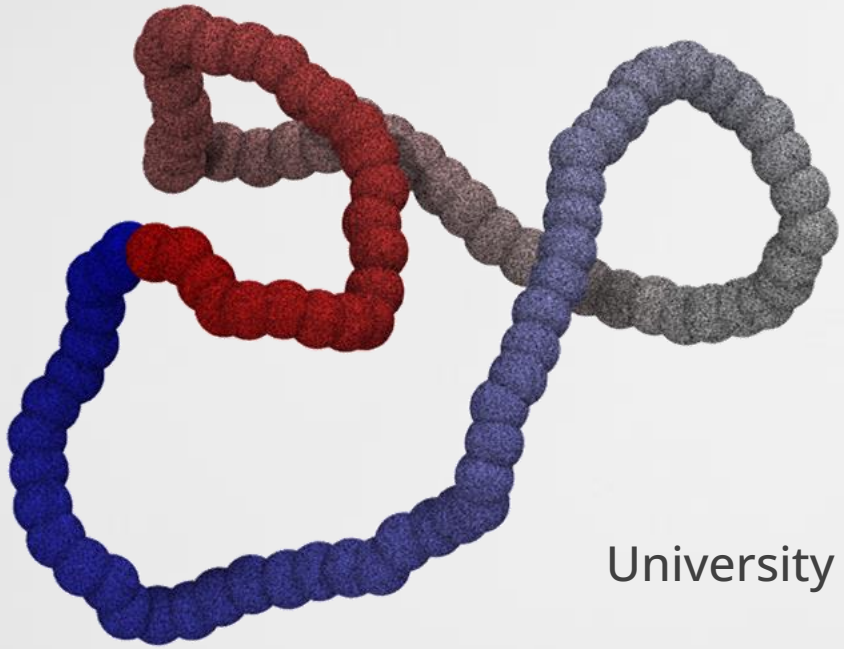
Classification of Knots using Deep Learning

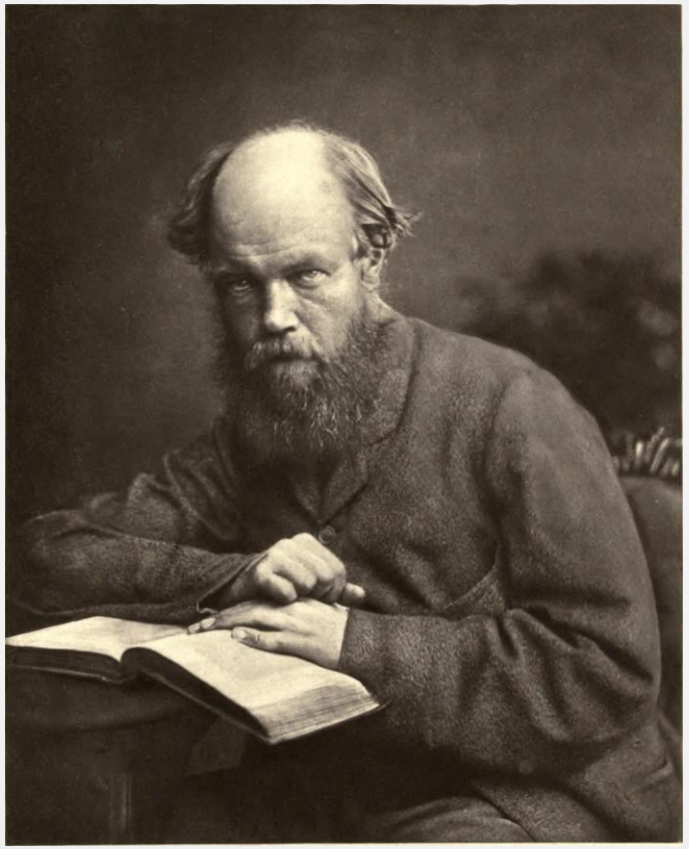


Joseph Sleiman

Supervisor: Dr. Davide Michieletto

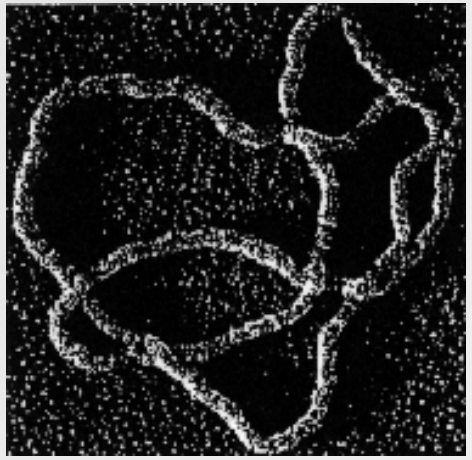
University of Edinburgh, School of Physics & Astronomy

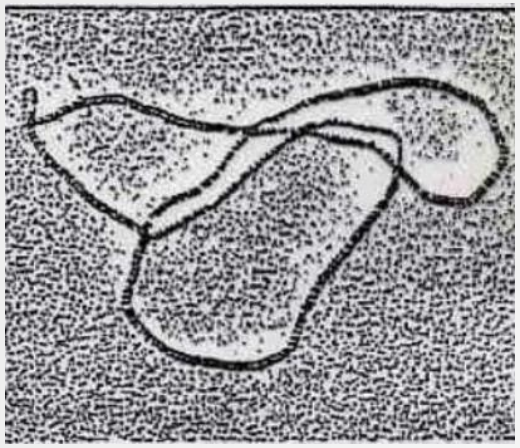
Motivation

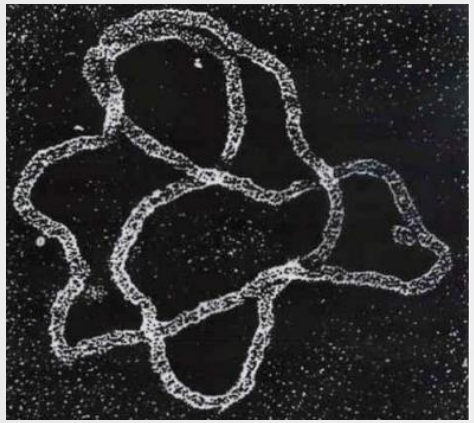


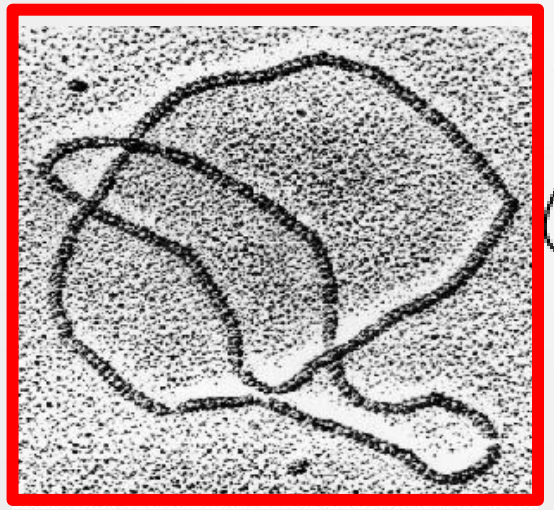
S.C. Dickson, Peter Guthrie Tait Road, Kings Buildings (2019
C. G. Knott, *Life and Scientific Work of Peter Guthrie Tait* (1911)
http://katlas.org/wiki/Other_Knot_Tables

