Minimum Complexity Echo State Networks For Genome and Sequence Analysis

Chris Neighbor

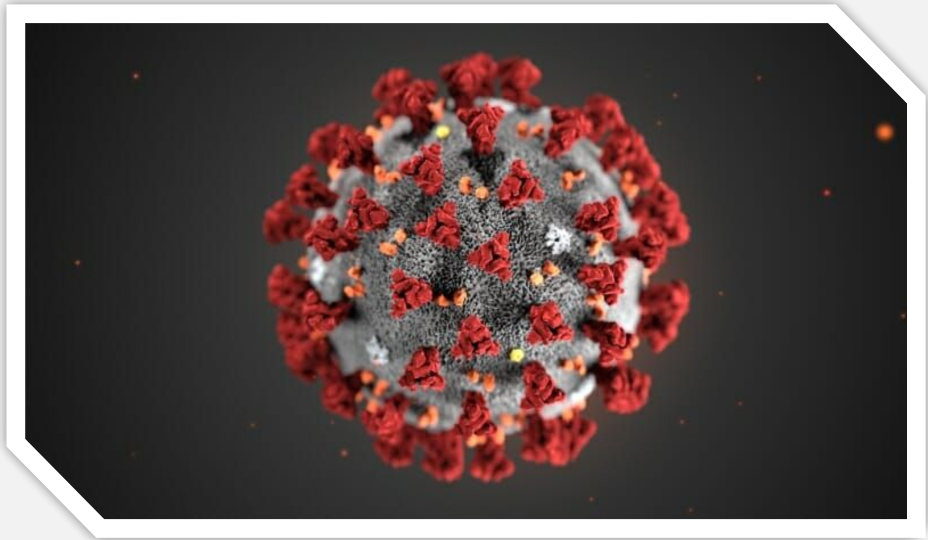
DR. CHRISTOF TEUSCHER
PORTLAND STATE UNIVERSITY
DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING (ECE)

This work was supported by the National Science Foundation under grant **no. 1518833**

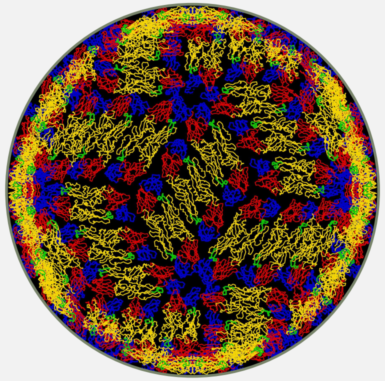
Introduction.

What's the problem?

- Increasing viral illnesses threatens global human health and welfare.
- Dengue virus infects 390 million people a year with reports increasing 6-fold from 2010 to 2016. Around 9,000 deaths per year. ***[WHO]***
- Developing tools which can directly detect the virus can aid in early detection.



The Coronavirus -- www.cdc.gov/coronavirus



https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue,

What's the solution?

DNA Computing. A tool implemented using DNA molecules which has promising applications in medical situations due to its biocompatibility and potential for the direct sensing of biological molecules.

How is it currently done?

Current Practices

- Require viral isolation and incubation for sensing and can still have an only **40%** detection rate. ***[Tang]***
- Other methods like Polymerase Chain Reaction ***(PCR)*** requires special equipment and trained technicians for detection.

→

DNA Computing

- Can perform sensing, computation, and classification in a self-contained system biocompatible with the body.
- Has potential for early detection before symptoms appear.

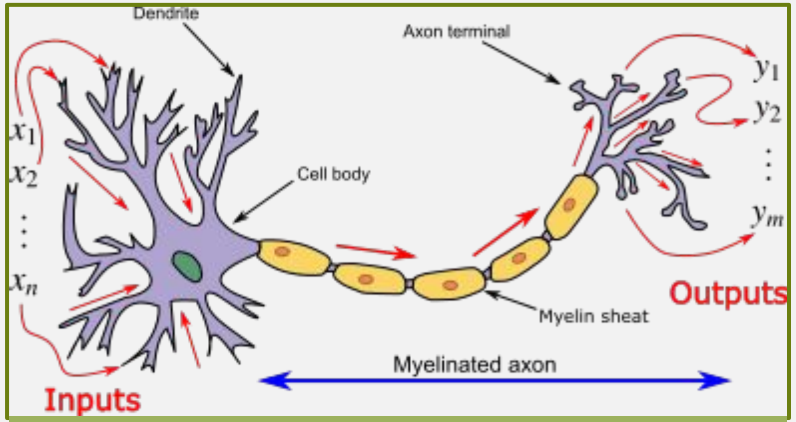
K. Tang et al, "Diagnosis of dengue: an update", Expert Review of Anti-infective Therapy, 2012

What are we trying to do?

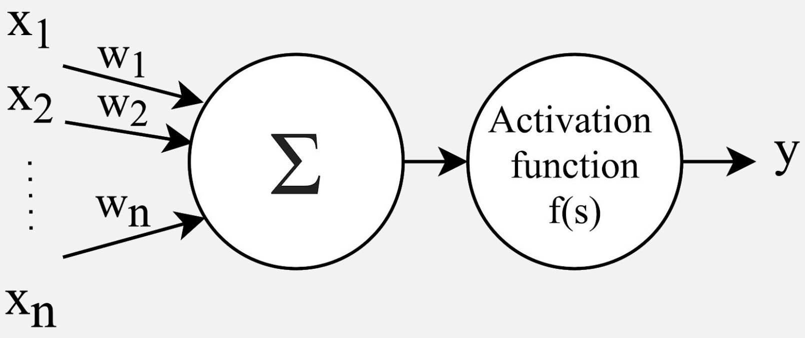
- Implement a machine learning architecture on DNA.
- Explore the use of dynamic randomized networks in order to accomplish the task of pathogen classification.
- With the aim that these kinds of randomized computing systems could be more easily implemented in DNA.
- Identify accuracy and complexity tradeoffs by exploring impact size has on network performance.

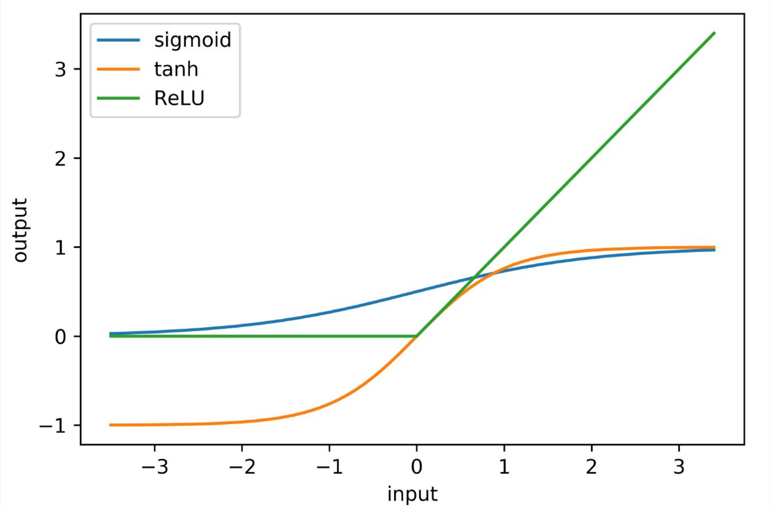
Neural Networks.

What is a “Neuron”?



