# Nonlinear Granger Causality Analysis Using Neural Network Architectures for Sequential Data

Diego N. Chacón Wilches, and Alvaro D. Orjuela-Cañón, *Senior Member, IEEE*

***Abstract*— This work presents an analysis of nonlinear Granger causality (NNGC) computation based on artificial neural networks (ANN) architectures. The study evaluates the impact of using computational intelligence models as ANN, and how the training parameters can modify the causality estimation. For this, the employing from three chaotic maps (Hénon, Ikeda, and Tinkerbell) and one neuron-like map (Rulkov) in bivariate scenarios were implemented. Three architectures from the ANN were used such as multilayer perceptron in a mode of the nonlinear autoregressive models, long-short memory term, and gated recurrent unit architectures were used to compute the NNGC, applying a forecasting on sequential data techniques. Results demonstrated that NNGC is highly sensitive to neural network parameters, such as the number of neurons, lag length, and batch size, with an optimal configuration by varying across chaotic maps. Comparisons with classical Granger causality were tested, revealing that neural networks effectively discover nonlinear relationships missed by linear methods, particularly in the Hénon and Rulkov maps.**

**Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/9993**

***Index Terms*—Granger causality, nonlinear models, artificial neural networks, chaotic maps, sequential data forecasting, neural networks Granger causality.**

## I. Introduction

CAUSALITY is a fundamental concept in time series or sequential data analysis, allowing researchers to determine whether one variable can predict or influence another. Specifically, Granger Causality (GC) was introduced by Clive Granger in 1969 [1], providing a statistical framework to assess such directional influences. According to GC, if past values of a variable $X$ (i.e., its lags) contain information that can help to predict another variable $Y$ beyond the information contained in the past values of $Y$ alone, then $X$ is said to cause $Y$ in the GC sense. This methodology is rooted in the principle of information flow, quantifying the extent to which knowledge from sequential data can improve the forecasting of another. Over the years, GC has found wide applications in multiple disciplines. In economics, it has identified predictive relationships between economic indicators [2]. In neuroscience, it elucidates directional interactions among different brain regions [3], and in the wider biomedical engineering field, it explores dependencies between physiological signals [4].

For example, in the field of biomedical engineering, Granger Causality has been applied to study directional interactions between the brain and heart in pediatric patients with obstructive sleep apnea (OSA) [5]. In this case, GC was used on electroencephalography (EEG) signals and heart rate variability (HRV) derived from electrocardiogram (ECG) data to evaluate how specific brain subbands influence cardiac activity, and vice versa, before and after clinical treatment. The results revealed significant changes in functional connectivity between the two systems post-intervention, demonstrating the real-world value of GC in understanding physiological effects of medical conditions and their therapies.

Despite its broad adoption, the original GC framework relies on linear autoregressive (AR) models. Real-world systems often show nonlinear dynamics that can make it difficult to find accurately the directional influences when only linear models are used [6]. Consequently, researchers have proposed various nonlinear extensions, including kernel-based approaches [7], approaches relying on statistical transformations such as transfer entropy [8], and solutions leveraging artificial neural networks (ANNs) for modeling [9]. These methods aim to capture more complex dependencies between $X$ and $Y$, providing richer insights into the flow of information that might otherwise remain obscured.

Chaotic time series, like those created by the Hénon or Ikeda maps, act as a useful platform for studying nonlinear Granger causality [10]. These maps exhibit sensitive dependence on initial conditions and nonlinear deterministic behavior, characteristics that place unique demands on causality assessment. For instance, the Hénon map is known to exhibit chaotic attractors when certain parameters are selected, making it a useful model for testing how well an estimator detects irregular yet deterministic structures in the data [11]. Likewise, the Ikeda map demonstrates significant sensitivity to initial conditions, challenging linear Granger models to capture the underlying coupling between variables [12]. Methods capable of identifying causality in such chaotic dynamical systems suggest broader applicability to real-world scenarios where processes are often highly nonlinear.

ANN were proposed to modify the traditional computation of Granger causality, leading to what is commonly referred to as Neural Network Granger Causality (NNGC) [13]. Using architectures such as multilayer perceptron (MLP) feedforward networks [14], radial basis function (RBF) networks [15], or recurrent networks like long short-term memory (LSTM) architectures [16], researchers have successfully demonstrated that these models can capture complex, nonlinear relationships among variables. In addition, the neural granger causality formulations includes restrictions and guidelines to help make it clearer when identifying directional influences [9]. Although, these

The associate editor coordinating the review of this manuscript and approving it for publication was Carlos Thomaz (*Corresponding author: Diego N. Chacón Wilches*).