Motivation

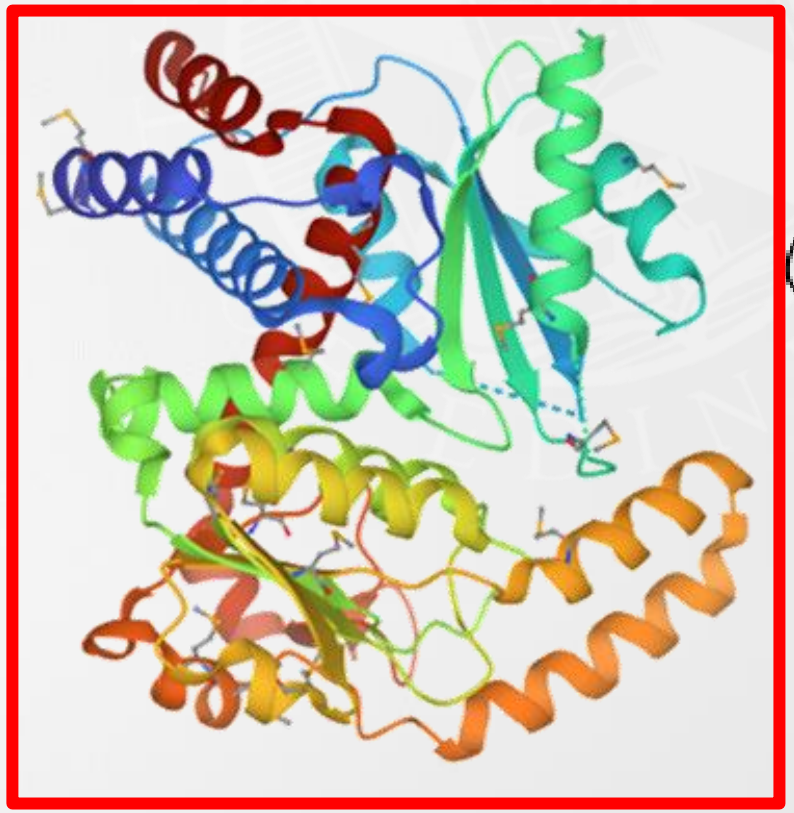






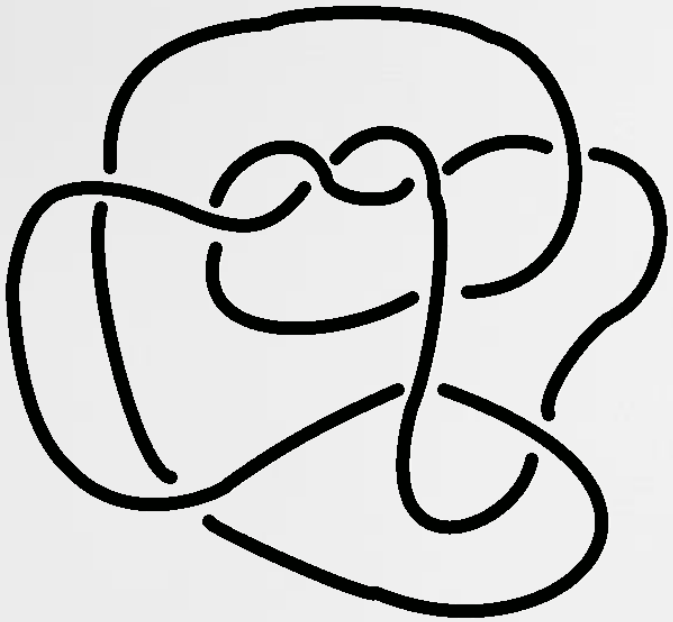


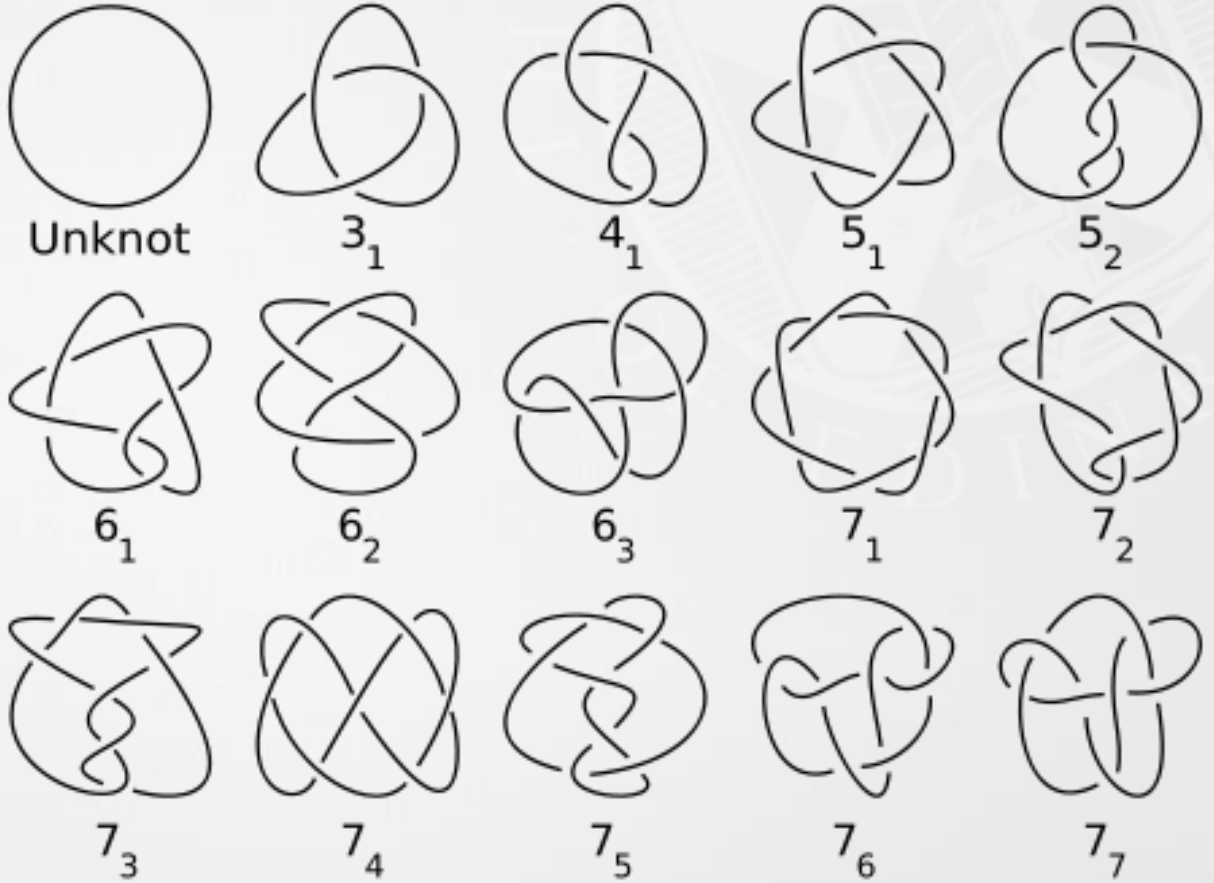
Structure of ygeW encoded protein from e. coli



Science 229, 171 (1985); copyright AAAS
Li, Y. et al. (2011). *Proteins*, *79*(7), 2327–2334.

Knot Theory



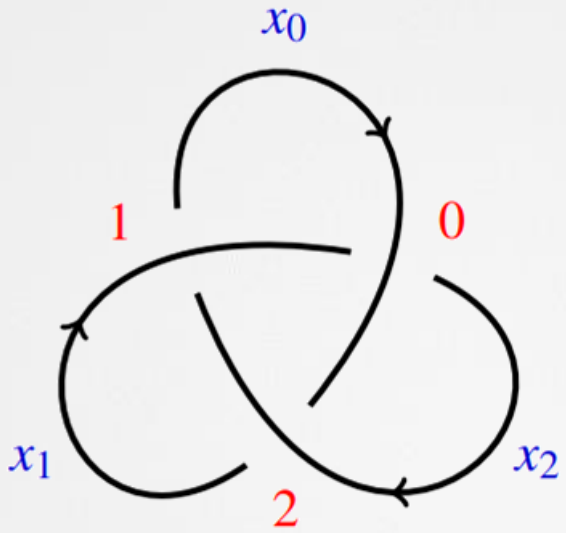


https://en.wikipedia.org/wiki/Knot_theory

Alexander Polynomial $\Delta(t)$

1.

2.

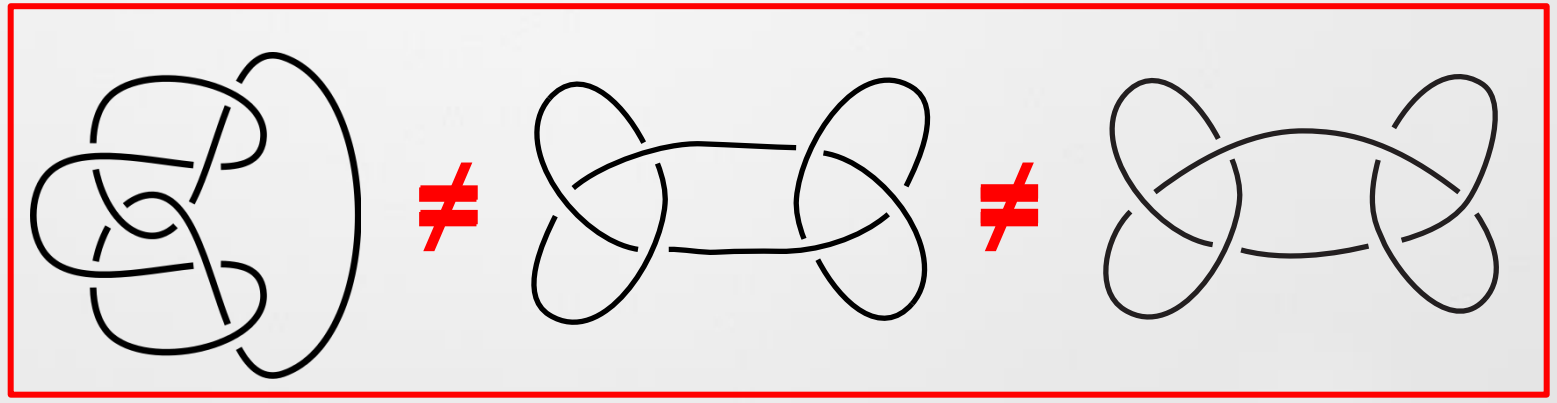


3.

$$\begin{pmatrix} 1-t & t & -1 \\ -1 & 1-t & t \\ t & -1 & 1-t \end{pmatrix}$$

4.

$$\Delta_{3_1}(t) = \det \begin{pmatrix} 1-t & t \\ -1 & 1-t \end{pmatrix} = (1-t)^2 + t = t^2 - t + 1$$



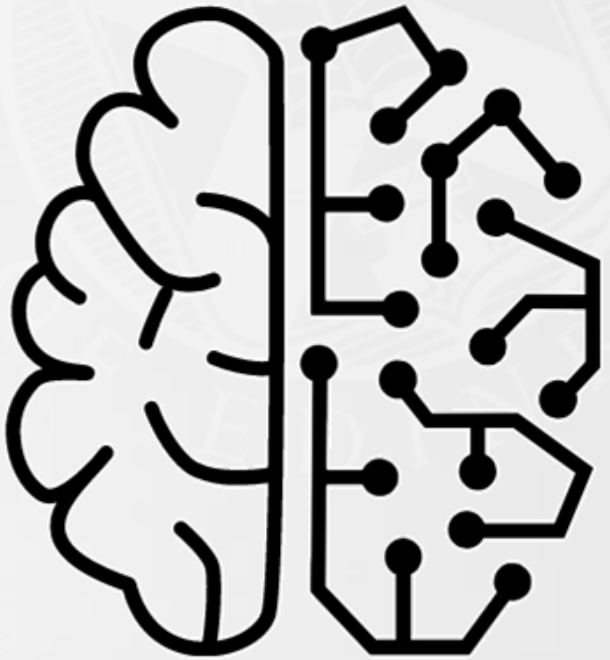
http://educ.jmu.edu/~taalmala/OJUPKT/lopez.pdf