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

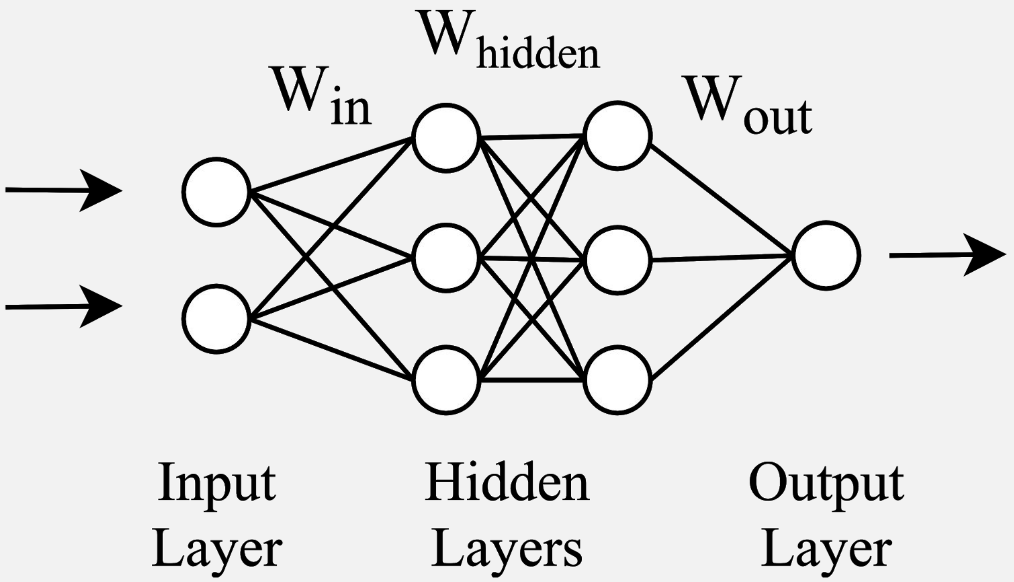




Activation Functions: we predominantly used ***tanh***

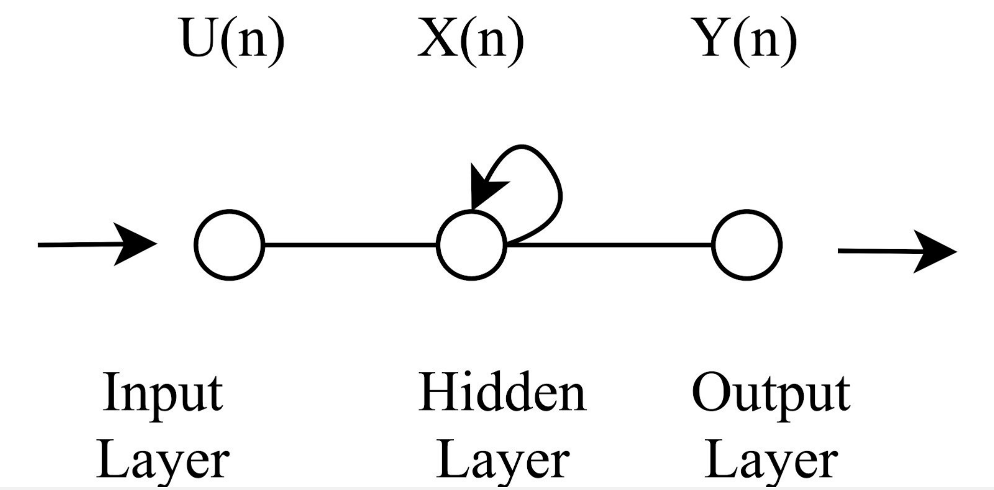
What is a “Neural Network”?

- Feed Forward multi-layered perceptron network, “Deep Learning”.
- The input layer is passed forward through the hidden layer then on to the output layer.
  - Each of the nodes within the hidden layer are a perceptron.
- Due to its multi-layered structure it is capable of solving non-linear separation problems.



What is a “Recurrent Neural Network”?

- By adding a feedback into the hidden layer, features across time can be extracted.
- In this model, memory and computation are *linked together* in the network.
- Other nodes can be used which can have a separate memory state.



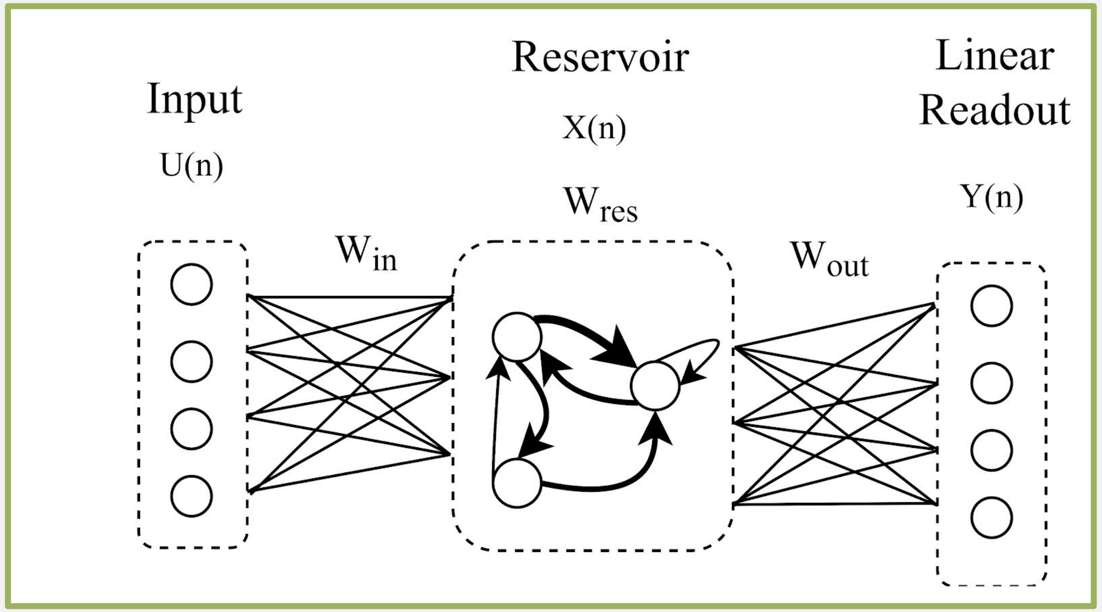
Reservoir Computing.

What is "Reservoir Computing"?

Three components:

- An input layer to the system which is randomly initialized.
- A fixed dynamic hidden layer known as the "reservoir" which extracts features.
  - Liquid State Machines: spiking nodes
  - Echo State Networks: simple recurrent nodes
  - Other dynamical systems and substrates, e.g. oscillators, self-assembled nanowires, and random chemical reaction network ***[Lukoševičius]***
- A final linear readout layer which can be optimized to match a target output.

- A **main advantage** of reservoir computing is the low learning complexity because only the output layer is trained.



$$X[n+1] = f(W_{in}U[n] + W_{res}X[n])$$
$$Y[n] = g(X[n]W_{out})$$

M. Lukoševičius, "A practical guide to applying echo state networks," in Neural Networks: Tricks of the Trade, 2012

Optimizing the Linear Readout Layer.

$$\underset{W}{argmin} \left\|XW - y\right\|^2 + \lambda \left\|W\right\|_2^2$$

$$W = (X^T X + \lambda I)^{-1} X^T y$$

Training the final readout layer: Ridge Regression

- **Classification:** Logistic Regression & Support Vector Machine.
  - Output probability of each class using softmax layer.
- Can use gradient descent to train the readout layer.

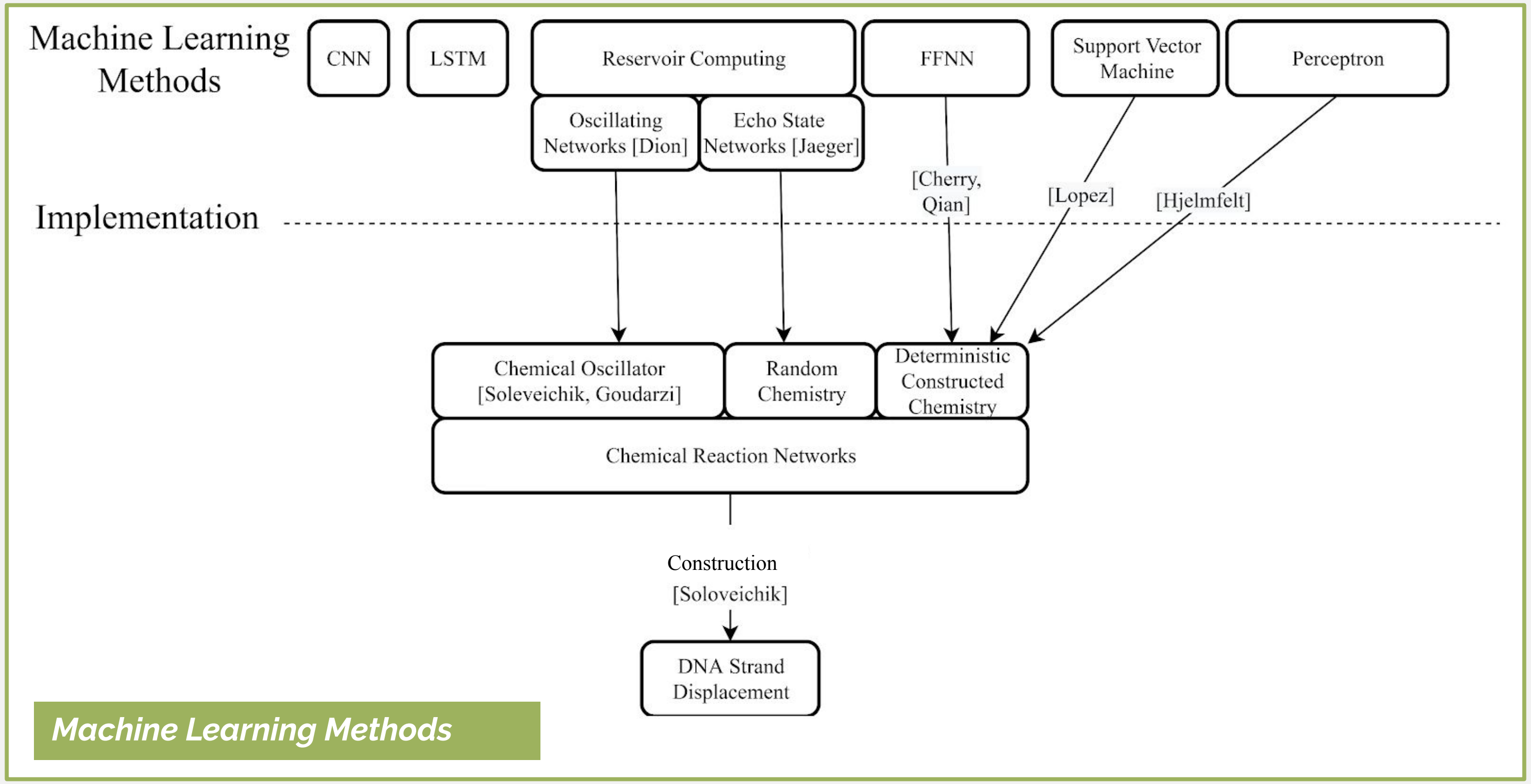
Why Reservoir Computing?

