Discovering cause-effect from observational data is an important but challenging problem in science and engineering. In this work, a recently proposed brain inspired learning algorithm namely Neurochaos Learning (NL) is used for the classification of cause-effect from simulated data. The data instances used are generated from coupled AR processes, coupled 1D chaotic skew tent maps, coupled 1D chaotic logistic maps and a real-world prey-predator system. The proposed method consistently outperforms a five layer Deep Neural Network architecture for coupling coefficient values ranging from 0.1 to 0.7. Further, we investigate the preservation of causality in the feature extracted space of NL using Granger Causality (GC) for coupled AR processes and and Compression-Complexity Causality (CCC) for coupled chaotic systems and real-world prey-predator dataset. This ability of NL to preserve causality under a chaotic transformation and successfully classify cause and effect time series (including a transfer learning scenario) is highly desirable in causal machine learning applications.

Keywords: Neurochaos Learning, Granger Causality, Compression-Complexity Causality, Coupled Auto Regressive Processes, Coupled Chaotic Maps, Causal Machine Learning, Transfer Learning

1 Introduction

Understanding causal factors responsible for the occurrence of a phenomena is vital for the progress of research in science and technology. The development of drugs/vaccines to combat epidemic from time to time requires to pass the test of trustworthiness. This can be done by understanding the casual influence of the drug in combating the epidemic. The gold standard method used for this test is the Randomized Control Trials (RCT). RCTs are used in clinical research to discover cause-effect relation via clinical interventions. However, RCTs are not a universal answer \[1\] and suffer from their own challenges \[2,3\]. In many cases, RCTs are not technically viable, time consuming, expensive and even unethical. In such a scenario, the need to develop novel causal inference and causal discovery algorithms to learn from observational data \[4\] plays a crucial role. The availability of high computational resources and big data motivates the development of data driven learning algorithms. Machine Learning (ML) and Deep Learning (DL) owes to the abundance of data and computational resources for its growth. Despite the success of ML/DL algorithms in the field of natural language processing, computer vision, speech recognition, these algorithms face difficulty in interpretability and trustworthiness. This is due to the fact that these algorithms are merely discovering associations between ‘input’ and ‘output’ in the name of ‘learning’. However, discovering associations alone is insufficient as an explanation to aid in the
decision making process. Often, what is needed is a causal explanation which goes over and beyond associations (or correlations). Hence, arises the need for causal machine learning algorithms where causality meets learning.

Judea Pearl’s ladder of causation provides a three level hierarchical order of a casual learner [6]. The three levels proposed are: (1) Association, (2) Intervention and (3) Counterfactuals. Most of the current causality testing methods are at the level of association. Some of the popular causality testing methods for time series data are Granger Causality (GC) [7], Transfer Entropy (TE) [8], data compression based causality testing method namely Compression-Complexity Causality [9]. GC and TE are popular causality testing methods used in econometrics [10][11], climatology [12][13], neuroscience [14][15] etc. However, both GC and TE lie at the lowest rung of the ladder -  associational causality. CCC is an example of interventional causality measure and corresponds to the second rung of this ladder.

In this research, we use a recently proposed brain inspired learning algorithm namely Neurochaos Learning [16] (NL) to learn generalized causal patterns from time series data. NL draws its inspiration from the chaotic firing of neurons in the brain [17]. We investigate the following: (a) can NL learn unique patterns of cause and effect from observational data?, (b) do the chaos based features extracted from the input layer of NL preserve causality? In order to tackle the first question, a binary classification problem is formulated for the classification of cause-effect. The efficacy of NL and DL in learning generalized cause-effect pattern from data and thereby performing classification tasks are investigated. The second question is addressed by studying the preservation of cause-effect for the features extracted from the input layer of NL. To this end, we use Granger Causality (GC) and Compression-Complexity Causality (CCC).

The sections in the paper are arranged as follows: Section 2 describes the proposed method used to do the cause-effect classification. Section 3 provides details of the simulated data used to carry out the experiments. Section 4 deals with experiments, results and discussions on simulated and real world prey-predator dataset, as well as a transfer learning scenario. Section 5 addresses the limitations and scope for future work. The concluding remarks are provided in Section 6.