Central Problem

There does not exist an algorithm that can uniquely identify different types of knots

Approach: Machine Learning

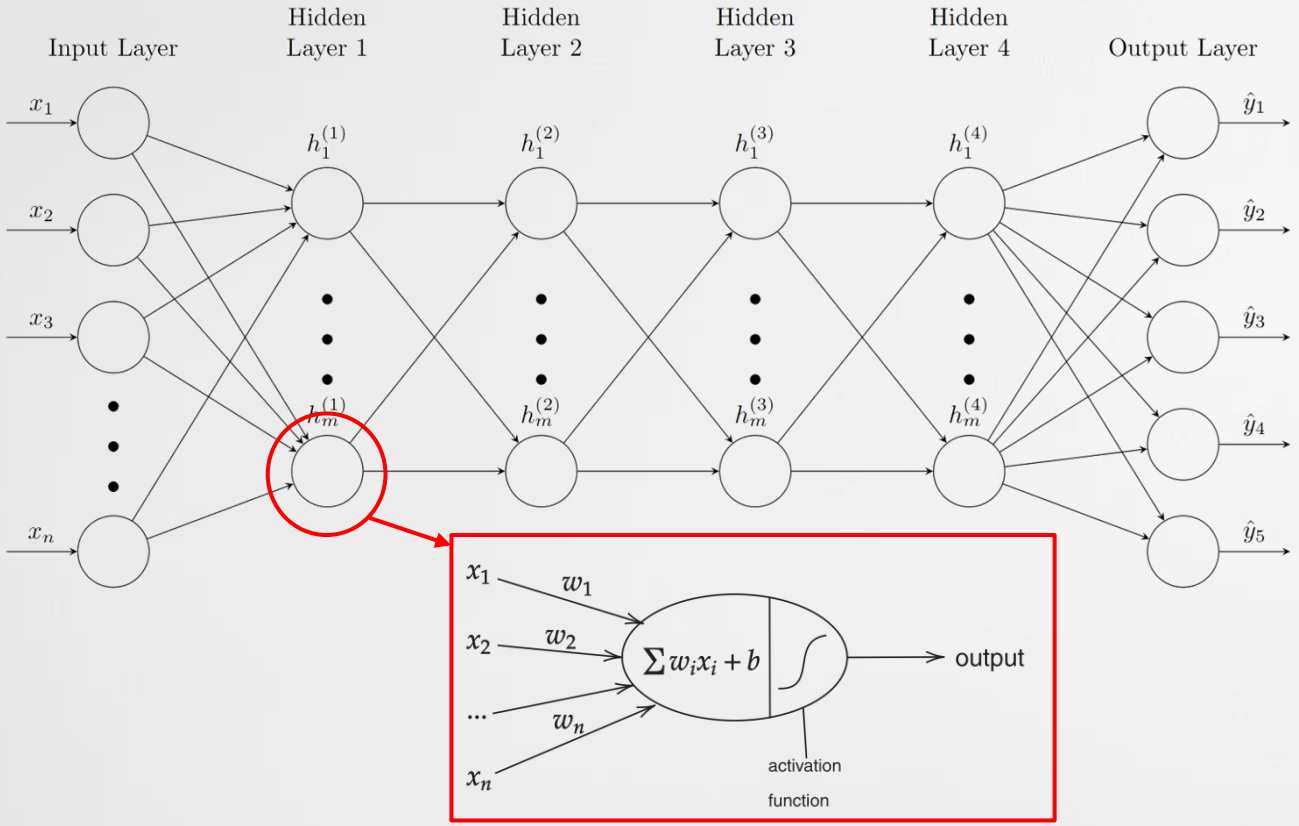
- The study and development of algorithms that learn from data in order to make predictions about new data



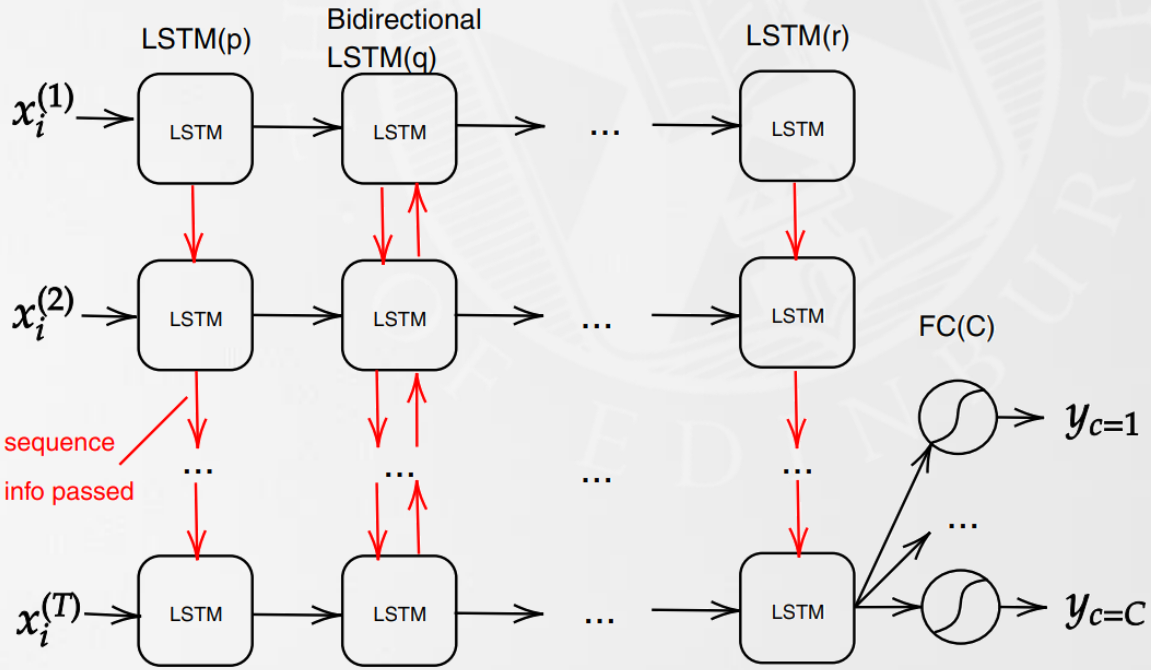
- **Supervised Learning**
  - Train model using labelled data
  - Classification or Regression tasks
- **Score Function**: Judges quality of fitted model
- **Optimisation Method**: Configure model parameters based on results of score function

Deep Learning

Feed-Forward Neural Network (FFNN)

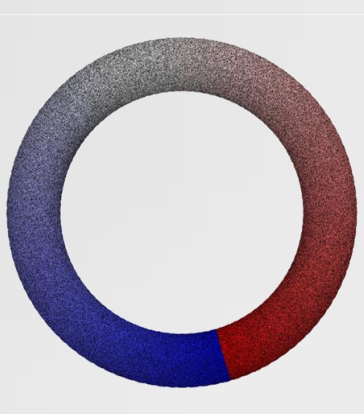


Recurrent Neural Network (RNN)



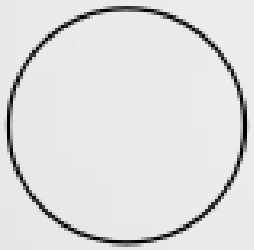
Vandans et al. (2020). *PRE*, 101(2), p.022502.