- Molecular circuits can be difficult to scale due to:
  - Cross-talk *(network pathway signal interference).*
  - Errors from manufacturing of DNA strands which require purification.
  - Requirement of orthogonality of the strands. ***[Wang]***
- The reservoir layer is unstructured allowing for easier biochemical implementation.
- Fully or partially random networks of chemical circuits with a trained linear output layer can be used for complex tasks such as pathogen detection.

F. Wang, et al, “Implementing digital computing with DNA-based switching circuits,” Nature Communications , 2020

Bridging the Gap:

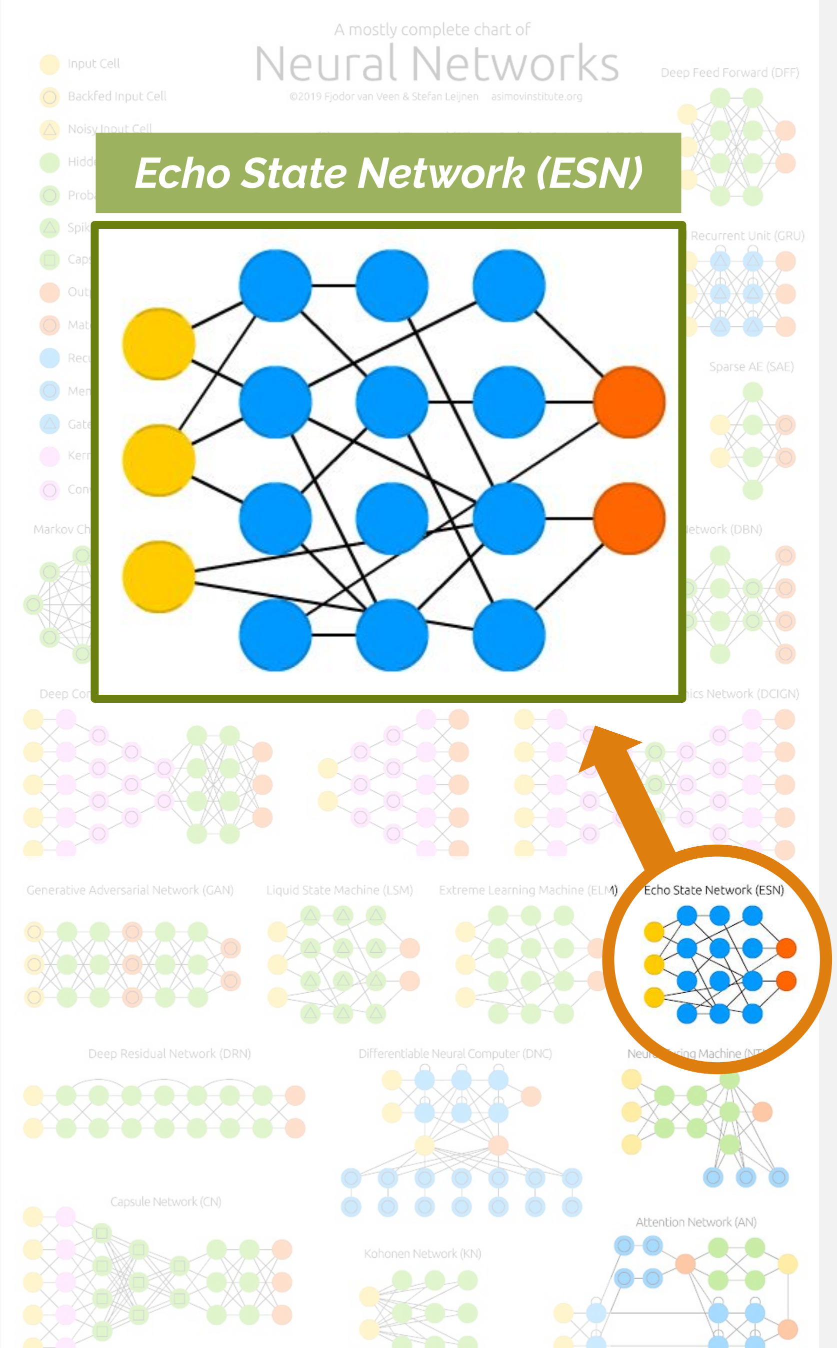
Reservoir Computing & Chemical Implementation.



G. Dion et al, "Reservoir computing with a single delay-coupled non-linear mechanical oscillator," Journal of Applied Physics, 2018
H. Jaeger, "The echo state approach to analysing and training recurrent neural networks-with an erratum note," GMD Technical Report, 2001
K. M. Cherry et al, "Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks," Nature, 2018
L. Qian et al, "Neural network computation with DNA strand displacement cascades," Nature, 2011.
A. Hjelmfelt et al, "Chemical implementation of neural networks and turing machines," Proceedings of the National Academy of Sciences, 1991.
R. Lopez et al, "A molecular multi-gene classifier for disease diagnostics," Nature Chemistry, 2018.
A. Goudarzi et al, "DNA reservoir computing: a novel molecular computing approach," in International Workshop on DNA-Based Computers, 2013
W. Yahiro et al, "A reservoir computing approach for molecular computing," in Artificial Life Conference Proceedings, 2018
D. Soloveichik et al, "DNA as a universal substrate for chemical kinetics, Proceedings of the National Academy of Sciences, 2010.

What is an “Echo State Network”?

- From this variety of Neural Networks, we zoom in on one particular architecture.
- ***Echo State Networks*** (ESN), the hidden layer consists of randomized recurrent neural networks with a trained output layer. The ESN is a subset of reservoir computing.



https://www.asimovinstitute.org/neural-network-zoo/

ESN – Initialization Property.

- Randomness allows for less precise constraints for implementation as the scale of these networks is increased.
- Readout complexity increases linearly as network size increases.

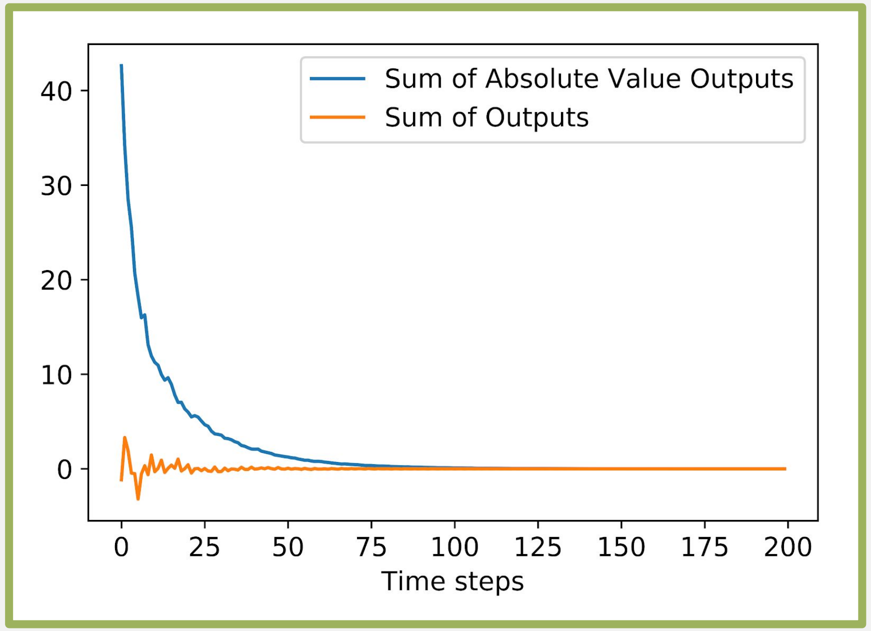
Echo State Property Initialization Process

- Initialize the weight matrix from Uniform[-1,1].
- Divide by the absolute value of the largest eigenvalue.
- Scale by the desired spectral radius value <1.

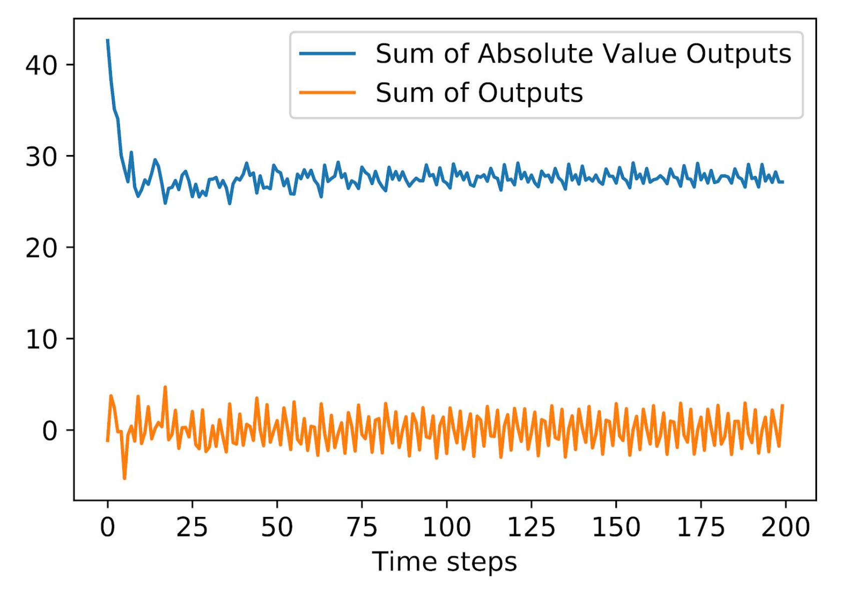
ESN – Effect of Spectral Radius.