2 Proposed Method

Neurochaos Learning (NL) is a novel brain inspired neuronal learning algorithm that has been recently proposed. The authors in [18][19] claim that NL is inspired from the chaotic firings of neurons in the brain and has mainly two architectures: (a) ChaosNet [19], (b) ChaosFEX+ML. In [20], the authors employ ChaosNet for continual learning. In another work [21], the authors propose deep ChaosNet for action recognition in videos.

Inspired by these recent developments, in this work, we employ ChaosNet architecture for the classification of cause-effect from observational data. The architecture consists of an input layer of Generalized Lüroth Series (GLS) neurons which are one-dimensional (1D) chaotic skew tent maps described as follows:

\[
T(x) = \begin{cases} 
\frac{x}{b}, & 0 \leq x < b, \\
\frac{(1-x)}{(1-b)}, & b \leq x < 1,
\end{cases}
\]  

(1)

where the skewness of the map is controlled by the parameter \( b \) (0 < \( b < 1 \)). Upon arrival of the input data/ stimulus \( (x_t) \), the chaotic GLS neurons in the input layer starts firing (from the initial value \( q \)) until the chaotic neural trace of the neurons reaches the \( \varepsilon \) neighbourhood of the corresponding stimulus. From the neural trace thus generated, the following features are extracted:

1. **Firing time (N):** The amount of time the chaotic neural trace takes to recognize the input stimulus.

2. **Firing rate (R):** Fraction of time the chaotic neural trace is above the discrimination threshold \( b \) so as to recognize the stimulus.

3. **Energy (E):** For the chaotic neural trace \( y(t) \) with firing time \( N \), energy is defined as:

\[
E = \sum_{i=1}^{N} |y(t_i)|^2.
\]  

(2)

4. **Entropy (H):** For the chaotic neural trace \( y(t) \), we first compute the binary symbolic sequence \( \text{Sym}(t) \) as follows:

\[
\text{Sym}(t_i) = \begin{cases} 
0, & y(t_i) < b, \\
1, & b \leq y(t_i) < 1,
\end{cases}
\]  

(3)

where \( i = 1 \) to \( N \) (firing time). We then compute Shannon Entropy of \( \text{Sym}(t) \) as follows:

\[
H(\text{Sym}) = -\sum_{i=1}^{2} p_i \log_2(p_i) \text{ bits},
\]  

(4)
where \( p_1 \) and \( p_2 \) refers to the probabilities of the symbols 0 and 1 occurring in \( \text{Sym}(t) \) respectively.

For each input value \( s \) (stimulus) of a data instance of class \( k \) is mapped to a 4D vector \([N_s, R_s, E_s, H_s]\). The collection of these 4D vectors forms the ChaosFEX feature space. If the input data is 1D with \( Z \) classes, the mean representation vector of the \( k \)-th class is given by \( \frac{1}{N} \sum_{i=1}^{N} [N_i, \sum_{i=1}^{N} R_i, \sum_{i=1}^{N} E_i, \sum_{i=1}^{N} H_i] \) where \( m \) is the number of data instances (training data) for each class.

In the case of ChaosNet, the classifier computes the cosine similarity of the ChaosFEX features extracted from the test sample with the pre-computed mean representation vectors (consisting of mean values of ChaosFEX features for each stimuli) of each class. The predicted class is assigned the label corresponding to the maximum cosine similarity. A detailed explanation of the ChaosNet architecture and its working is provided in [18].

### 3 Datasets

To evaluate the efficacy of ChaosNet and deep learning for the classification of cause-effect, we used simulated datasets from (a) Coupled autoregressive (AR) processes, (b) Coupled 1D chaotic maps in master-slave configuration (1D skew tent maps and the 1D logistic maps) and real-world dataset from a (c) prey-predator system.

#### 3.1 Coupled AR processes

The governing equations for the coupled AR processes are the following:

\[
M(t) = a_1 M(t-1) + \gamma r(t),
\]
\[
S(t) = a_2 S(t-1) + \eta M(t-1) + \gamma r(t),
\]

where \( M(t) \) and \( S(t) \) are the independent and the dependent time series respectively; \( a_1 = 0.8, a_2 = 0.9 \), the noise intensity \( \gamma = 0.03 \) and \( r(t) \) is the i.i.d additive gaussian noise drawn from a standard normal distribution. The coupling-coefficient \( \eta \) is varied from 0 to 1 in steps of 0.1. We generated 1000 independent random trials for each value of \( \eta \). Each of the data instances are of length 2000, after removing the initial 500 samples (transients) from the time series.