Diego N. Chacón Wilches and A. Orjuela-Cañón are with the School of Medicine and Health Sciences from Universidad del Rosario, Bogota, Colombia (e-mails: diegon.chacon@urosario.edu.co and alvaro.orjuela@urosario.edu.co).

network-based methods have a lot of flexibility, raising questions about choosing hyperparameters that can significantly impact both predictive metrics and the causality estimation.

More recent work highlights the need for taking into account considerations about architectural and training choices to avoid overfitting and to maintain statistical reliability in causal inference [17]. For example, the number of hidden neurons in a neural network can determine whether it sufficiently represents the nonlinearities in chaotic maps or time series generated by physiological signals. Likewise, the number of lags (i.e., historical inputs) or the batch size during training can change the model's ability to account for temporal dependencies and to converge to stable solutions [18]. When these factors are suboptimal, misinterpretation of causal relationships may arise, emphasizing the importance of systematic parameter tuning.

In this work, the principal objective is to analyze how specific neural network training parameters influence the computation of nonlinear GC when only two time series (bivariate case) are considered. To achieve this, we employ the Python tool developed by Rosoł et al. [4]. The experiments were carried out using bivariate chaotic time series to illustrate whether chaotic dynamics is related to GC and the nonlinear GC one. Three key parameters were analyzed: the number of hidden-layer neurons, the number of lags to incorporate in the model, and the batch size used during training. By systematically varying these parameters, it was measured the impact on the ability of neural network-based methods to establish the causal relationships in nonlinear settings, highlighting both methodological insights and practical implications. The paper is organized into Section II, which outlines the time series, models, and tools used to conduct the comparison. Section III presents the results regarding the calculated causalities, with a focus on the neural network models, followed by a discussion in Section IV. Finally, Section V summarizes the conclusions obtained from the completion of this work.

## II. Material and Methods

### A. Sequential Data Generation

To develop the experiments and compare the proposed models for connectivity identification, four scenarios with different chaotic time series generation were simulated: the Hénon chaotic map, the Ikeda chaotic map, the Tinkerbell chaotic map, and the Rulkov chaotic map. First, the Hénon map generates time series according to the following equations:

$$x_{x+1} = 1 - ax_n^2 + y_n \tag{1}$$

$$y_{n+1} = bx_n$$

In these equations, $x_n$ and $y_n$ represent the generated time series at time step $n$. The constants $a$ and $b$ are set to $a = 1.4$ and $b = 0.3$, which induce chaotic behavior in the generated time series.

Second, the Ikeda map is defined by the following equation:

$$x_{n+1} = 1 + u(x_n sin(t_n) - y_n cos(t_n))$$

$$y_{n+1} = u(x_n \cos t_n - y_n \sin t_n) \tag{2}$$

Here, $u$ is a constant set to 0.9, and $t_n$ is defined as $t_n = 0.4 - \frac{6}{1+x_n^2+y_n^2}$. The variables $x_n$ and $y_n$ again represent the time series at time step $n$, and this formulation captures the dynamics of the coupled chaotic system.

Third, the Tinkerbell map was built according to the following equations:

$$x_{n+1} = x_n^2 - y_n^2 + by_n$$
$$y_{n+1} = 2x_n y_n + cx_n + dy_n \tag{3}$$

In this case, the constants are set to $a = -1.4$, $b = 1.0$, $c = 0.3$, and $d = 0.1$, which contribute to the generation of intricate and complex patterns in the time series, characteristic of chaotic behavior.

Finally, the Rulkov map was employed for simulating a coupled time series in the way:

$$x_{n+1} = \frac{\propto}{1 + y_n^2} + x_n$$
$$y_{n+1} = y_n - \mu(x_n + \beta y_n - \gamma) \tag{4}$$

In this formulation, $\alpha = 0.02$, $\mu = 0.001$, $\beta = 0.001$, and $\gamma = 0.15$ are parameters that enable the generation of robust chaotic time series.

For all scenarios, the time series were composed of 500 data points, allowing for the development of computational experiments with different algorithms without considering hardware limitations. Additionally, the time series were normalized to the interval [−1,1] before applying the methods used in the analysis.

### B. Granger Causality Analysis

The classical Granger causality test was applied to the bivariate time series $(X, Y)$. The test assessed whether past values of $X$ or $Y$ contributed statistically significant information for predicting the future values of $X$ or $Y$, depending on the direction of causality being analyzed, beyond what could be inferred from their own past values alone. For this purpose, linear AR models were estimated using the ordinary least squares (OLS) method, while the statistical significance of Granger causality was evaluated using the F-test and the chi-square test [1].