The spectral radius is the largest eigenvalue of the hidden weight matrix.

Input: $\delta[n]$, (1 at n=0 and 0 afterwards), Nodes=100



Spectral radius = 0.95

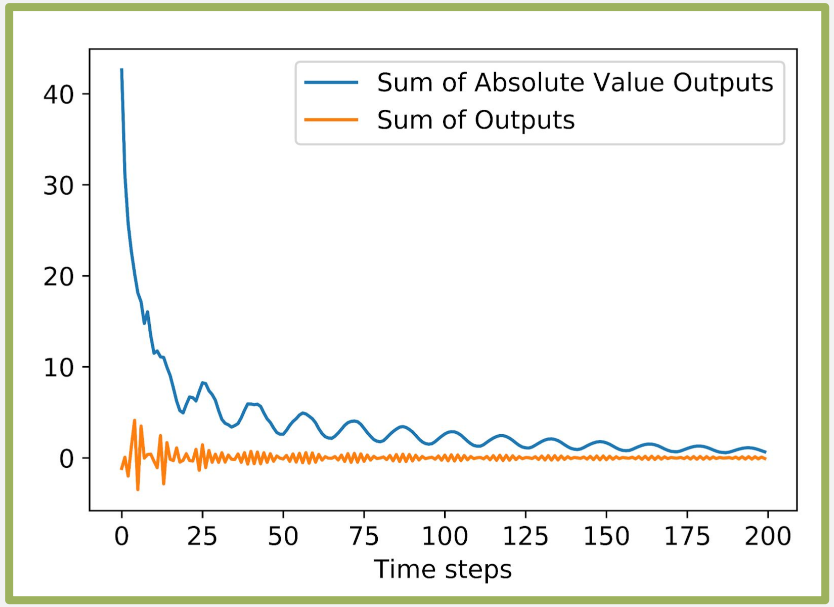


Spectral radius = 1.1

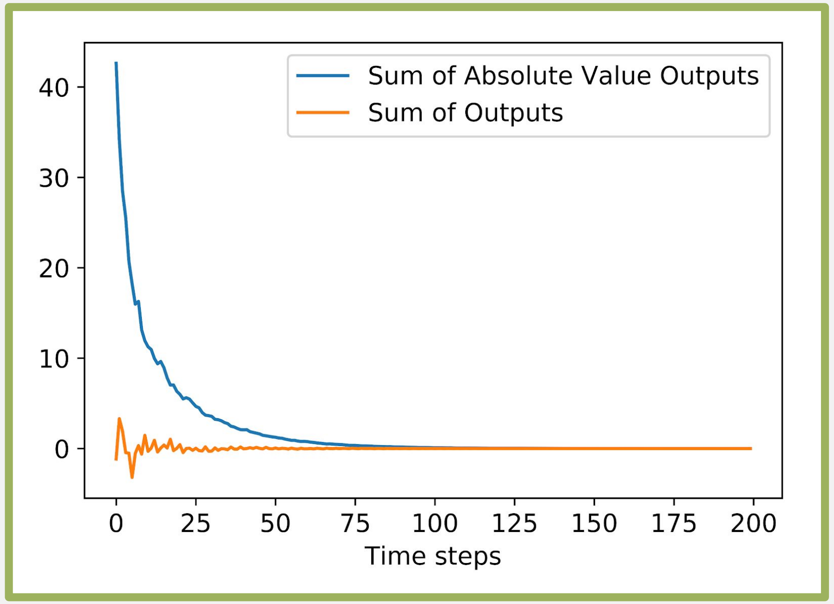
ESN – Effect of Sparsity.

Local networks can form with oscillations between them. ***[Lukoševičius]***

Input: $\delta[n]$ N=100



Sparsity = 0.1

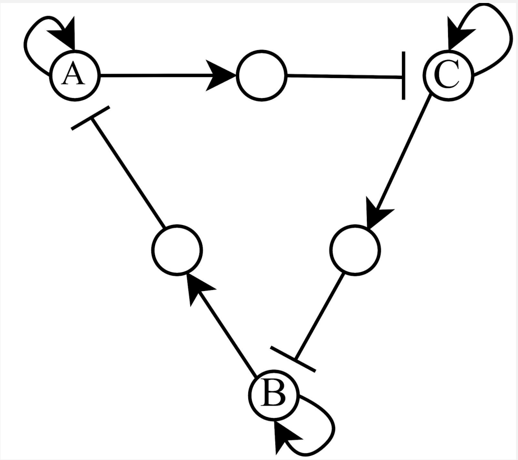


Fully Connected, Sparsity = 1.0

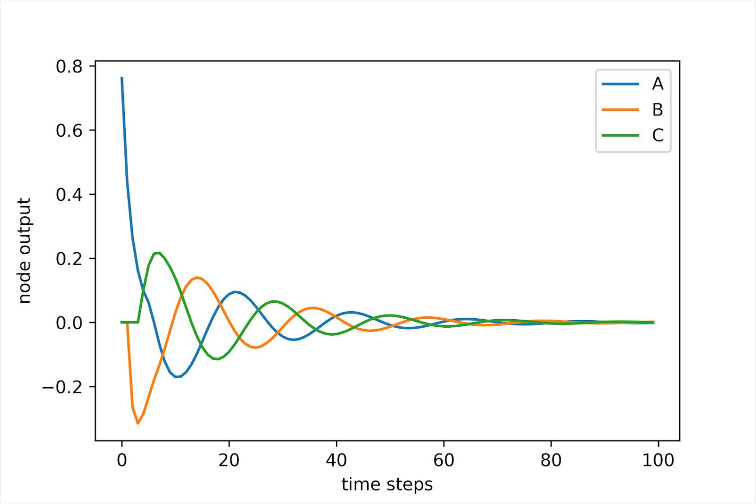
M. Lukoševičius, “A practical guide to applying echo state networks,” in Neural Networks: Tricks of the Trade, pp. 659–686, Springer, 2012.

Oscillating Neural Networks.

Oscillating Neural Network



Based on three DNA inhibitor circuits with a product and substrate reaction. ***[Goudarzi, Yahiro]***



- Direct conversions between weights and reaction pathways.
- Chemical Reaction Concepts:
  - **Autocatalysis:** Feedback to its own reaction,
  - **Inhibition:** Negative feedback to other reactions,
  - **Activation:** Positive feedback to reactions

A. Goudarzi et al,, "DNA reservoir computing: a novel molecular computing approach," in International Workshop on DNA-Based Computers, 2013
W. Yahiro et al, "A reservoir computing approach for molecular computing," in Artificial Life Conference Proceedings, 2018

Why ONN?

- Examining impact of size and spectral radius on the computational capabilities of an oscillating system, we can translate these insights into chemical oscillator scale, rate of reactions, and efflux rate.
- Implement in chemistry as it has a direct one to one conversion capability.
- These ONNs can have qualitatively similar outputs to chemical oscillators once equilibrium is shifted and normalized.

Tasks.

- NARMA-Task A
- ***NARMA-10***
- MNIST
- Splice Junction Gene Sequences
- ***Dengue Virus Genome Serotypes***

For others, please see Thesis.

Results.

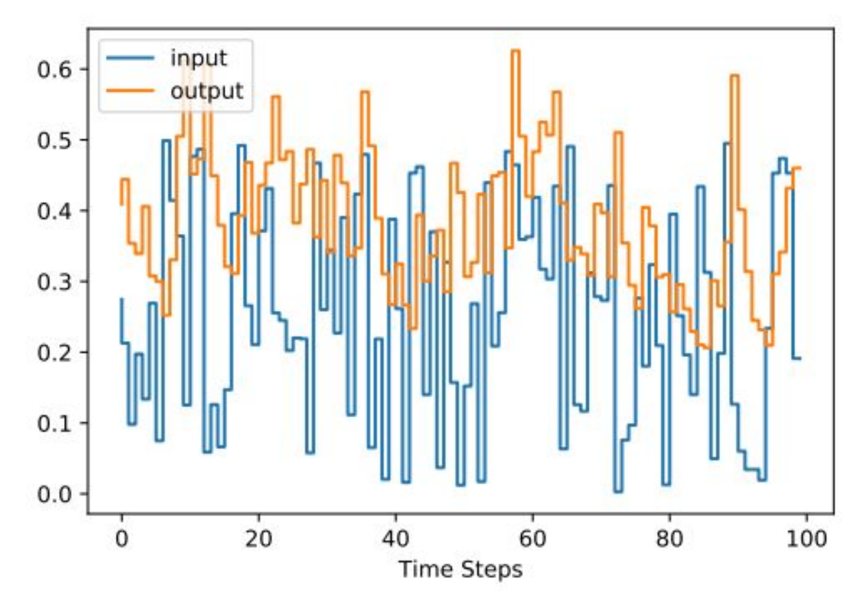
What is “NARMA”?