#### 3.2 Coupled 1D Chaotic maps in Master-Slave configuration

##### 3.2.1 Coupled Skew-tent maps

The governing equations used to generate the master and slave time series for the coupled 1D skew-tent maps are the following:

\[
M(n) = T_1(M(n-1)),
\]
\[
S(n) = (1-\eta)T_2(S(n-1)) + \eta M(n-1),
\]

where \( M(n) \) is the master and \( S(n) \) is the slave system. \( M(n) \) influences the dynamics of the slave system (equation [8]). The coupling coefficient is given by \( \eta \) is varied from 0 to 0.9 with a step size of 0.1. \( T_1(n) \), and \( T_2(n) \) are skew tent maps with skewness \( b_1 = 0.65 \), and \( b_2 = 0.47 \) respectively. The initial values are chosen randomly for the master-slave system in the interval \((0, 1)\). We generated 1000 independent random trials for each value of \( \eta \). Each of the data instances are of length 2000, after removing the initial 500 samples (transients) from the time series. As an example, the attractor for this master-slave system for \( \eta = 0.4 \) is provided in Figure [1].

##### 3.2.2 Coupled Logistic maps

The 1D Logistic map is a widely used model to study population dynamics [22]. The governing dynamics for coupled logistic maps in master-slave configuration is given by:

\[
M(n) = L_1(M(n-1)),
\]
\[
S(n) = (1-\eta)L_2(S(n-1)) + \eta M(n-1),
\]

The coupling coefficient \( \eta \) is varied from 0 to 0.9. \( L_1(n) = 4 \cdot L_1(n-1)(1 - L_1(n-1)) \), and \( L_2(n) = 3.82 \cdot L_2(n-1)(1 - L_2(n-1)) \). The attractor for this coupled dynamical system is provided in Figure [14].

For both systems, 1000 data instances \( (M(n), S(n)) \) are generated and grouped as class-0 \( (M(n)): \text{Cause}) \) and class-1 \( (S(n): \text{Effect}) \) respectively. Table [7] gives details of the train-test split for the classification task.
Figure 1: Attractor for the coupled 1D chaotic skew tent maps in master slave configuration with $b_1 = 0.65$ and $b_2 = 0.47$.

Table 1: Train-Test distribution for the simulated datasets.

| Class    | Traindata | Testdata |
|----------|-----------|----------|
| Class-0  | 801       | 199      |
| Class-1  | 799       | 201      |
| Total    | 1600      | 400      |

4 Experiments, Results and Discussions

In this section, we begin with a description of hyperparameter tuning for NL and DL followed by a demonstration of causality preservation by ChaosFEX for coupled AR processes (and the failure of DL). Subsequently, macro F1-scores for ChaosNet and DL for the cause-effect classification for each $\eta$ are plotted. For all results in this paper, software implementation is performed using Python 3, scikit-learn [23], keras [24], ChaosFEX toolbox [16], Multivariate Granger Causality (MVGC) toolbox [25], CCC toolbox [9] and MATLAB.

4.1 Hyperparameter tuning for NL

Every ML algorithm has a set of hyperparameters that needs to be tuned for efficient performance. In the case of ChaosNet, there are three hyperparameters - initial neural activity ($q$), discrimination threshold ($b$), and noise intensity ($\varepsilon$) [16]. The hyperparameter tuning is done only once with the traindata corresponding to $\eta = 0.4$ (Table 1) separately for the coupled AR processes and coupled skew tent maps.

For a fixed value of $b = 0.499$, and $\varepsilon = 0.171$, $q$ was varied from 0.01 to 0.98 with a stepsize of 0.01 for both coupled AR processes and coupled chaotic skew tent maps. In the case of coupled AR processes, a maximum average macro F1-score $= 0.605$ is obtained for $q = 0.78$. In the case of coupled skew tent maps, a maximum average macro F1-score $= 1.0$ was obtained for the following values of $q = [0.16, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.32, 0.34, 0.36, 0.37, 0.38, 0.48, 0.51, 0.52, 0.56, 0.57, 0.72, 0.76, 0.77, 0.78, 0.79, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.98]$ in a five-fold cross validation using traindata. We choose $q = 0.56$ for further experiments.

