Medical & Biological Engineering & Computing manuscript No.
(will be inserted by the editor)

Supplementary Information for 'Classification of SARS-CoV-2 Viral Genome Sequences using Neurochaos Learning'

Harikrishnan NB[a,b] · **Pranay SY**[b] · **Nithin Nagaraj**[b]

This is the supplementary information pertaining to the main manuscript. We list the properties of neurochaos learning, and data preprocessing of genome sequences.

S1 Properties of Neurochaos Learning

We list below the properties of Neurochaos Learning (NL) architecture.

S1.1 Nonlinearity

Neurons in the brain are known to have a non-linear response and further found to exhibit chaotic behaviour [1], [2]. However, existing ANN architectures do not exhibit chaos at the level of neurons. In the case of NL, the neurons fire chaotically upon encountering an input sample. The notion of firing (trajectory) is absent in traditional ANNs (Table 1: Main Manuscript).

S1.2 Topological transitivity and existence of a dense orbit

Every GLS neuron in NL is a 1D chaotic map and exhibits topological transitivity which is defined below.

Definition: A dynamical system (Σ) is transitive if for each two points $x, y \in \Sigma$ and for $\varepsilon > 0$, there exist a $z \in \Sigma$ such that on finite number of iterations, z reaches the ε neighbourhood of x and y.

Harikrishnan N. B. thanks "The University of Trans-Disciplinary Health Sciences and Technology (TDU)" for permitting this research as part of the PhD programme.

Harikrishnan NB, Pranay SY, Nithin Nagaraj
[a]PhD scholar, The University Of Trans-Disciplinary Health Sciences And Technology,
Bengaluru, India.
[b]Consciousness Studies Programme,
National Institute of Advanced Studies,
Indian Institute of Science Campus, Bengaluru, India.
E-mail: harikrishnannb@nias.res.in, mail@pranaysy.com, nithin@nias.res.in

Definition: A set Y is a dense subset of X if, for any point $x \in X$, there is a point y in the subset Y arbitrarily close to x [3].

We use the Topological Transitivity property of chaos and the existence of a dense orbit to prove the Universal Approximation Theorem (UAT) for GLS neurons. An example of a dense orbit on the binary map is the real number that has a binary expansion given by '0.0 1 00 01 10 11 000 001...' [3]. The skew-tent map can be proved to have a dense orbit by the fact that there exists a conjugacy with the binary map.

S1.3 Chaos for feature engineering

ChaosFEX feature extraction maps the data into a high dimensional space. The input data matrix with m data instances and n points per data instance ($m \times n$) is mapped to $m \times 4n$. We extract 4 features from the chaotic trajectory. In general, the input data matrix with size $m \times n$ is mapped to chaotic feature space where the new matrix dimension is $m \times kn$, where k represents the number of distinct features extracted from the chaotic firing of the neurons in the input layer of NL. In `ChaosNet` [4], $k = 1$ because only a single feature (Firing Rate) was extracted. Thus, chaos based feature engineering can be incorporated in either NL (as we have done in this work) or in conventional ANNs.

S1.4 Other properties

Deterministic chaos provides a rich variety of behaviours - periodic, quasi-periodic, eventually periodic and non-periodic orbits or trajectories, multiple co-existing attractors, fractal boundaries, various types of synchronization - all of which are being exploited in engineering applications. This adaptability of chaotic systems allows applications in several fields such as chaotic computing [5], lossless compression and memory encoding [6, 7], chaos based cryptography [8], chaotic neural multiplexing [9], chaotic neural networks [10], fractal image compression [11] etc. Despite having a positive Lyapunov exponent, two chaotic systems still synchronizes under specific conditions [12].

The GLS neurons used in NL is proven to be useful in memory encoding [6], lossless compression [13], cryptography [14], [15] and error control coding [16]. These rich properties of chaotic neurons employed in NL (possibly in conjunction with traditional ML architectures) makes it a very good competitor for ANNs.