NARMA-10 - The **N**onlinear **A**uto**r**egressive **M**oving **A**verage has long term non-linear dependencies and is generated using the equation:

$$y(t) = \alpha y(t-1) + \beta y(t-1) \sum_{i=1}^{n} y(t-i) + \gamma u(t-n)u(t-1) + \delta$$

γ = 1.5
δ = 0.1
n = 10

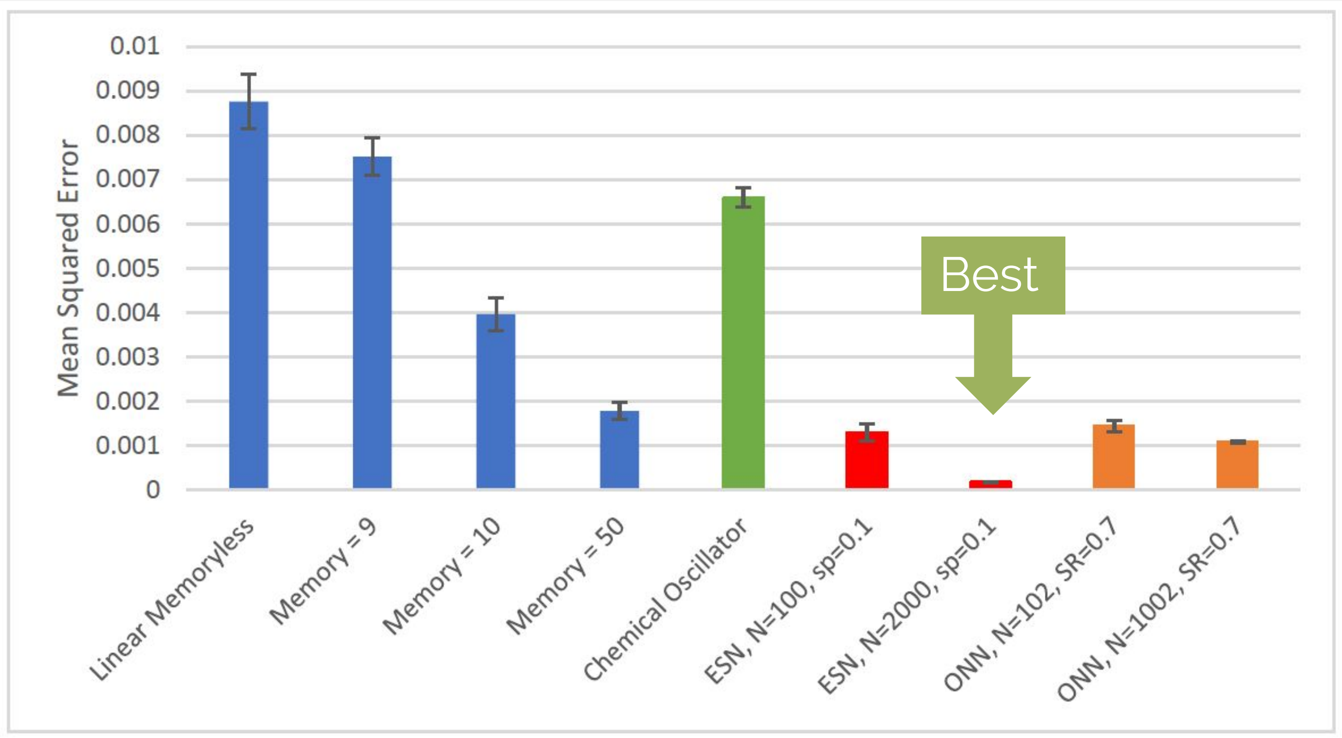
α = 0.3
β = 0.05
u is a random vector~U [0,0.5]



Results: NARMA-10

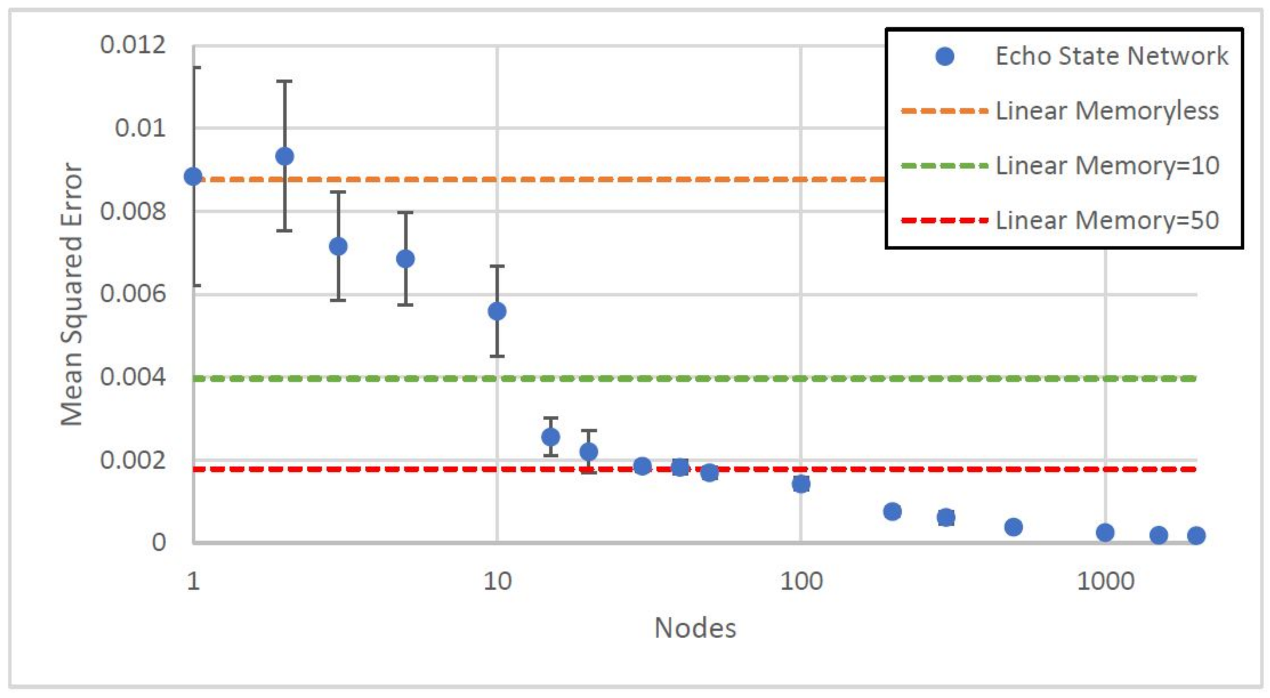
- Mean Squared error for:
- Ridge regression linear models
- Reservoir Layers:
  - Chemical Oscillator
  - Echo State Network
  - Oscillating Neural Network

MSE = mean((y-ŷ)2)