Knot Types



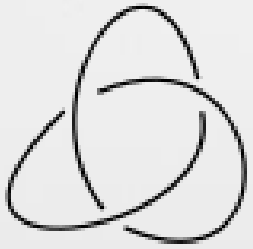
0_1

Unknot



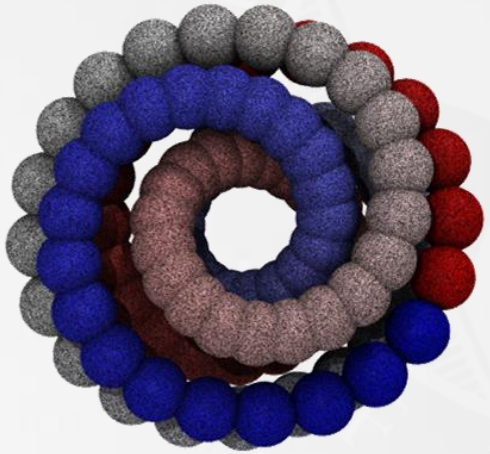
3_1

Trefoil Knot



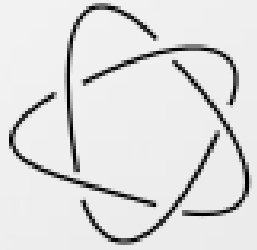
4_1

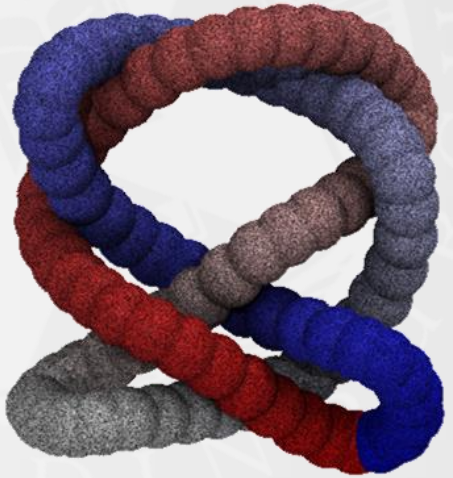
Figure-Eight Knot



5_1

Pentafoil Knot

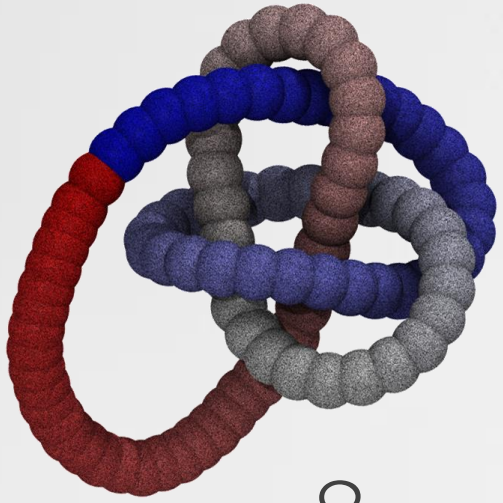




5_2

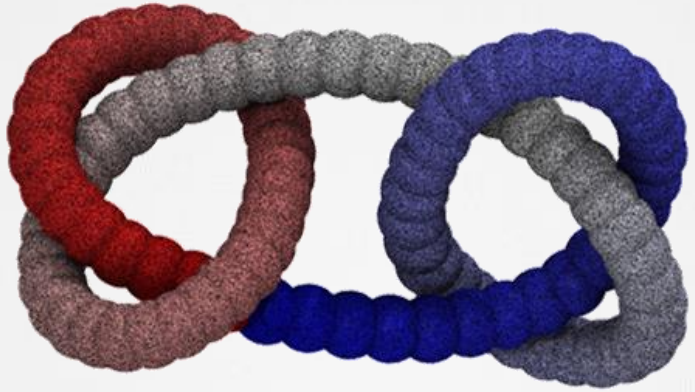
Three-twist Knot

Knot Equivalence



8_{20}

Pretzel Knot



$3_1 \# 3_1$

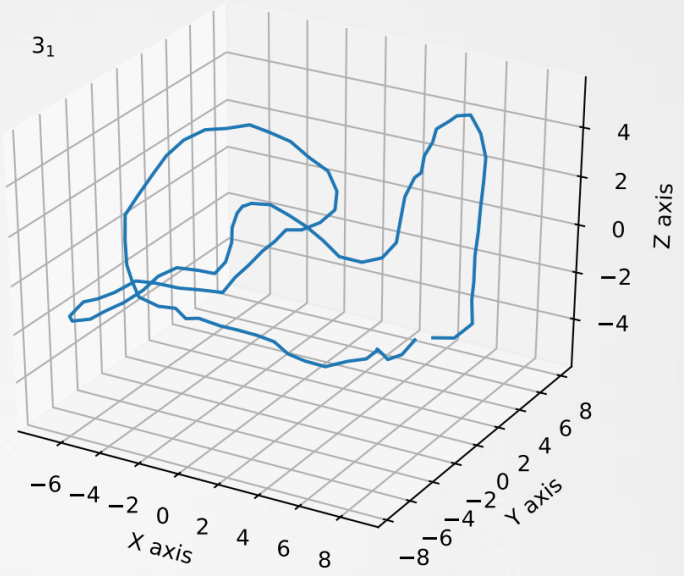
Granny Knot

$3_1 \# 3_1$

Square Knot

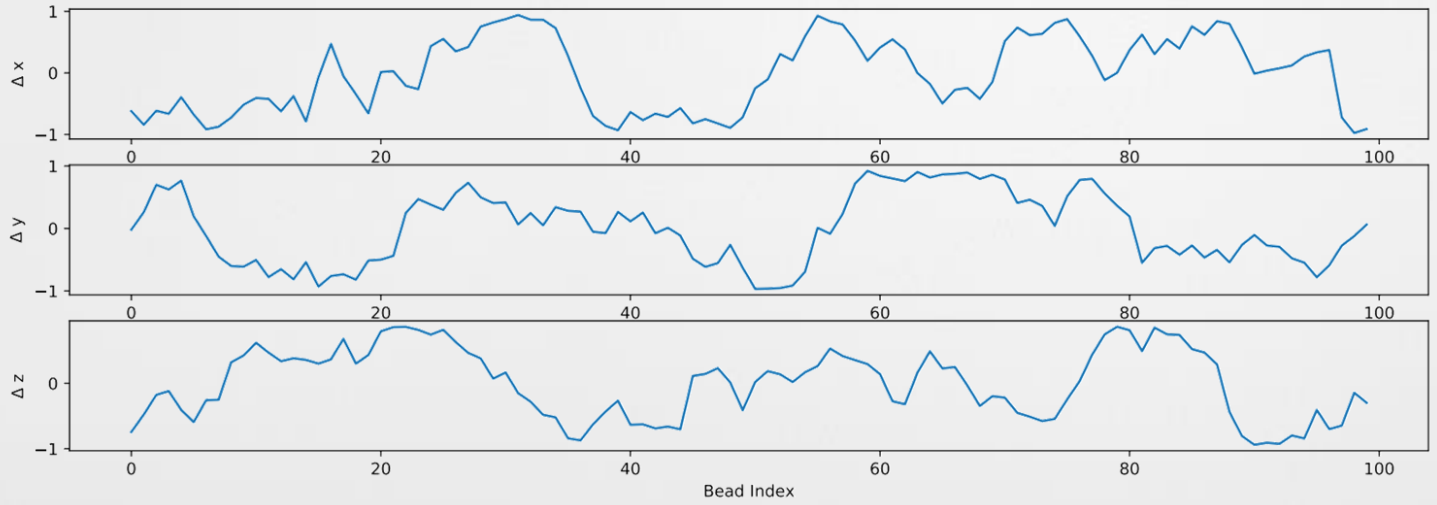
Knot Descriptors

XYZ Coordinates



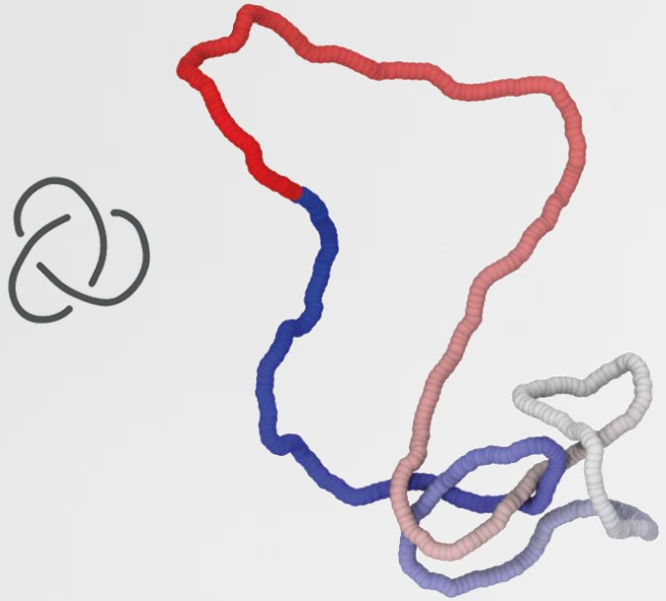
Global

Adjacent Bead Distances



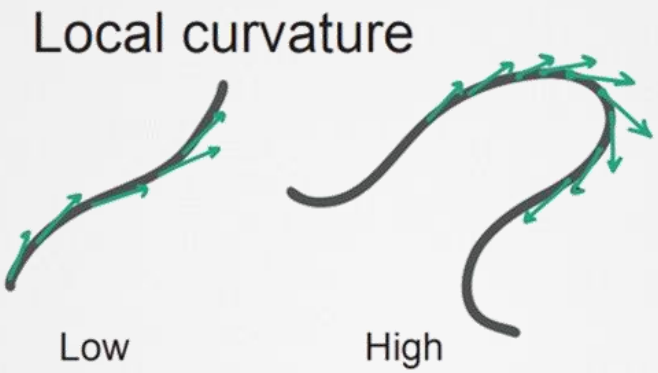
Local

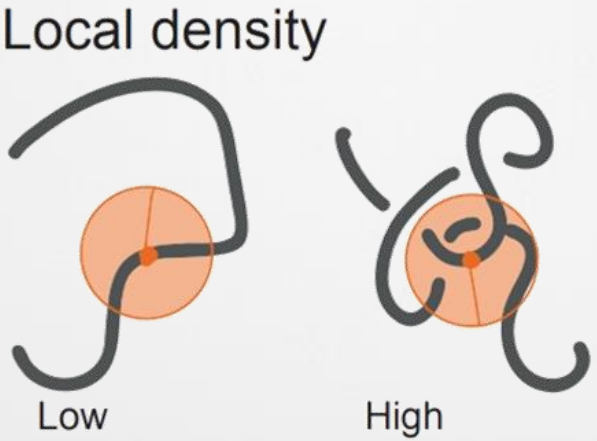
More Local Descriptors

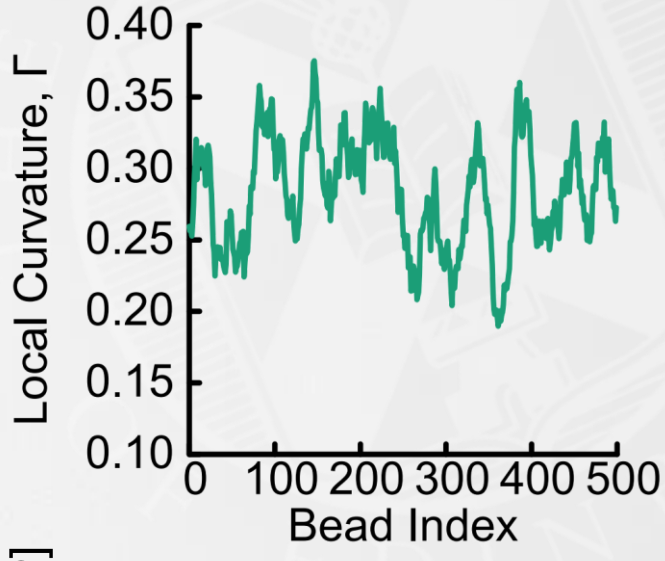


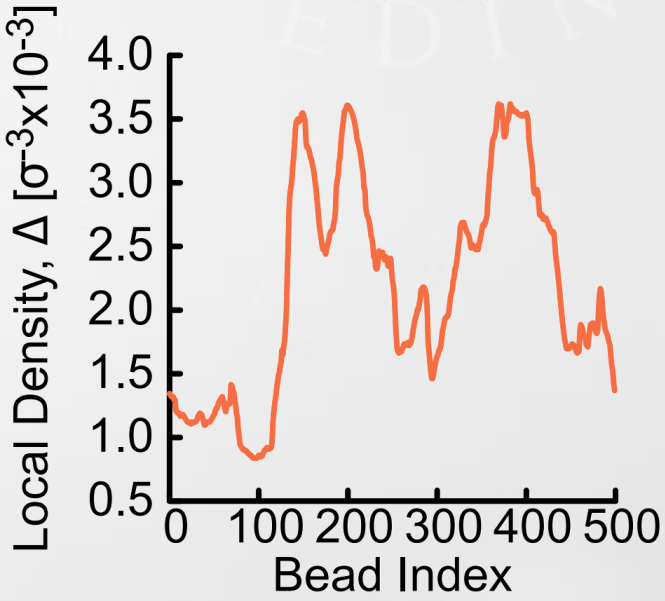
$$\Gamma_{LCi} = \arccos\left(\frac{\vec{t}_{i-1,i} \cdot \vec{t}_{i,i+1}}{|\vec{t}_{i-1,i}||\vec{t}_{i,i+1}|}\right)$$

$$\Delta(i) = \frac{1}{V_{thresh}} \sum_{j \neq i}^{N} \Theta\left(r_{thresh} - |\vec{r}_i - \vec{r}_j|\right)$$