S2 Data Preprocessing for Genome Sequence

Data preprocessing is the first step in any machine learning task. The following data preprocessing is carried out for genome sequence data.

- Step 1: Conversion of nucleotide sequence into numeric format. We choose the same numeric conversion as mentioned in [17]. The numeric conversion is as follows: Cytosine (C) = 0.25, Thymine (T) = 0.50, Guanine (G) = 0.75, Adenine (A) = 1.0.
- Step 2: Compute the absolute value of the Fast Fourier Transform (FFT) of the numeric genome sequence.
- Step 3: The FFT coefficients of the nucleotide sequences are normalized independently so as to lie in the range [0,1].

References

1. Philippe Faure and Henri Korn. Is there chaos in the brain? i. concepts of nonlinear dynamics and methods of investigation. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, 324(9):773–793, 2001.
2. Henri Korn and Philippe Faure. Is there chaos in the brain? ii. experimental evidence and related models. *Comptes rendus biologies*, 326(9):787–840, 2003.
3. Robert L Devaney, Peter B Siegel, A John Mallinckrodt, and Susan McKay. A first course in chaotic dynamical systems: theory and experiment. *Computers in Physics*, 7(4):416–417, 1993.
4. Harikrishnan Nellippallil Balakrishnan, Aditi Kathpalia, Snehanshu Saha, and Nithin Nagaraj. Chaosnet: A chaos based artificial neural network architecture for classification. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(11):113125, 2019.
5. Behnam Kia, K Murali, Mohammad-Reza Jahed Motlagh, Sudeshna Sinha, and William L Ditto. Synthetic computation: chaos computing, logical stochastic resonance, and adaptive computing. In *International Conference on Theory and Application in Nonlinear Dynamics (ICAND 2012)*, pages 51–65. Springer, 2014.
6. Aditi Kathpalia and Nithin Nagaraj. A novel compression based neuronal architecture for memory encoding. In *Proceedings of the 20th International Conference on Distributed Computing and Networking*, pages 365–370, 2019.
7. Zainab Aram, Sajad Jafari, Jun Ma, Julien C Sprott, Sareh Zendehrouh, and Viet-Thanh Pham. Using chaotic artificial neural networks to model memory in the brain. *Communications in Nonlinear Science and Numerical Simulation*, 44:449–459, 2017.
8. Ljupco Kocarev. Chaos-based cryptography: a brief overview. *IEEE Circuits and Systems Magazine*, 1(3):6–21, 2001.
9. Nithin Nagaraj and KR Sahasranand. Neural signal multiplexing via compressed sensing. In *2016 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–5. IEEE, 2016.
10. Kazuyuki Aihara, T Takabe, and Masashi Toyoda. Chaotic neural networks. *Physics letters A*, 144(6-7):333–340, 1990.
11. Yuval Fisher. *Fractal image compression: theory and application*. Springer Science & Business Media, 2012.
12. Louis M Pecora and Thomas L Carroll. Synchronization in chaotic systems. *Physical review letters*, 64(8):821, 1990.
13. Nithin Nagaraj, Prabhakar G Vaidya, and Kishor G Bhat. Arithmetic coding as a non-linear dynamical system. *Communications in Nonlinear Science and Numerical Simulation*, 14(4):1013–1020, 2009.
14. Kwok-Wo Wong, Qiuzhen Lin, and Jianyong Chen. Simultaneous arithmetic coding and encryption using chaotic maps. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 57(2):146–150, 2010.
15. Nithin Nagaraj. *Novel applications of chaos theory to coding and cryptography*. PhD thesis, NIAS, 2008.
16. Nithin Nagaraj. Using cantor sets for error detection. *PeerJ Computer Science*, 5:e171, 2019.
17. Alejandro Lopez-Rincon, Alberto Tonda, Lucero Mendoza-Maldonado, Eric Claassen, Johan Garssen, and Aletta D Kraneveld. Accurate identification of sars-cov-2 from viral genome sequences using deep learning. *bioRxiv, doi: https://doi.org/10.1101/2020.03.13.990242*, 2020.