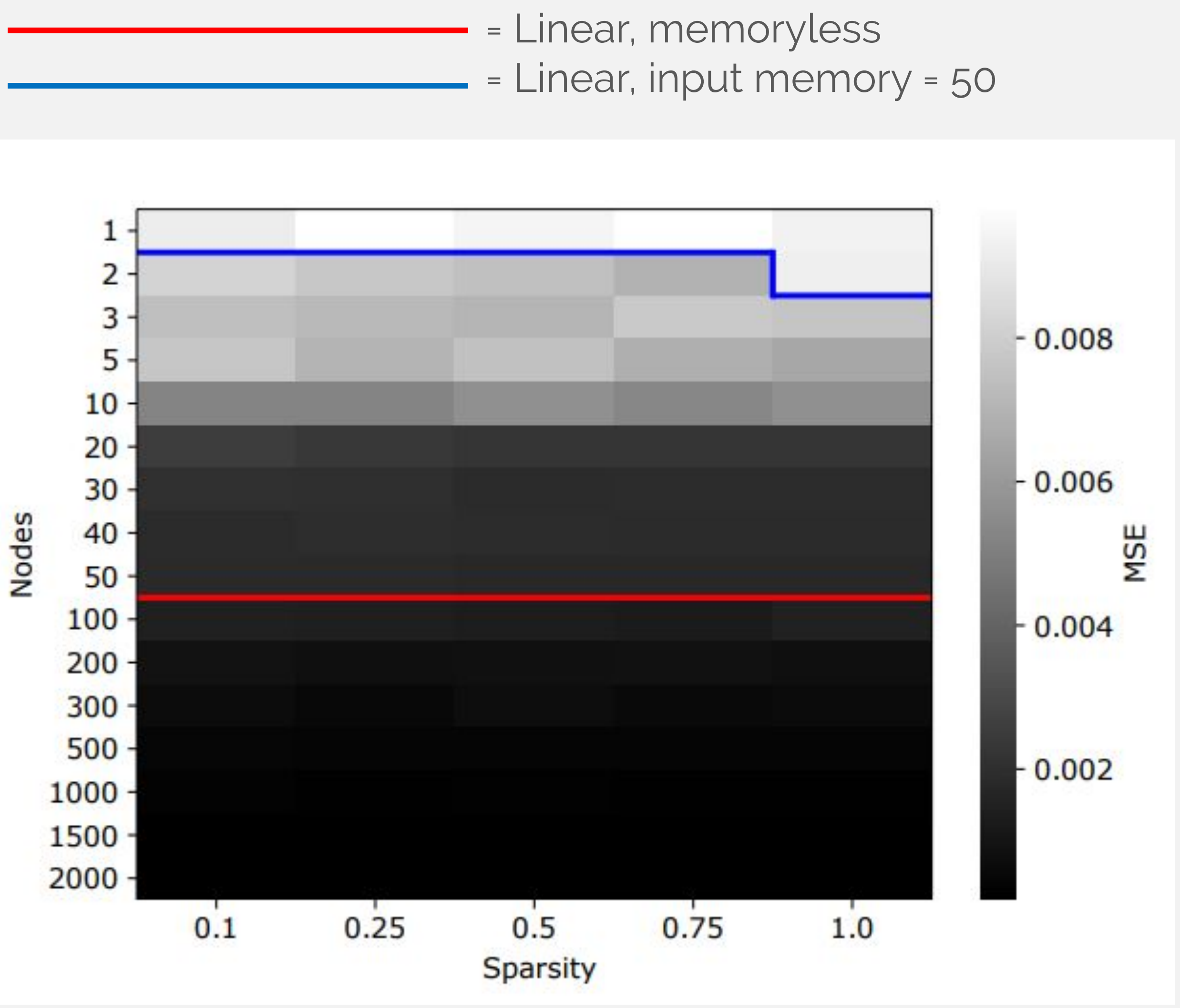
Results: NARMA-10 ESN

- Plot showing the relation of size of the network with performance.
- This was for a fully connected ESN.
- ESN's tend to have larger error bars for smaller networks as the random initializations can have a bigger impact on performance.



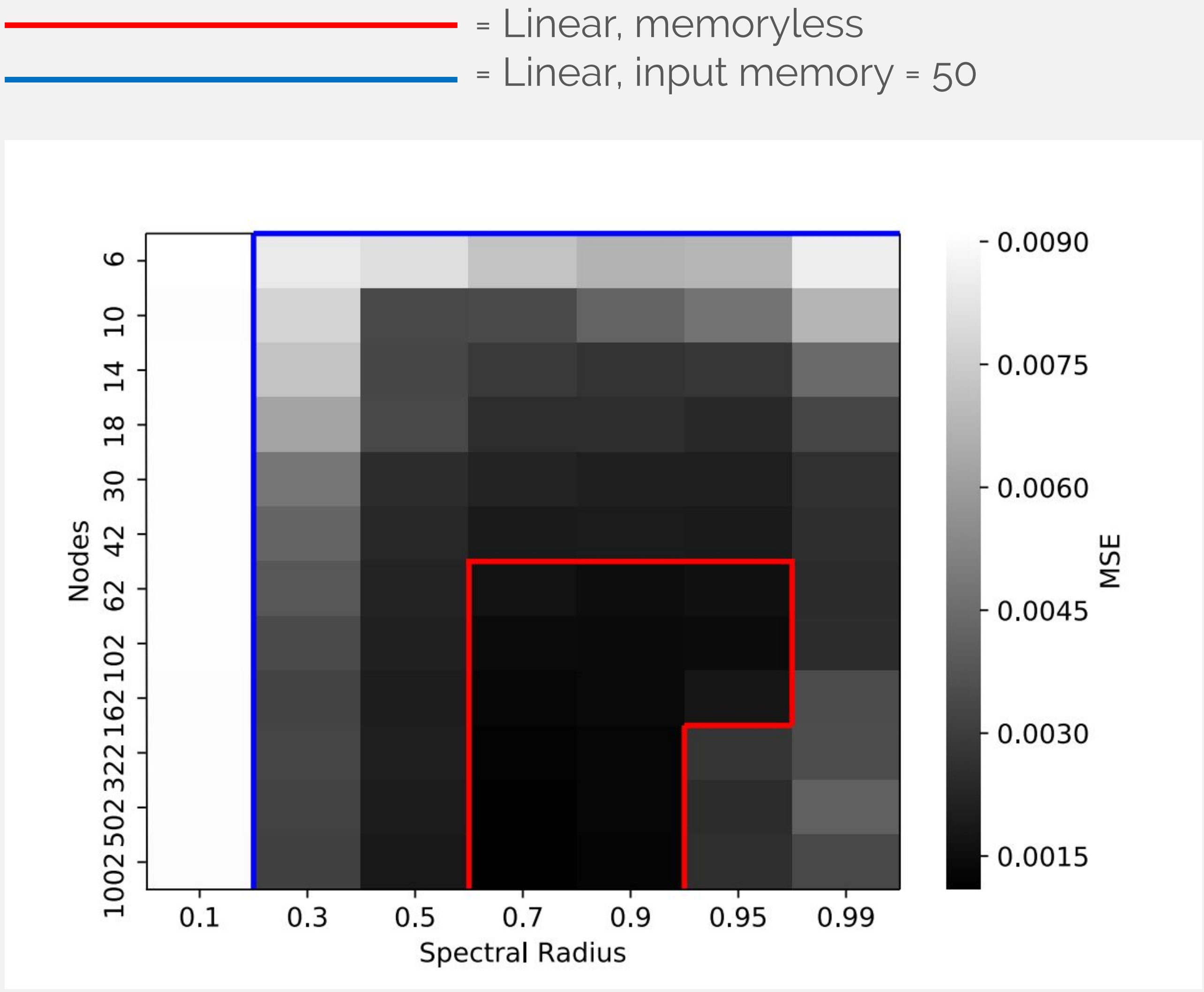
Results:
NARMA-10 ESN

- Plot showing the relation of size and sparsity of the network with performance.
- Sparsity = 1.0 is fully connected.
- There is not a significant deterioration in accuracy as network connectivity is reduced, network with fewer connections could be easier to implement.



Results: NARMA-10 ONN

- Plot showing the relation of size and spectral radius of the network with performance.
- This was for an oscillating neural network
- The spectral radius could translate to characterizing the chemical efflux rate from a micro-reactor setup



Dengue Virus.

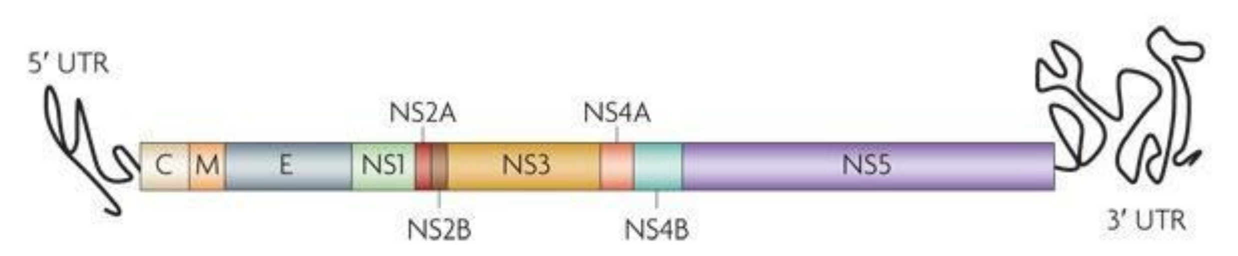
Structure of Dengue Virus

- ~11,000 base pairs, 4 Serotypes.
- Immunity to one serotype does not grant immunity to others. 65% genome similarity. 3-5% variation in same class.
- 200 base pair subsampling of the ***untranslated region*** (UTR) ***capsid*** regions of the genome.

Untranslated Region (UTR): Important in the regulation of translation and replication of the viral genome.

Capsid: Protein shell that protects the genetic material of a virus. ***[Guzman]***

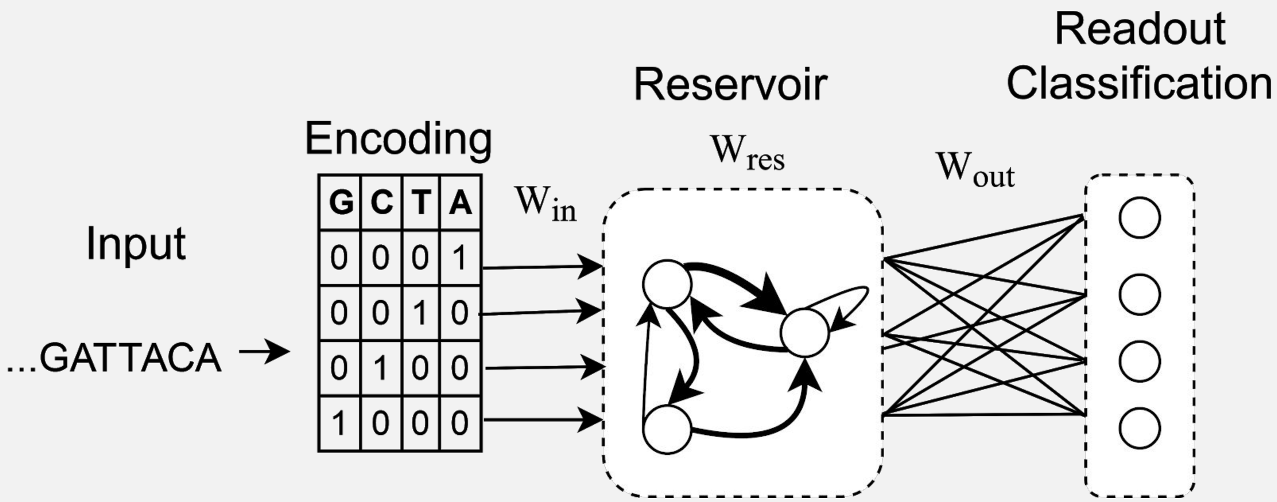
Task: Classify 200 base pair subsample from the ***untranslated region*** (UTR) ***capsid*** regions of the genome



M.G. Guzman et al. "Dengue: A continuing global threat." Nature Reviews Microbiology, 2010.