4.2 Deep Learning Parameters

A five layer Deep Learning architecture was used to evaluate the efficacy of cause effect classification. The number of nodes in the input layer $= 2000$, followed by first hidden layer with 5000 neurons and sigmoid activation function. The output from this layer is passed to second hidden layer with 500 neurons and ReLU activation function. This is followed by 100 neurons with Relu activation function in the third hidden layer. The fourth hidden layer contains 30 neurons with ReLu activation function. The output layer contains 2 neurons with softmax activation function. Training was done for 30 epochs.

4.3 Preservation of Granger Causality for coupled AR processes under a chaotic transformation

Accurate estimation of causality for coupled AR processes is ideally suited for Granger Causality (GC) since GC models time series as AR processes. This is the reason GC is very popular in causal analysis of financial time series.
climatology and neuroscience. ChaosFEX features are extracted after a chaotic transformation of the input time series. It is important to verify whether Granger Causality is preserved under such a nonlinear transformation. To test this, we perform the following experiment. For $q = 0.78$, $b = 0.499$, and $\varepsilon = 0.171$, the firing time has been extracted from ChaosFEX for time series from coupled AR processes. The GC vs. coupling coefficient plot for firing time depicted in Figure 2 reveals that indeed Granger Causality is nicely preserved. The GC values shown here are obtained from 50 random trials. This indicates the reliability of the chaotic transformation of NL in preserving granger causality and hence very desirable in applications which employ GC. Note that such a property is not available for DL (Figure 3) making NL a very attractive candidate for causal ML applications.

Figure 2: GC vs. coupling coefficient for the firing time feature extracted from the coupled AR processes. The ChaosFEX settings are $q = 0.78$, $b = 0.499$, and $\varepsilon = 0.171$. The GC F-statistic is computed from 50 trials.

4.4 Classification of Cause Effect for Coupled Skew-Tent maps in Master-Slave Configuration

The performance of ChaosNet and 5 layer DL for varying coupling coefficient ($\eta$) is depicted in Figure 4. ChaosNet and DL give identical performance (a macro F1-score = 1.0) for $\eta$ values up to 0.5. However, for $\eta = [0.6, 0.7]$, ChaosNet outperforms DL. Beyond $\eta > 0.6$, the synchronization error < 0.013 indicating that the two time series are practically identical. Hence, classification fails as there is essentially nothing to distinguish between the two time series owing to synchronization.

4.5 Preservation of causality in ChaosFEX feature space

To check if causality is preserved in the ChaosFEX feature space of unidirectionally coupled skew-tent maps, we use the measure Compression-Complexity Causality (CCC) [9]. CCC is ideal for application to non-linear time series,

\[^{1}\text{The maximum model order setting in the MVGC toolbox was set to 30 for ChaosFEX features and to 20 for DL features.}\]
Figure 4: Performance comparison of ChaosNet and five layer DL for the classification of cause-effect for 1D coupled skew tent map in master-slave configuration.

where often GC can face issues. Figure 5 shows the CCC estimates for original (raw) time series. The estimates plotted are averaged over 50 trials with CCC parameters set to $L = 100, w = 15, \delta = 50, B = 4$. As expected, the magnitude of CCC values from the master to the slave increases with increasing coupling and begins to decrease as the time-series become synchronized and effectively no transfer of information can be detected. As discussed in [9], CCC can take negative values and its magnitude denotes the strength of causation. CCC values in the direction of causation from slave to master are much lower in magnitude and remain close to zero.

Figure 5: CCC vs Coupling Coefficient for the raw data corresponding to 1D chaotic skew tent map in master-slave configuration.

CCC for the corresponding firing time feature of ChaosFEX for these coupled maps is depicted in Figure 6. These values are also averaged over 50 trials and computed with CCC parameters set to $L = 120, w = 15, \delta = 60, B = 2$. Here, master to slave CCC does not perfectly preserve the increasing trend with increasing values of coupling, however decreases just as the estimates for raw data, when the processes proceed to synchronization. The slave to master estimates for the coupling range 0.1 – 0.9 are quite low in magnitude and remain close to zero as expected. Even though the estimates for zero coupling are not very close to zero or take exactly the same value and the increasing trend for increasing coupling is not perfectly preserved, CCC estimates in the direction for which coupling exists and the its opposite are well differentiated and hence it can be said that ChaosFEX features do a reasonably good job in preserving causality even for skew chaotic tent maps. Surrogate based causality analysis might help to reveal a more adequate picture of the differentiation and of the existence of causality, but is out of the scope of this work.