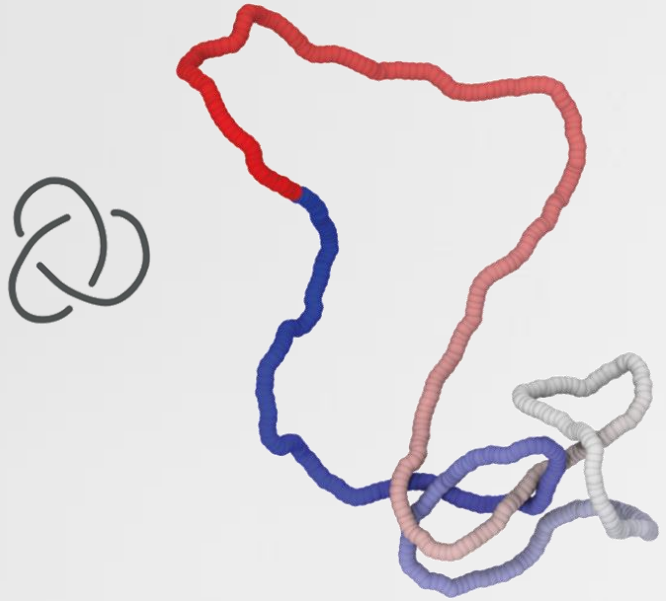


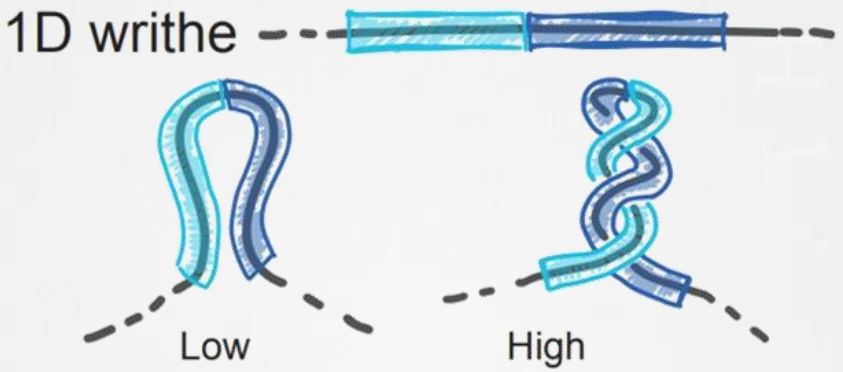


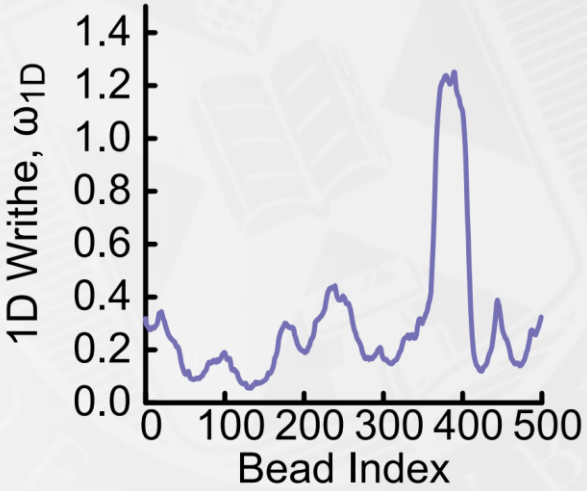


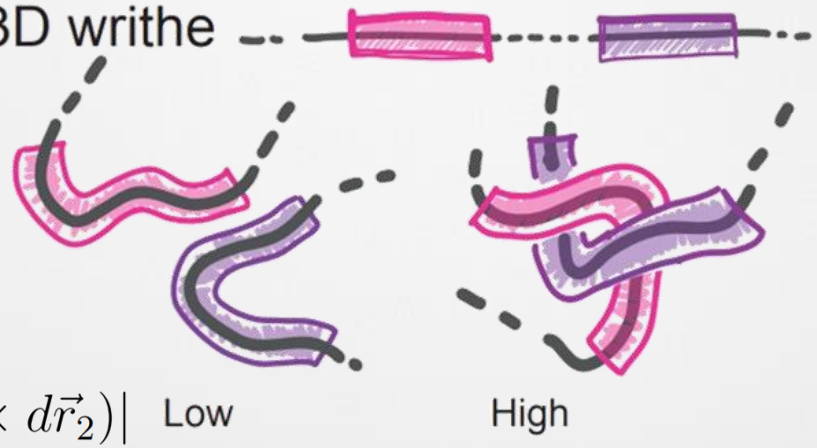


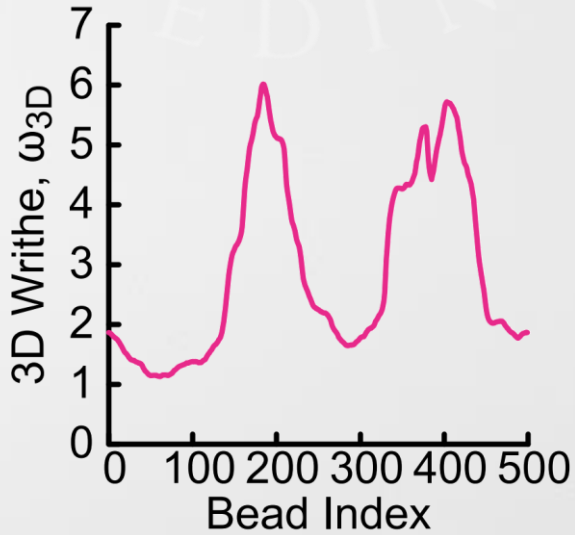
More Local Descriptors









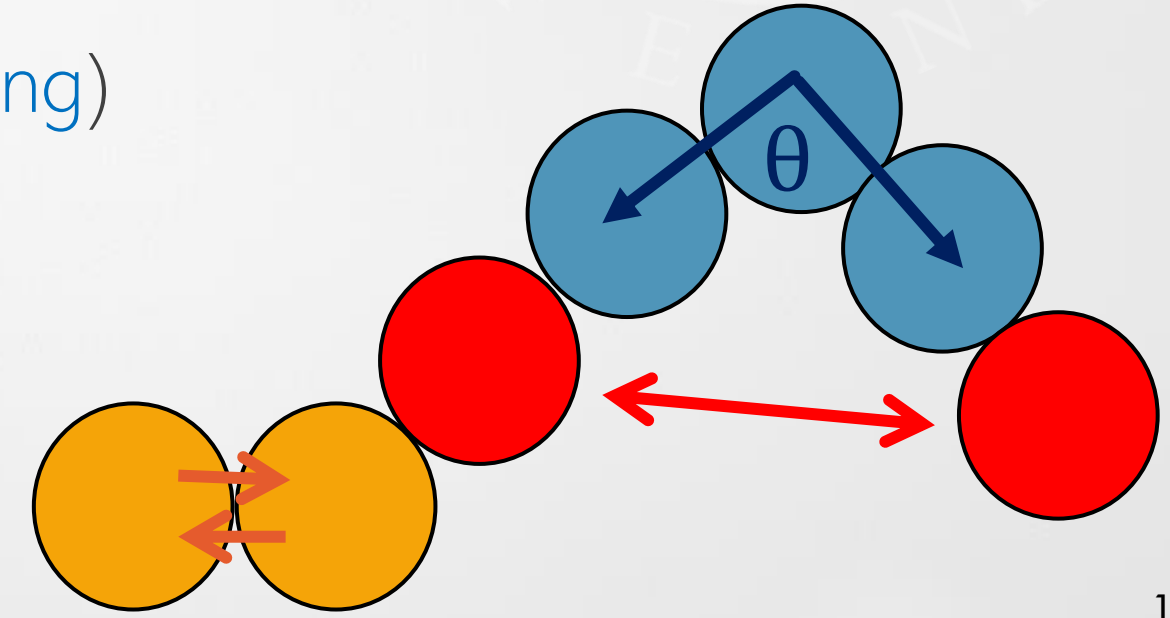


$$Wr_{1D}(k) = \frac{1}{4\pi} \int_{k-\frac{d}{2}}^{k} \int_{k}^{k+\frac{d}{2}} C \frac{|(\vec{r}_1 - \vec{r}_2) \cdot (d\vec{r}_1 \times d\vec{r}_2)|}{|\vec{r}_1 - \vec{r}_2|^3}$$

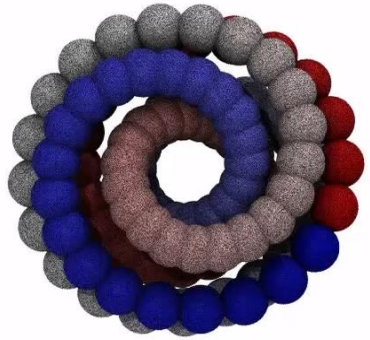
$$Wr_{3D}(k) = \frac{1}{4\pi} \int_{k-\frac{d}{2}}^{k+\frac{d}{2}} \int_{C} \frac{|(\vec{r}_1 - \vec{r}_2) \cdot (d\vec{r}_1 \times d\vec{r}_2)|}{|\vec{r}_1 - \vec{r}_2|^3}$$

Molecular Dynamics Simulations

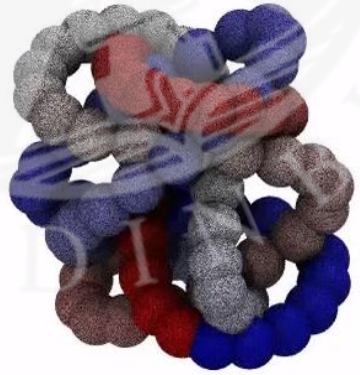
- Large Atomistic Many-body Molecular Physics Simulation (LAMMPS)
- Langevin Equation $m\vec{a}(t) = -\gamma\vec{v}(t) + \vec{\eta}(t)\sqrt{2\gamma k_B T} - \vec{\nabla}\phi(r)$
- Potentials (L-J, bead-spring, bending)
- $N_{beads} = 100$



