CNN for the Classification and Grouping of Segmented Viruses

Sam Klauer, Bhavya Papudeshi, Susanna R. Grigson, Vijini Mallawaarachchi, Robert A. Edwards
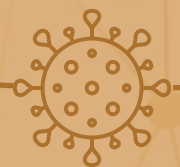
The first neural network that can distinguish between segmented and non-segmented viruses for greater viral understanding and disease prevention

Github + References

NCBI Virus
Sequences for discovery
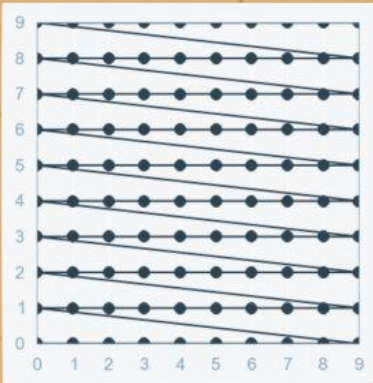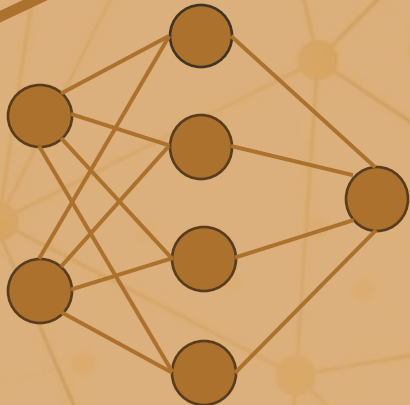
4725 segments

6569 non-segmented viruses

9035 training sequences
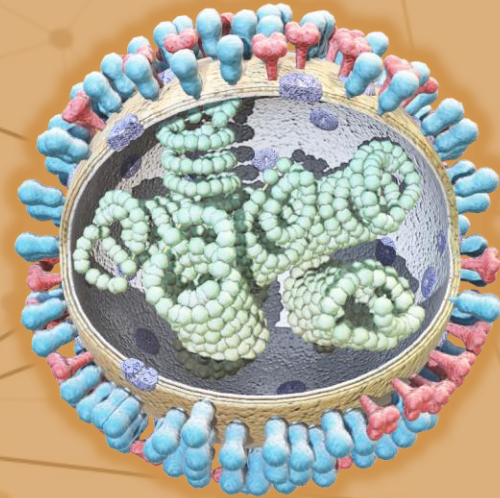
2259 testing sequences

Segmented versus Non-Segmented Viruses:
Over 600,000 unknown human threatening viruses
Influenza is the most well-known segmented virus and COVID is the most well-known non-segmented virus
Segmented viruses can evolve rapidly with different combinations of altered segments
The number and type of segments needed for infection vary across viruses
Unknown viruses can also play a critical role in microbial communities

	A	C	G	T
A	1	0	0	0
C	0	1	0	0
G	0	0	1	0
T	0	0	0	1

84%

| One-Hot Encoding | → | CNN Model Architecture | → | Extensive Grid Search | → | Segment Classification | → | Clique Detection for Segment Grouping | → | Segmented Virus Final Classification |