---

These were chosen using the selection criteria described in [9].
4.6 Transfer Learning for Cause-Effect Classification

We have demonstrated the possibility of cause-effect classification for coupled chaotic maps in master-slave configuration. However, it is interesting to explore whether we can transfer this ‘learning’ to scenarios where the master-slave systems are different from the ones for which the method was trained. Specifically, we shall change the skewness of both the master and slave systems from the original parameter values used in the training phase. A more drastic case of transfer learning would be to test on an entirely different nonlinear map, for example, coupled logistic maps without training afresh (using the same learned parameters as the coupled skew tent maps). These would help us determine to what extent the learning is generalizable for both NL and DL.

We consider the following cases for transfer causal learning:

- **Case I:** Train with master-slave coupled skew tent map system ($b_1 = 0.65, b_2 = 0.47$) and test with master-slave coupled skew tent map system with $b_1 = 0.6$ and $b_2 = 0.4$ (classification results are in Figure 7). The attractor for skew tent map master slave testdata with $b_1 = 0.6$ and $b_2 = 0.4$ is provided in Figure 8.

- **Case II:** Train with master-slave coupled skew tent map system ($b_1 = 0.65, b_2 = 0.47$) and test with master-slave coupled skew tent map system with $b_1 = 0.49$ and $b_2 = 0.52$ (classification results are in Figure 11). The attractor for skew tent map master slave testdata with $b_1 = 0.49$ and $b_2 = 0.52$ is provided in Figure 12.

![Figure 6: CCC vs Coupling Coefficient for firing time (ChaosFEX feature) corresponding to 1D chaotic coupled skew tent maps in master-slave configuration.](image)

![Figure 7: Transfer learning for case I: comparative performance of ChaosNet and five layer DL evaluated using macro F1-score for $\eta$ in the range 0 to 0.9 with a stepsize of 0.1.](image)
CASE-EFFECT PRESERVATION AND CLASSIFICATION USING NEUROCHAOS LEARNING

Figure 8: Case I: Attractor for the coupled 1D chaotic skew tent maps in master slave configuration with $b_1 = 0.6$ and $b_2 = 0.4$.

Figure 9: Transfer Learning for Case II: comparative performance of ChaosNet and five layer DL evaluated using macro F1-score for $\eta$ in the range 0 to 0.9 with a stepsize of 0.1.

Figure 10: Case II: Attractor for the coupled 1D chaotic skew tent maps in master slave configuration with $b_1 = 0.1$ and $b_2 = 0.3$.

- **Case IV:** Train with skew tent map master-slave coupled skew tent map system ($b_1 = 0.65$, $b_2 = 0.47$) and test with logistic map master-slave system with $A_1 = 4.0$ and $A_2 = 3.82$ (Figure 13). The attractor for logistic map master slave testdata with $A_1 = 4.0$ and $A_2 = 3.82$ is provided in Figure 14.

In the case of testing with data generated from different models, ChaosNet completely outperforms DL for Case I (Figure 7) and Case II (Figure 9) for the entire range of $\eta$. For Case III (Figure 11) and IV (Figure 13), ChaosNet and
Figure 11: Transfer Learning for Case III: comparative performance of ChaosNet and five layer DL evaluated using macro F1-score for $\eta$ in the range 0 to 0.9 with a stepsize of 0.1.

Figure 12: Case III: Attractor for the coupled 1D chaotic skew tent maps in master slave configuration with $b_1 = 0.49$ and $b_2 = 0.52$.

Figure 13: Transfer Learning for Case IV: comparative performance of ChaosNet and five layer DL evaluated using macro F1-score for $\eta$ in the range 0 to 0.9 with a stepsize of 0.1.

DL shows similar trends (with DL outperforming ChaosNet for some values of $\eta$). A high performance of ChaosNet in classification shows the separability of the mean representation vectors of cause and effect.
4.7 Real Data

The efficacy of ChaosFEX features in cause-effect preservation was evaluated on a real world dataset from a prey-predator system as well. The data consists of 71 data points of predator (Didinium nasutum) and prey (Paramecium aurelia) populations [26, 27] (Figure 15). This is a system of bidirectional causation as the predator population directly influences the prey population and then itself gets influenced by a change in the prey population. It is expected that the direct causal influence from the predator to the prey should be higher than in the opposite direction.