Molecular Dynamics Simulations



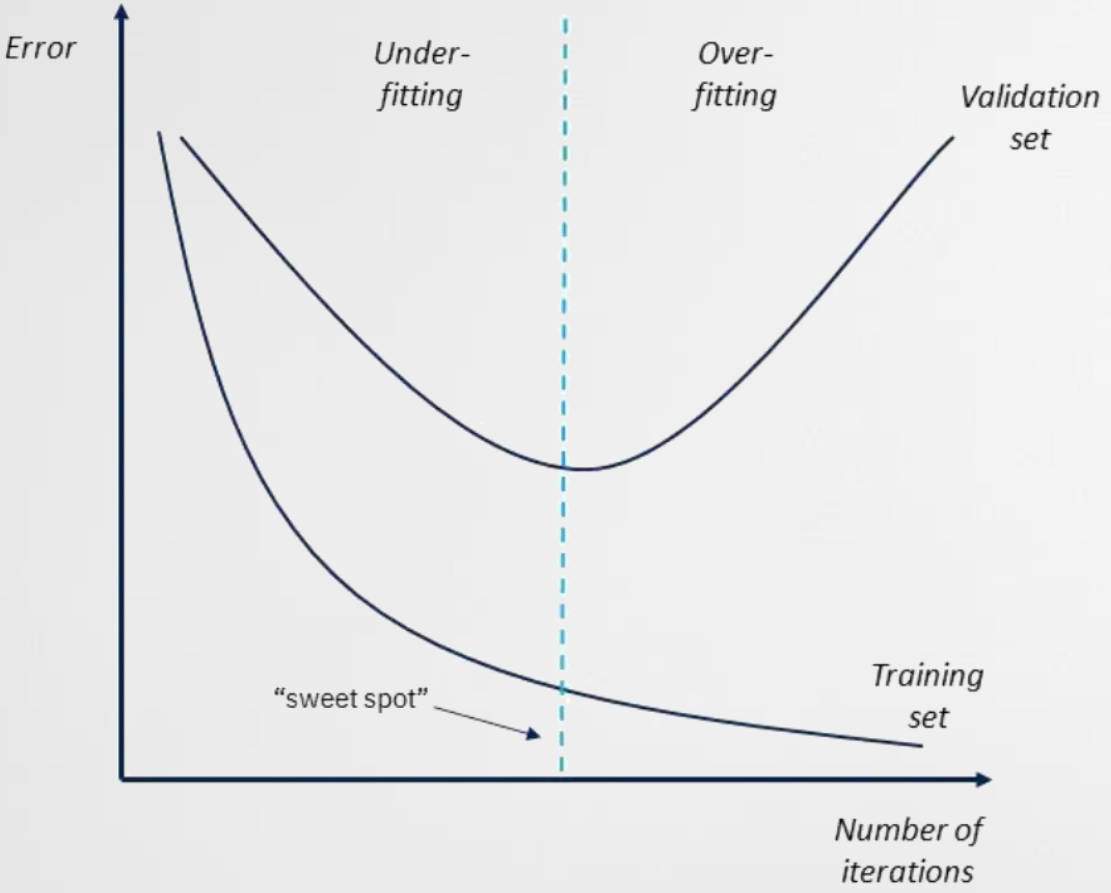
Molecular Dynamics Simulation

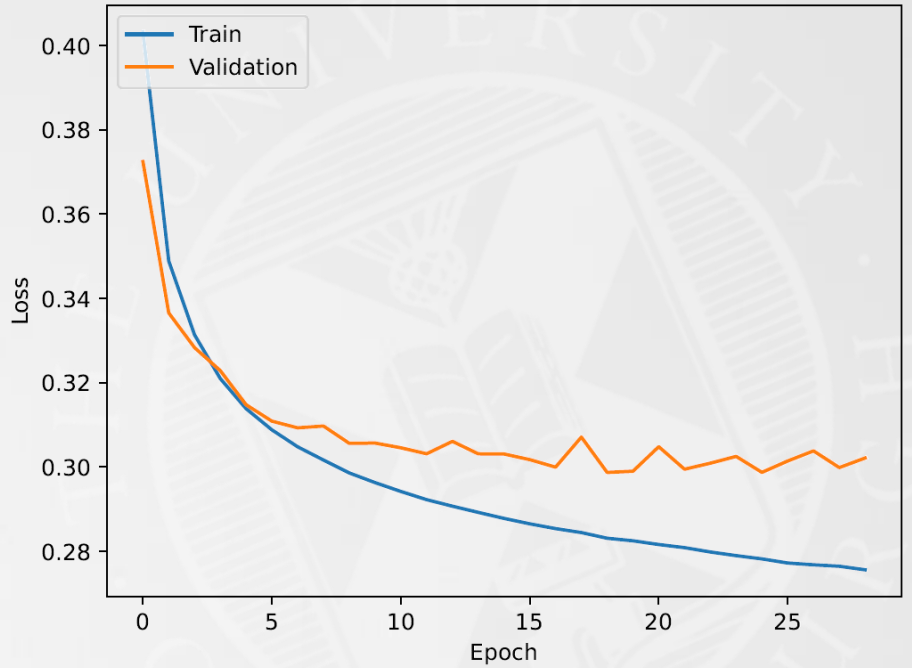


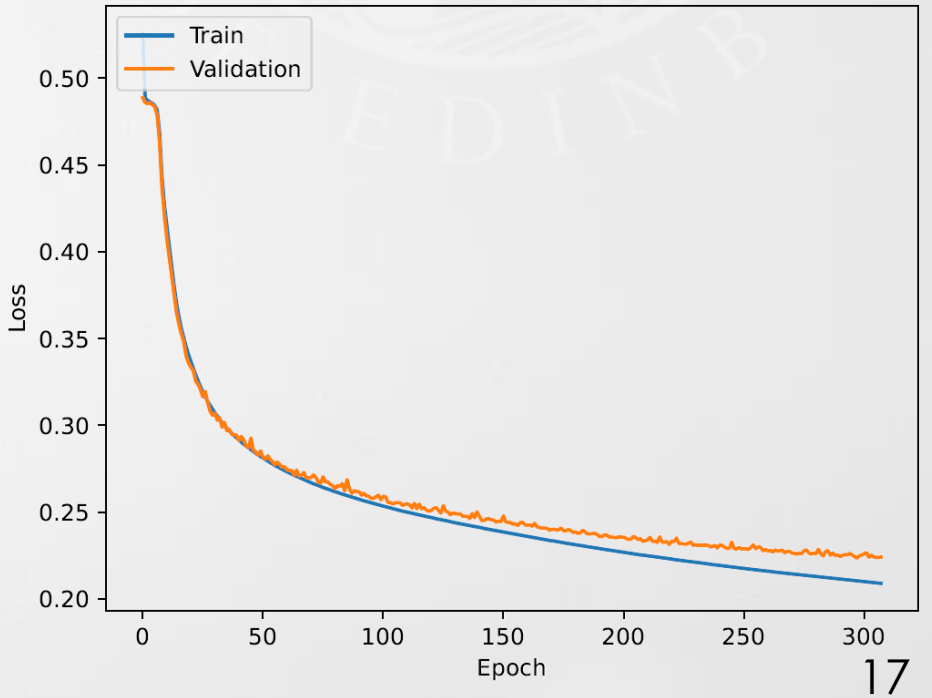
Training

Data	1,000,000 knots per type 100 or 300 dimensions each
