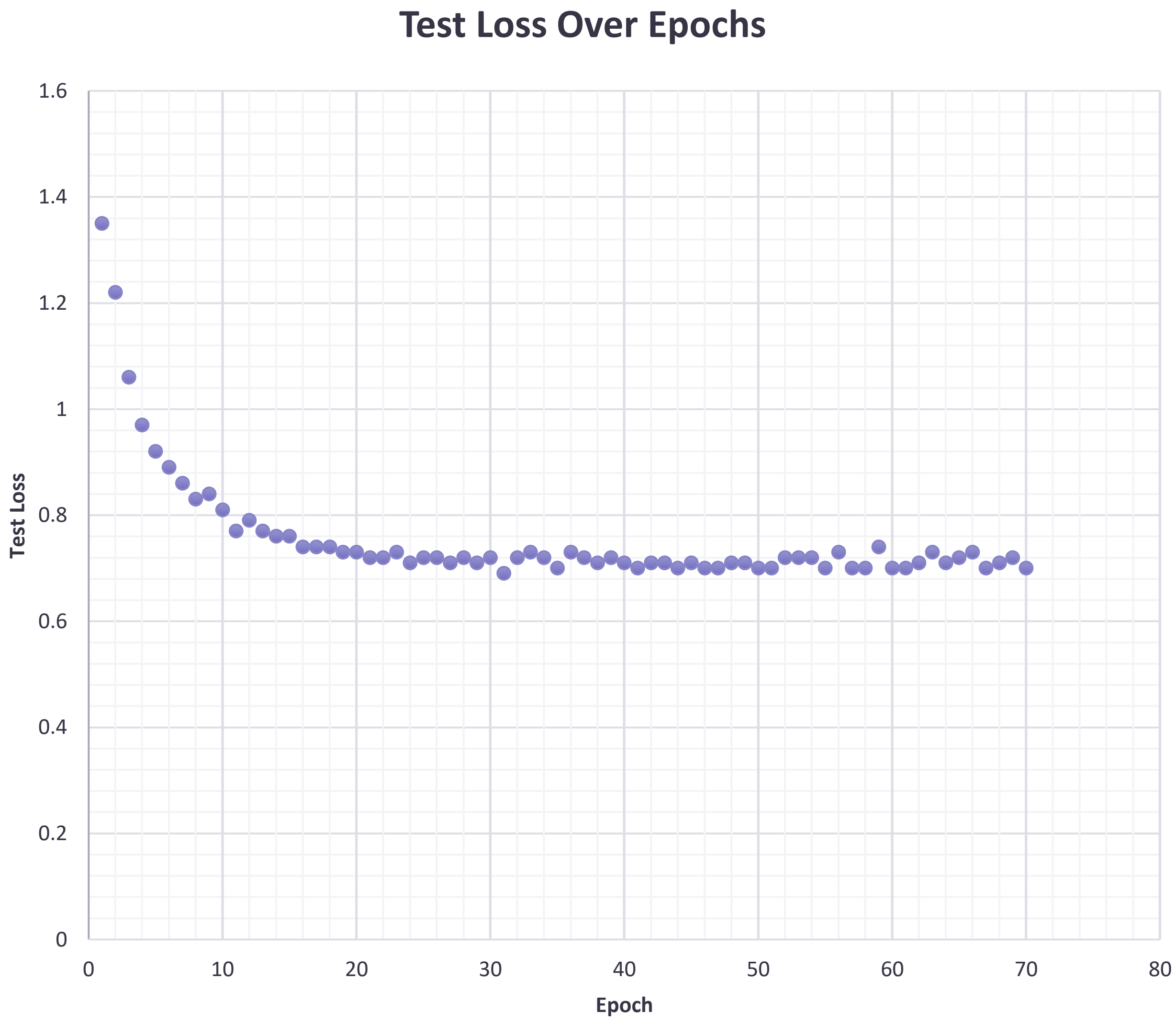
Balance Existing Data



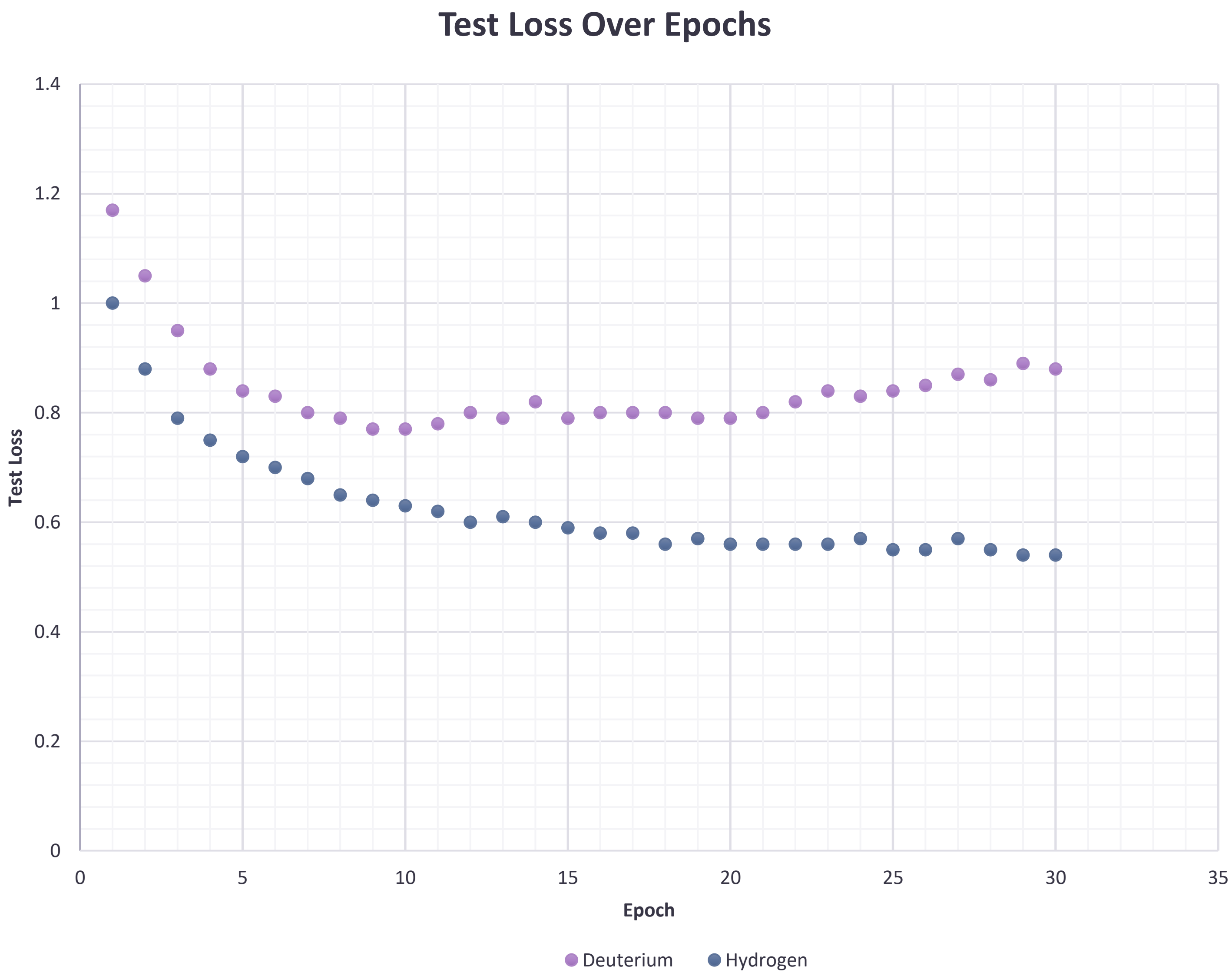
50k Data Augmentation



200k Data Augmentation



Deuterium Substitution



Confusion Matrix

ICSD + CSD

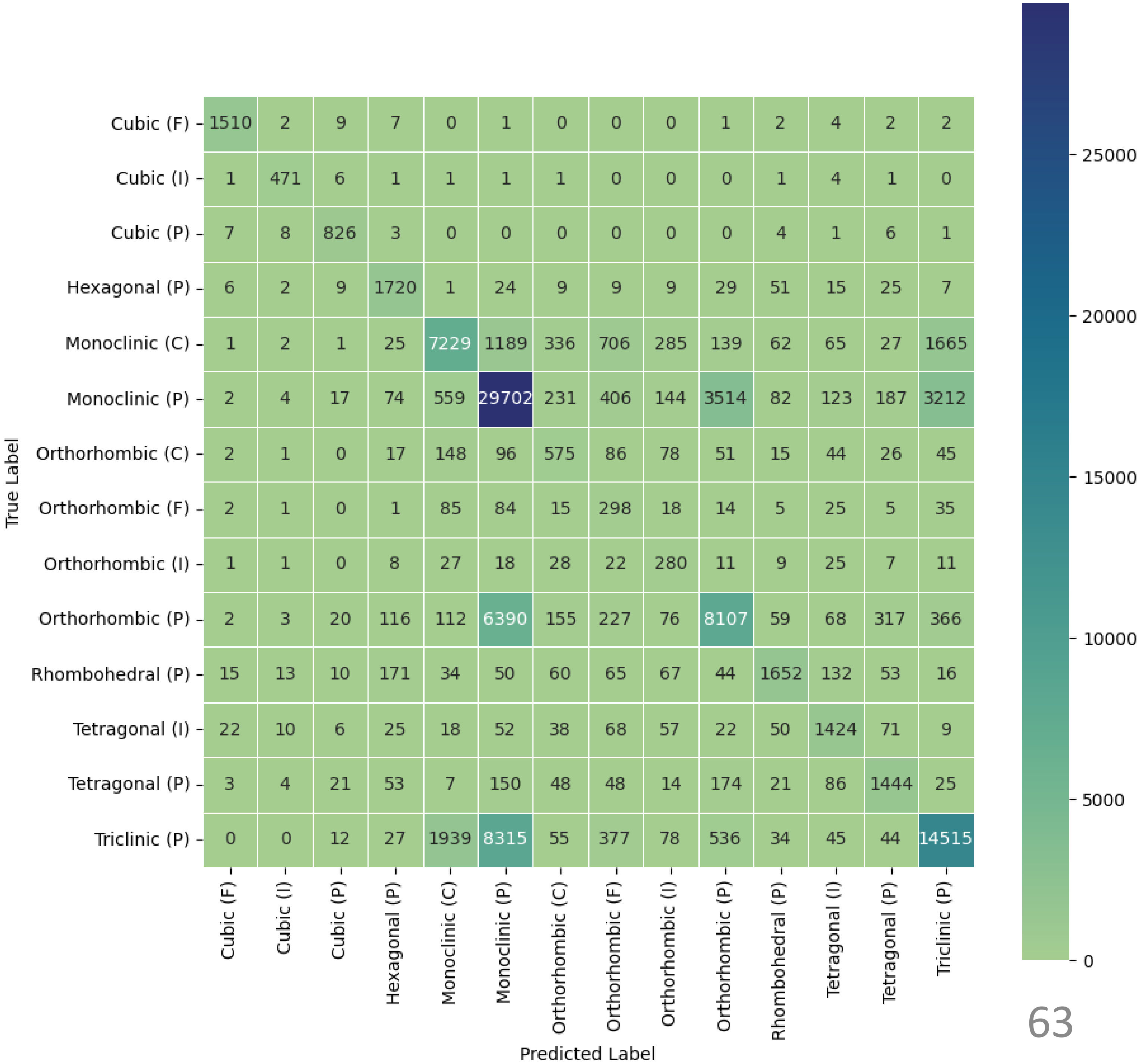
80% accuracy

True Label \ Predicted Label	Cubic (F)	Cubic (I)	Cubic (P)	Hexagonal (P)	Monoclinic (C)	Monoclinic (P)	Orthorhombic (C)	Orthorhombic (F)	Orthorhombic (I)	Orthorhombic (P)	Rhombohedral (P)	Tetragonal (I)	Tetragonal (P)	Triclinic (P)