For our analysis, initial 9 transients were removed. With the remaining 62 data points, CCC values are computed for the raw data and ChaosFEX feature (firing time). The parameters of CCC used for the raw data are $L = 40, w = 15, \delta = 4, B = 8$. In the case of ChaosFEX, firing time feature was extracted for the following NL hyperparameters: $q = 0.56, b = 0.499, \varepsilon = 0.1$. The CCC parameters chosen for ChaosFEX are $L = 40, w = 15, \delta = 4, B = 4$. The results for the cause-effect preservation for the raw data and ChaosFEX firing time feature is provided in Table 2. CCC rightly captures the higher causal influence from predator to prey population and finds a lower influence in the opposite direction, for both raw data and ChaosFEX feature - firing time.

| Class          | CCC (raw data) | CCC (firing time) |
|----------------|----------------|-------------------|
| Predator $\rightarrow$ Prey | 0.1160         | 0.0484            |
| Prey $\rightarrow$ Predator  | -0.0210        | 0.0050            |

These values were chosen using the selection criteria described in [9].
5 Limitations

In the case of coupled 1D chaotic maps, NL consistently performed well up to $\eta = 0.5$ for classification. However, the same is not true for the classification of data generated from coupled AR processes. For $q = 0.78$, $b = 0.499$, and $\varepsilon = 0.171$, the classification results are depicted in Figure 16. In the same figure, it can be seen that DL performance is worse than NL. We have used the exact same architecture for DL as we have used for the cause-effect classification of data from coupled chaotic skew-tent maps in master-slave configure (section 4.2).

![Figure 16: F1-score vs. coupling coefficient for the classification of the coupled AR processes using ChaosNet.](image)

A maximum macro F1-score = 0.656 was obtained for $\eta = 1.0$ implying that ChaosNet was not able to find mean representation vectors that could separate the two classes. Choosing a more sophisticated classifier for NL (instead of the simplistic cosine-similarity metric) could solve this problem and improve classification results. We shall explore these possibilities in a future study.

In this research, we have shown the classification and preservation of causality for unidirectional causation of two variables. A detailed study needs to be undertaken for the classification and causal discovery for coupled high dimensional systems and real world datasets in the future.

6 Conclusion

In this work, Neurochaos Learning architecture - ChaosNet has been used in the classification of cause-effect for data generated from coupled chaotic maps. ChaosNet outperforms a five layer deep learning architecture in several cases. In the experiments, ChaosNet consistently outperformed a five layer DL up to a coupling coefficient of 0.7. Causality testing using Granger Causality (for coupled AR processes) and Compression-Complexity Causality (for coupled chaotic systems and for a real-world prey-predator system) on the firing times extracted from the chaotic neural traces reveals the preservation of cause-effect in the NL feature extracted space. Further, the efficacy of the proposed method was observed in transfer learning of the classification of cause-effect from the master-slave time series generated from different chaotic unimodal maps (skew-tent maps with different skews and logistic map with different parameters). This motivates future research direction of NL in lifelong learning framework and classification of cause-effect using ChaosNet on real world datasets.

The preservation of causality can be attributed to the rich properties of the nonlinear chaotic transformation of GLS neurons in NL (ChaosNet). Unlike traditional ANNs, NL is intrinsically a nonlinear deterministic algorithm that performs a point-by-point chaotic transformation, in fact, a non-linear embedding of the input raw features in to a high dimensional space. Deterministic Chaos combines the best of both the worlds - pseudo-randomness and determinism. The ergodic, ‘random-like’ structure of the chaotic neural traces enables an effective transformation of the input data (stimuli) preserving causality that is inherent in the input space and at the same time ensuring separability in the chaotic feature space for efficient classification.

---

4 We have performed some amount of hyperparmater tuning for DL architecture, however a more extensive tuning needs to be performed.
Acknowledgment

Harikrishnan N. B. thanks “The University of Trans-Disciplinary Health Sciences and Technology (TDU)” for permitting this research as part of the PhD programme. The authors gratefully acknowledge the financial support of Tata Trusts. A. Kathpalia acknowledges the financial support of the Czech Science Foundation, Project No. GA19-16066S and the Czech Academy of Sciences, Praemium Academiae awarded to M. Paluš. Nithin Nagaraj gratefully acknowledge the financial support of Cog. Sci. Res. Initiative (CSRI), Dep. of Sci. & Tech., Govt. of India under Grant No. DST/CSRI/2017/54(G).