Labels	Knot Type
Training-Validation-Testing Split	72-18-10

Early Stopping

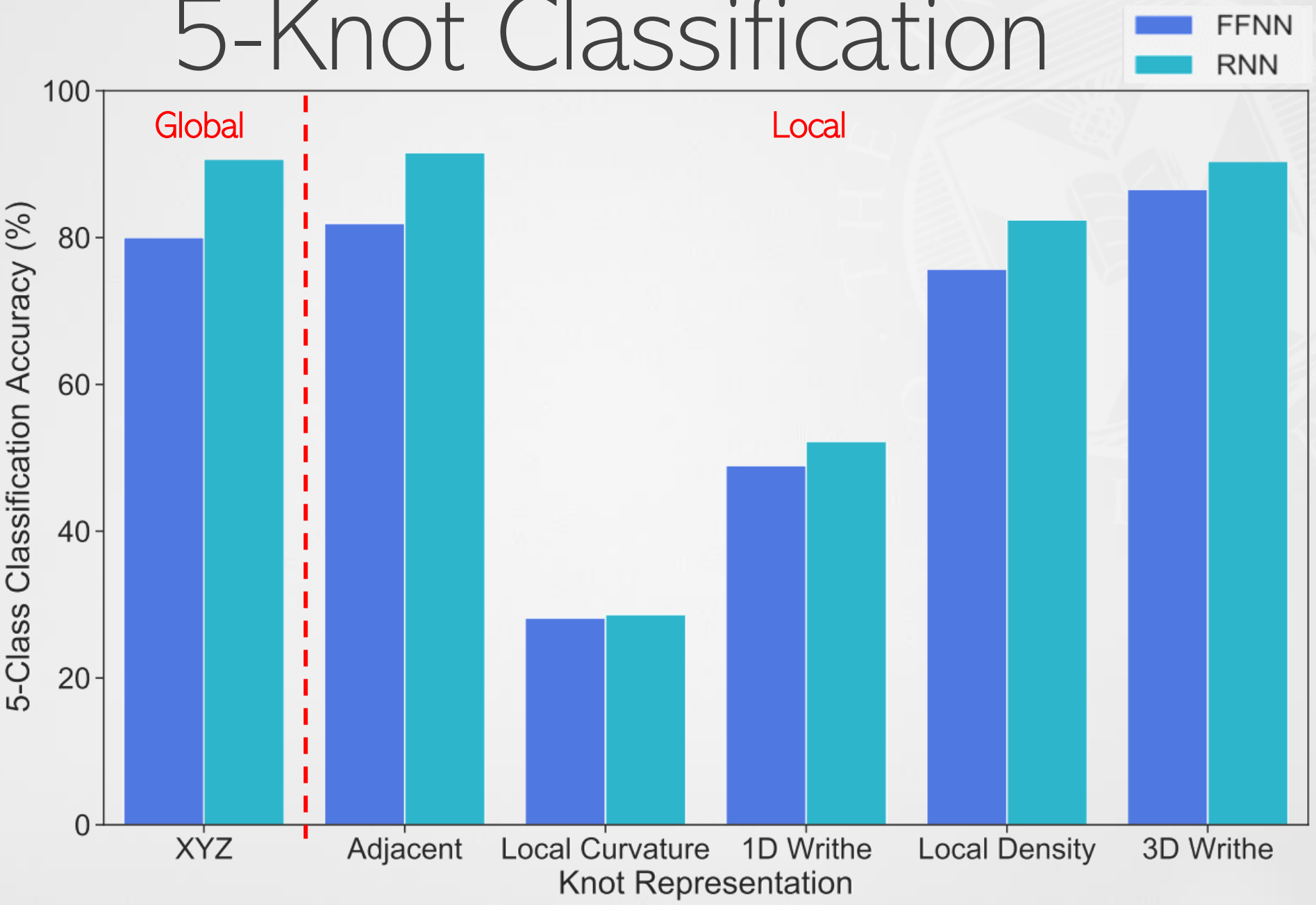






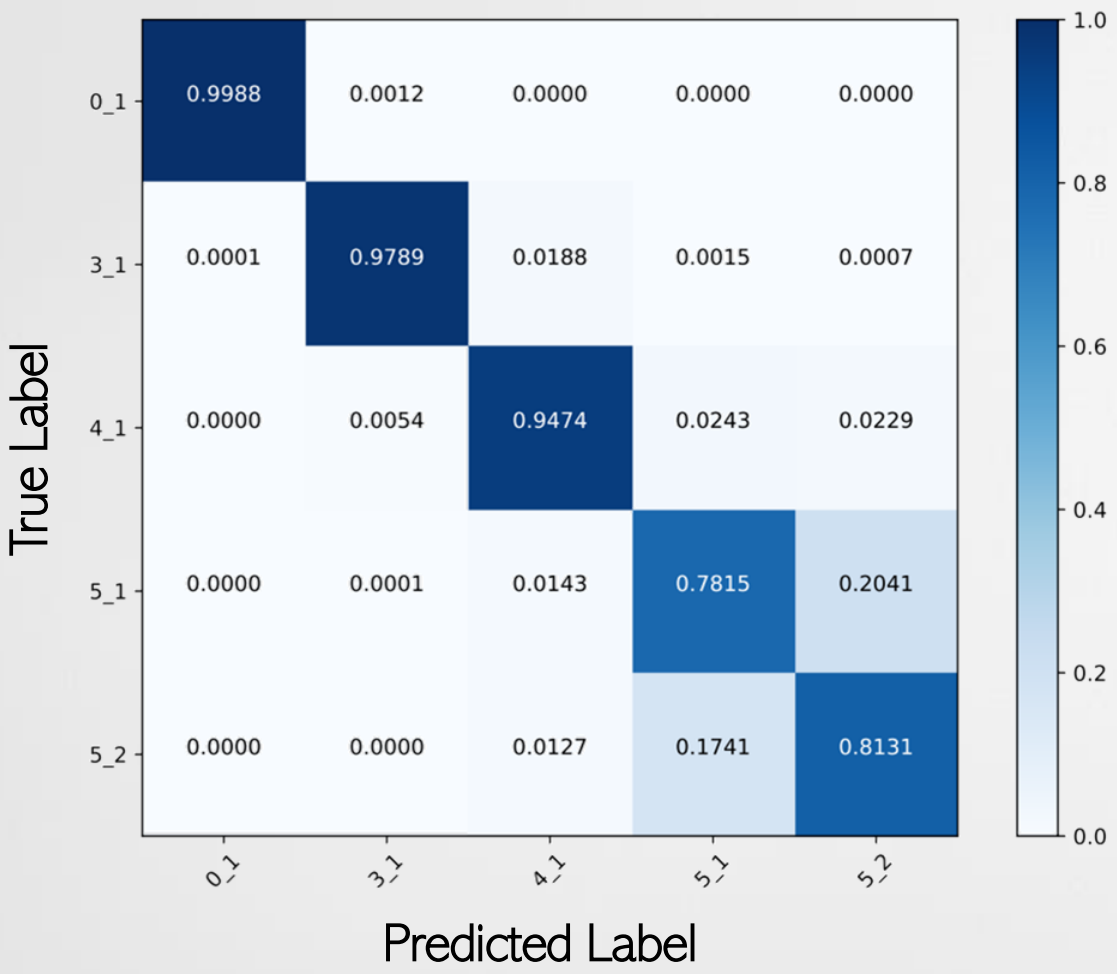
https://www.ibm.com/cloud/learn/overfitting