Cubic (F)	740	2	3	5	0	0	0	1	0	0	2	11	0	0
Cubic (I)	3	235	8	0	0	0	0	0	0	0	4	3	3	0
Cubic (P)	5	5	394	0	0	0	0	1	0	2	0	1	3	2
Hexagonal (P)	1	0	8	883	0	4	1	1	0	27	27	9	12	3
Monoclinic (C)	0	1	0	2	4375	345	34	42	13	50	14	20	1	1044
Monoclinic (P)	0	0	0	7	112	15994	14	11	5	1317	9	9	26	1460
Orthorhombic (C)	0	0	1	7	137	50	265	7	11	28	18	21	11	18
Orthorhombic (F)	0	0	0	2	81	50	4	103	4	13	0	9	1	35
Orthorhombic (I)	0	0	1	3	34	15	9	0	115	5	10	14	2	6
Orthorhombic (P)	1	0	3	13	28	1589	12	6	4	6154	11	6	75	181
Rhombohedral (P)	1	2	3	52	25	23	9	3	8	20	970	30	7	13
Tetragonal (I)	4	4	1	10	23	33	11	8	8	19	28	766	20	7
Tetragonal (P)	2	0	11	14	6	55	4	1	1	98	8	20	807	12
Triclinic (P)	2	0	2	5	340	1783	2	4	1	160	3	0	6	10722

Color scale: 0, 2000, 4000, 6000, 8000, 10000, 12000, 14000

50,000 Data Augmentation

66% accuracy



200,000 Data Augmentation

65% accuracy