DNA Encoding

For physical implementation, this translates into an amplification step with the conversion of each nucleotide into a corresponding oligonucleotide.

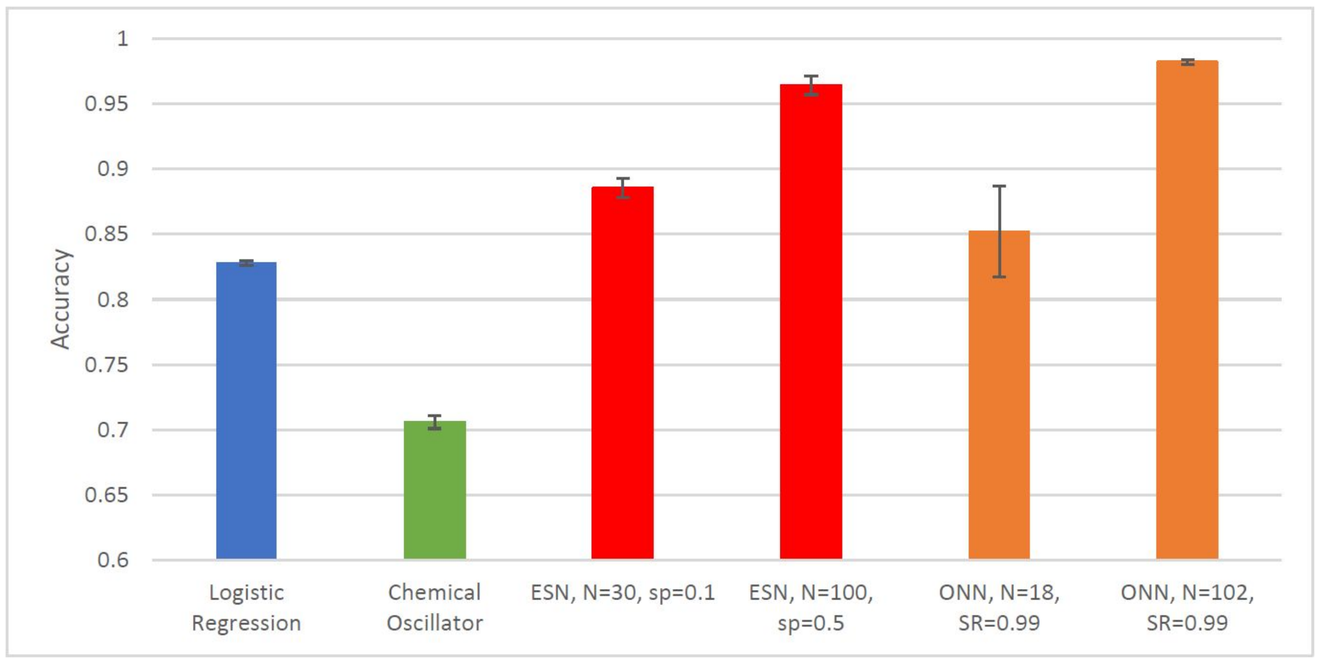


G	C	T	A
0	0	0	1
0	0	1	0
0	1	0	0
1	0	0	0

- One-hot encoding of the nucleotides:
    - A = [1, 0, 0, 0]
    - T = [0, 1, 0, 0]
    - C = [0, 0, 1, 0]
    - G = [0, 0, 0, 1]

Results: Dengue Virus

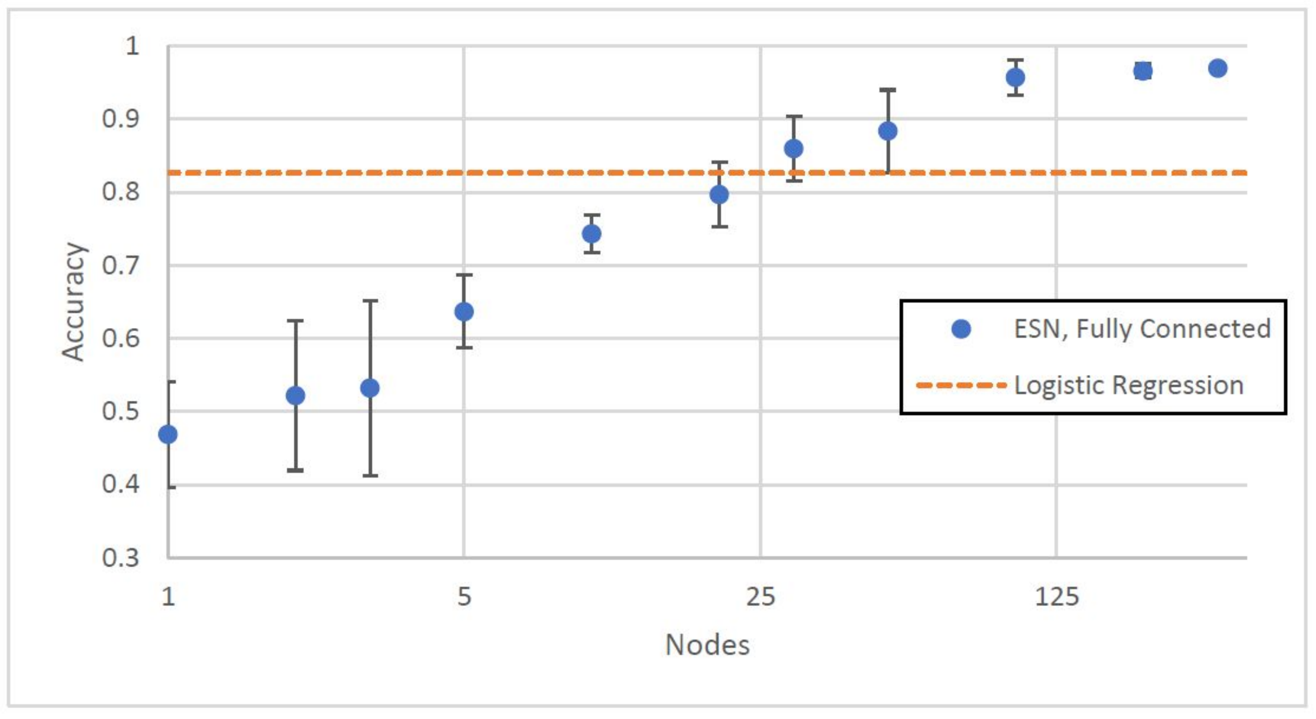
- While the simulated chemical oscillator struggled to effectively differentiate the classes, relatively small echo state networks and oscillating neural networks provided improvements in separability of the classes.



Accuracies of the different classifiers on the task of Dengue serotype classification.