5-Knot Classification

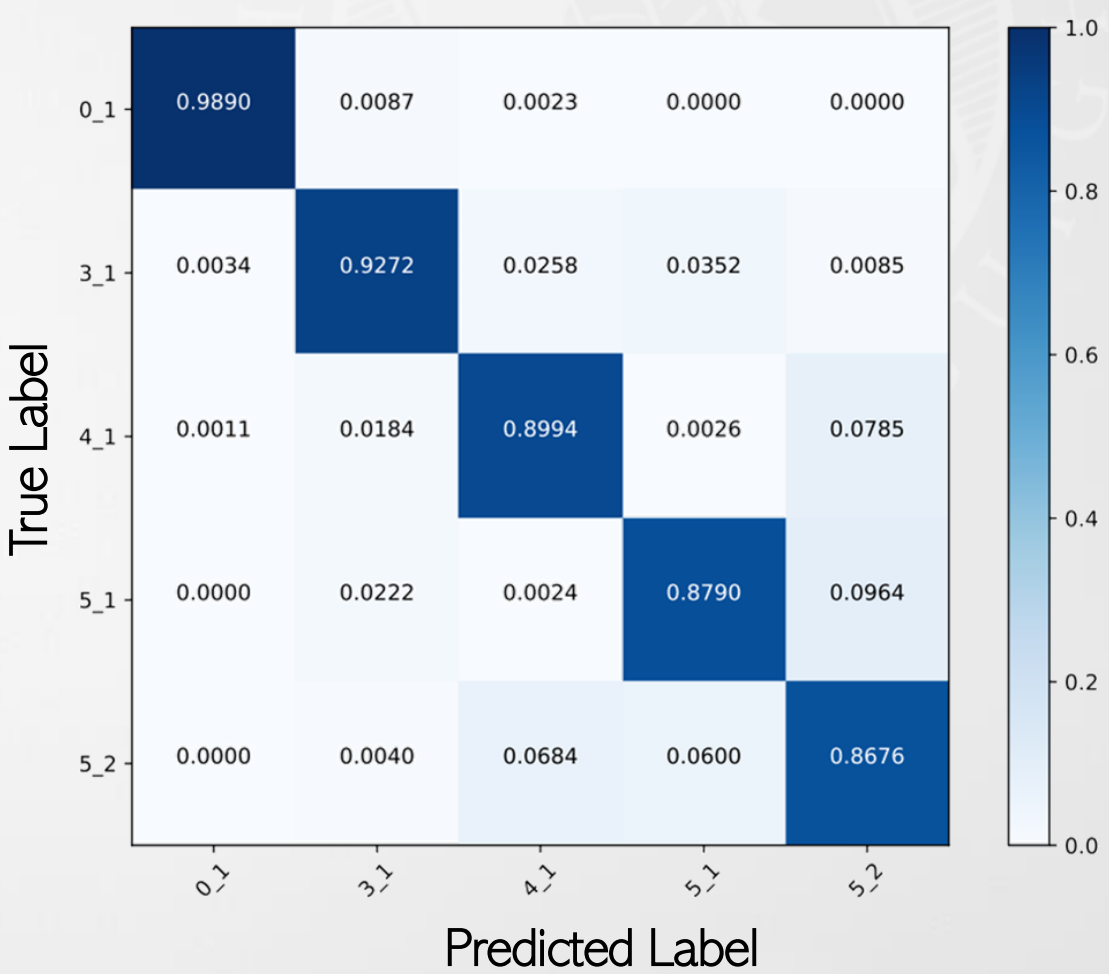


5-Knot Classification

RNN – Normalised 3D Writhe: 90.4%

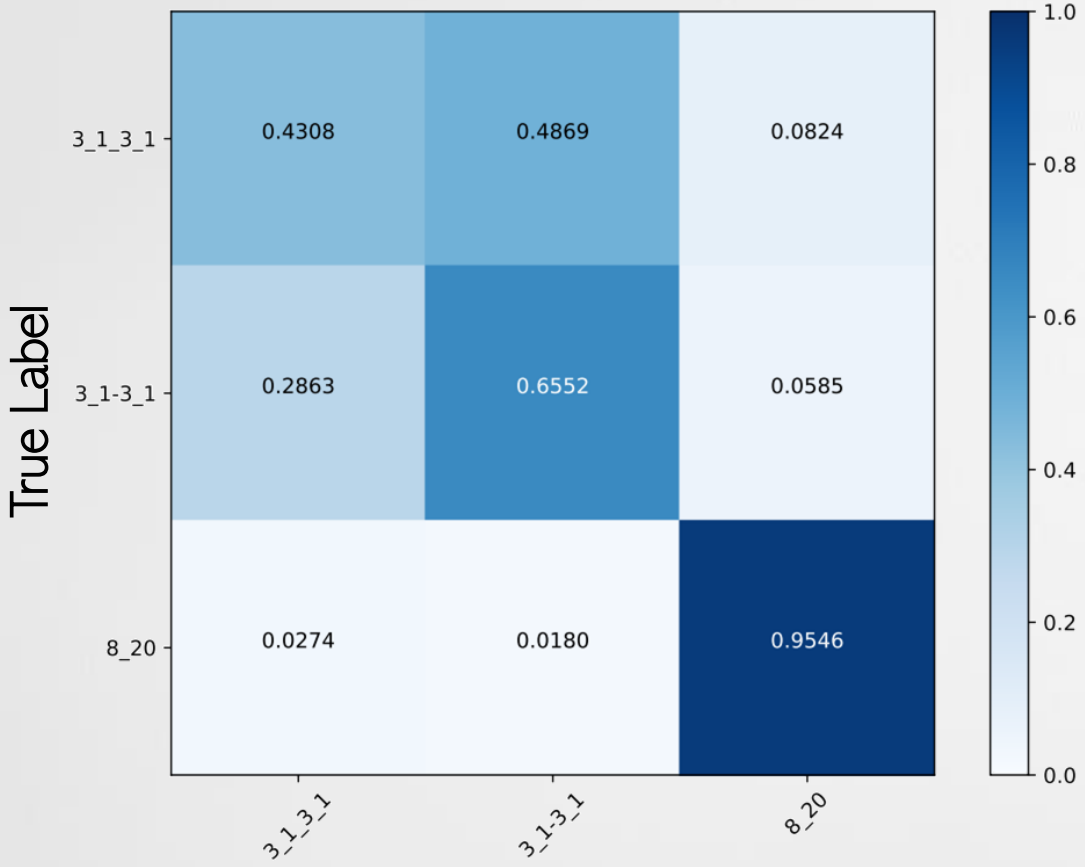


RNN – Normalised Adjacent: 91.2%

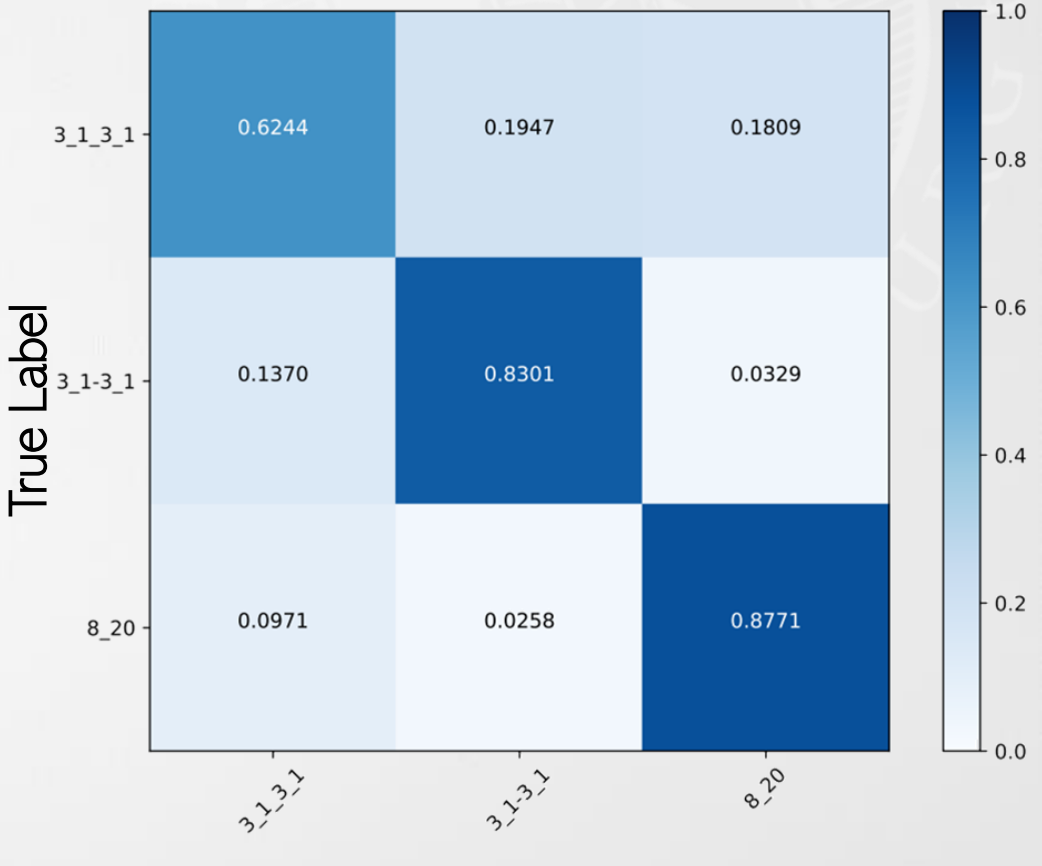


Knot Equivalence Classification

Normalised 3D Writhe: 68.0%



XYZ Coordinates: 77.7%



Conclusions & Future Work

- Machines can learn knot topology via geometric descriptors
- Sequentiality of the knot informs global topology
- Different-sized polymers (N_{beads} = 250, 500, 1000)
- Different Deep Learning models (Generative Adversarial Networks)
- Data fusion