References

[1] Bertrand Graz, Elaine Elisabetsky, and Jacques Falquet. Beyond the myth of expensive clinical study: Assessment of traditional medicines. Journal of ethnomedecology, 113(3):382–386, 2007.
[2] AD Nichol, M Bailey, DJ Cooper, On behalf of the POLAR, et al. Challenging issues in randomised controlled trials. Injury, 41:S20–S23, 2010.
[3] John Concato. Observational versus experimental studies: what’s the evidence for a hierarchy? NeuroRx, 1(3):341–347, 2004.
[4] Pushya A Gautama. Rcts and other clinical trial designs in ayurveda: A review of challenges and opportunities. Journal of Ayurveda and Integrative Medicine, 12(3):556–561, 2021.
[5] Joris M Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. The Journal of Machine Learning Research, 17(1):1103–1204, 2016.
[6] Judea Pearl and Dana Mackenzie. The book of why: the new science of cause and effect. Basic books, 2018.
[7] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, pages 424–438, 1969.
[8] Thomas Schreiber. Measuring information transfer. Physical review letters, 85(2):461, 2000.
[9] Aditi Kathpalia and Nithin Nagaraj. Data-based intervention approach for complexity-causality measure. PeerJ Computer Science, 5:e196, 2019.
[10] Craig Hiemstra and Jonathan D Jones. Testing for linear and nonlinear granger causality in the stock price-volume relation. The Journal of Finance, 49(5):1639–1664, 1994.
[11] Song Zan Chiou-Wei, Ching-Fu Chen, and Zhen Zhu. Economic growth and energy consumption revisited—evidence from linear and nonlinear granger causality. Energy Economics, 30(6):3063–3076, 2008.
[12] Timothy J Mosedale, David B Stephenson, Matthew Collins, and Terence C Mills. Granger causality of coupled climate processes: Ocean feedback on the north atlantic oscillation. Journal of climate, 19(7):1182–1194, 2006.
[13] Adolf Stips, Diego Macias, Clare Coughlan, Elisa Garcia-Gorriz, and X San Liang. On the causal structure between co2 and global temperature. Scientific reports, 6(1):1–9, 2016.
[14] Anil K Seth, Adam B Barrett, and Lionel Barnett. Granger causality analysis in neuroscience and neuroimaging. Journal of Neuroscience, 35(8):3293–3297, 2015.
[15] Raul Vicente, Michael Wibral, Michael Lindner, and Gordon Pipa. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. Journal of computational neuroscience, 30(1):45–67, 2011.
[16] NB Harikrishnan and Nithin Nagaraj. When noise meets chaos: Stochastic resonance in neurochaos learning. Neural Networks, 143:425–435, 2021.
[17] 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.
[18] Harikrishnan Nelliappallil 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.
[19] NB Harikrishnan and Nithin Nagaraj. Neurochaos inspired hybrid machine learning architecture for classification. In 2020 International Conference on Signal Processing and Communications (SPCOM), pages 1–5. IEEE, 2020.
[20] Touraj Laleh, Mojtaba Faramarzi, Irina Rish, and Sarath Chandar. Chaotic continual learning. 4-th Lifelong Learning Workshop at ICML, 2020.
[21] Huafeng Chen, Maosheng Zhang, Zhengming Gao, and Yunhong Zhao. Deep chaosnet for action recognition in videos. *Complexity*, 2021, 2021.

[22] Robert M May. Simple mathematical models with very complicated dynamics. *The Theory of Chaotic Attractors*, pages 85–93, 2004.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[24] François Chollet et al. Keras. [https://keras.io](https://keras.io), 2015.

[25] Lionel Barnett and Anil K Seth. The mvgc multivariate granger causality toolbox: a new approach to granger-causal inference. *Journal of neuroscience methods*, 223:50–68, 2014.

[26] Brendan G Veilleux. The analysis of a predatory interaction between didinium and paramecium. *Master’s thesis. University of Alberta, Edmondton*, 1976.

[27] Christian Jost and Stephen P Ellner. Testing for predator dependence in predator-prey dynamics: a non-parametric approach. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1453):1611–1620, 2000.