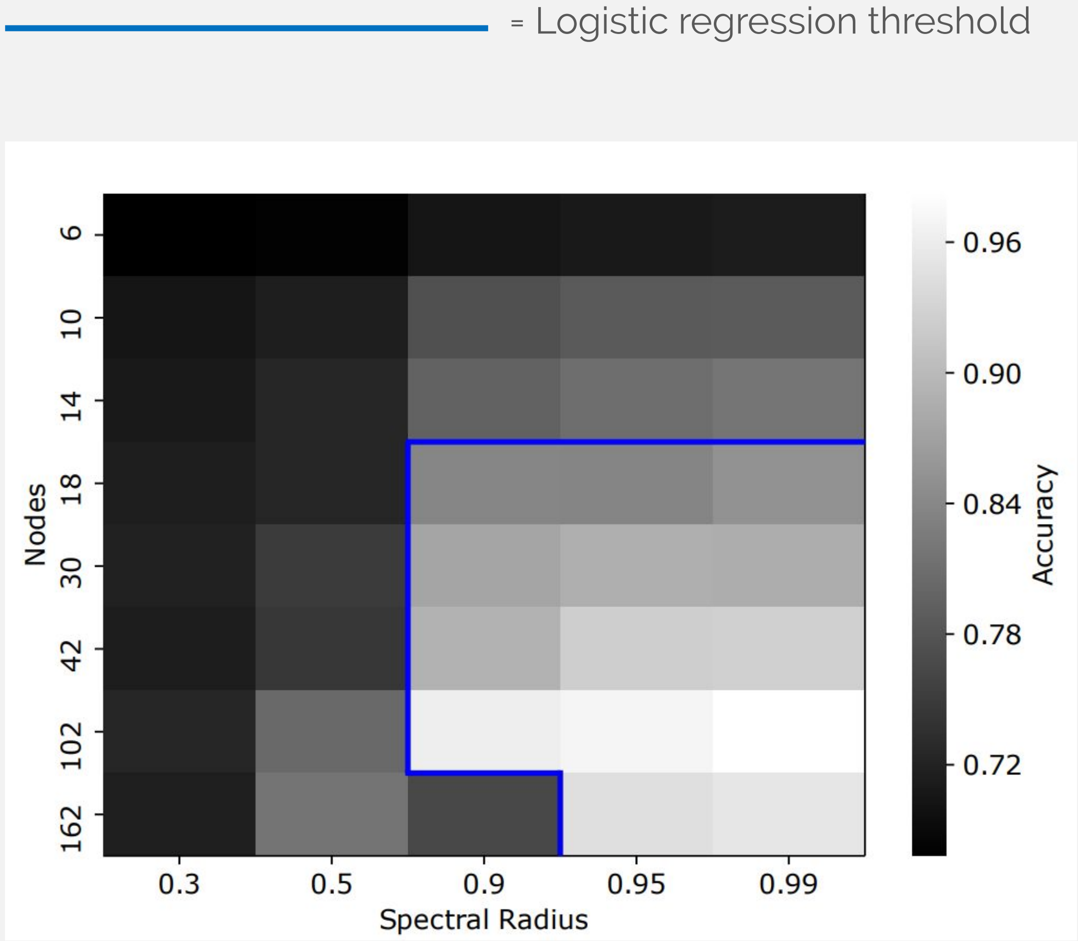
Results:
Dengue Virus ESN

- Large error bars evident in smaller networks due to initialization variation.
- Networks larger than 50 nodes, improvements over linear models are evident.



Results:
Dengue Virus ONN

- The improved performance at higher spectral radius shows that having a longer memory of previous inputs to the network can improve performance.
- This increased memory and computation accomplished with higher node values allows for the extraction of effective genomic features for distinguishing serotype.



Chemical Reaction Networks.

The Belousov-Zhabotinsky Reaction

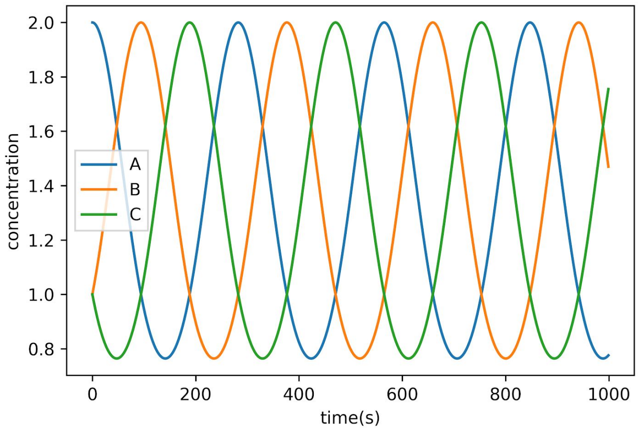
Molecular Dynamics

- DNA can be used to construct arbitrary chemical reactions.
- Rock, paper, scissors reaction as described capable of non-linear dynamics. ***[Srinivas]***
- Stochastic simulation can also be explored for small numbers of molecules using the Gillespie algorithm ***[Rao]***.

$$A + B \xrightarrow{\text{k1}} 2B$$

$$B + C \xrightarrow{\text{k2}} 2C$$

$$C + A \xrightarrow{\text{k3}} 2A$$



$$\frac{dA}{dt} = -k1[A][B] + k3[C][A]$$

$$\frac{dB}{dt} = -k2[B][C] + k1[A][B]$$

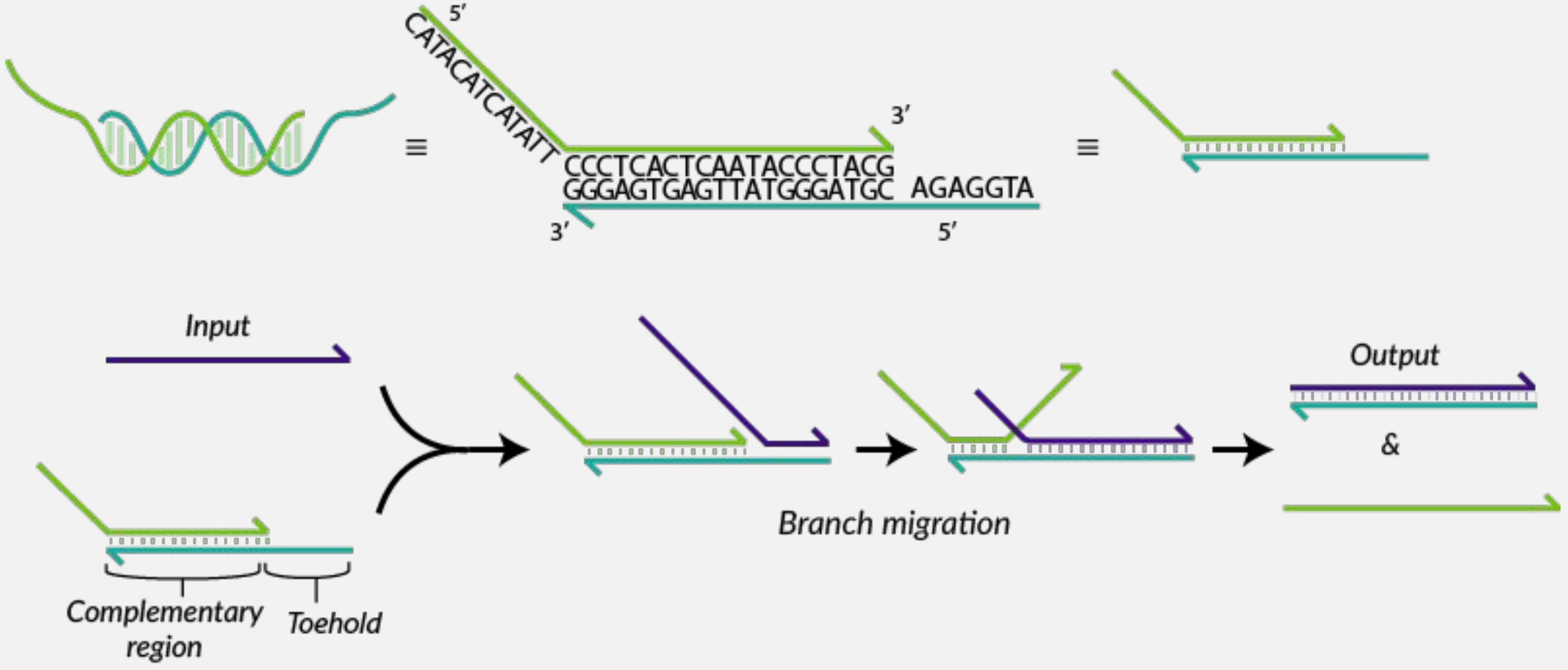
$$\frac{dC}{dt} = -k3[C][A] + k2[B][C]$$

N. Srinivas, et al, "Enzyme-free nucleic acid dynamical systems," *Science*, 2017.
C. V. Rao et al, "Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm," The Journal of Chemical Physics, 2003.

Toehold Mediated Strand Displacement *(TMSD)*

Boolean logic and arbitrary chemical reaction equations can be implemented by cascading these circuits together. ***[Zhang, Srinivas]***

How this type of network would be physically implemented.



D. Y. Zhang et al,. Dynamic DNA nanotechnology using strand displacement reactions. Nature Chemistry, 2011.
N. Srinivas, et al, “Enzyme-free nucleic acid dynamical systems,” *Science*, 2017.

Conclusion.

Contributions.

- Prepared tutorials for others in the lab for reading and encoding genomic data files from the splice junction gene sequences dataset. Along with examples for creating echo state networks using PyTorch and Keras at ***https://github.com/Neighborc/tlabTutorial***
- Evaluate the simulated use of a chemical oscillator as a reservoir layer for sequence prediction tasks. Showing that these chemical oscillators can improve performance above the memoryless linear baseline for **NARMA-A** from MSE= 0.42 to 0.34. and for **NARMA-10** from MSE=8.8×10^3 to 6.6×10^3.
- Evaluate the simulated use of a chemical oscillator as a reservoir layer for the MNIST dataset achieving **90.6%** classification accuracy
- Explored the use of Echo State Networks for the multivariate sequence classification task for dengue genome serotype classification achieving **96.4%** accuracy
- Presented a research poster at the **2019 International Conference on DNA Computing and Molecular Programming** in Seattle, WA, where I published a peer-reviewed abstract for my poster in the conference poster proceedings. Title: "**Reservoir Computing for Genome Sequence Classification**."
- Currently drafting a paper based on this research for submission to the **Artificial Life** journal

Conclusions.

- By implementing these networks in DNA and perturbing them with a nucleotide input sequence, the task of pathogen classification can be accomplished
- This work demonstrates that randomized non-linear networks of limited size and complexity can achieve an improvement over linear baseline models

Future work

- Expand the section of genome sampled.
- Adjust subsample length to be smaller and characterize accuracy loss.
- Encoding input in a different way such as through k-mer one-hot encoding.

Questions?