### C. Nonlinear Granger Causality with Neural Networks

To capture nonlinear dependencies between time series, we employed a neural-network-based model to compute the nonlinear GC approach. This method involves comparing the prediction errors of two distinct models: the restricted model and the full model. The restricted model predicts the time series $X$ using only its past values, while the full model predicts $X$ using both its past values and the past values of another time series $Y$. A reduction in prediction error when moving from the restricted model to the full model indicates the presence of Granger causality from $Y$ to $X$.

In the analysis, three ANN architectures were implemented to perform the causality tests: multilayer perceptron (MLP), long short-term memory (LSTM), and gated recurrent unit (GRU). All models were constructed with a single hidden layer, as we aimed to develop a simple model with low error and low complexity. The MLP is a fully connected network that captures complex relationships in the data, implementing a nonlinear AR provided by activation functions with nonlinearity. The LSTM is a recurrent neural network designed to learn long-term dependencies, making it particularly effective for time series data. The GRU is a

streamlined variant of LSTM that uses fewer parameters, thus simplifying the model while maintaining performance.

Each neural network architecture was configured with specific parameters to optimize performance. We tested various numbers of neurons in the hidden layer, specifically 10, 50, and 100, to assess their impact on model accuracy. These values were selected based on preliminary experimental trials with a wider range of neuron counts. The three final values were chosen to represent low, medium, and high model complexity while keeping the total number of training experiments computationally feasible. Additionally, we evaluated the effect of different lag values, specifically 5, 10, and 20, to capture temporal dependencies effectively. The batch sizes were set to either 16 or 32, and the training process was standardized across models with a fixed number of epochs set at 50 to avoid excessively long training times. A learning rate of 0.0001 was chosen to ensure stable convergence during training.

For each chaotic map and each pair of time series $(X, Y)$, we followed a systematic training procedure. The time series data was split into a training set (70%) and a test set (30%) to evaluate out-of-sample prediction performance. Both the restricted and full models were trained on the training set using the specified neural network architectures and parameters. To address the initialization problem with neural network models, we tested each architecture with 100 initializations. To assess causality, the Wilcoxon signed-rank test was used to determine if the prediction error from the full model was significantly lower than that of the restricted model, thereby inferring the nonlinear GC.

### D. Software and Tools

All analyses were conducted using Python version 3.7.10. To run this code, a virtual Python environment was created, and the following versions of TensorFlow, Keras, and scikit-learn were used to avoid version conflicts, as these were the existing versions at the time of the creation of the Custom Neural Network Package inspired by Rosoł et al. [4] for nonlinear Granger causality analysis: TensorFlow 2.13.1, Keras 2.13.1, and scikit-learn 1.3.1.

The following libraries and packages were employed: statsmodels (version 0.14.4) [19] for classical Granger causality tests, Custom Neural Network Package for nonlinear Granger causality analysis, NumPy and Pandas for data manipulation and preprocessing, and TensorFlow and Keras for building and training neural network models.

### E. Statistical Analysis

Statistical analysis was performed to assess whether the past values of $X$ provided useful information for predicting $Y$, or vice versa, depending on the direction of causality being analyzed. The analysis yielded $p$-values and test statistics for different lag orders, and a *p-value* threshold of 0.05 was used to determine significant Granger causality. A *p-value* below 0.05 indicated significant Granger causality in the tested direction.

In addition, causality was evaluated using neural network models and the Wilcoxon signed-rank test, which was employed to compare the errors from the restricted and full models. A *p-value* threshold of 0.05 was used to ascertain significant causality.

## III. Results

The results for connectivity based on Granger causality (GC) and nonlinear neural network-based Granger causality (NNGC) are presented separately. In the case of nonlinear Granger causality with neural networks, we emphasize the errors obtained from the trained models and specify the values for the number of neurons in the hidden layer, the lags, and the batch size.

### A. Granger Causality

The results in Table I shows the connectivity, indicating a flow of information from $X$ to $Y$ or $Y$ to $X$ in specific cases. This calculation was derived from a statistical test that assessed the connectivity between the subsystems. For the GC calculation, lag values of 5, 10, and 20 were tested. Significant information flow was observed in the Ikeda and Tinkerbell maps across all lags, while the Hénon map showed connectivity only at lag 5. The Rulkov map did not demonstrate significant connectivity at any lag.

### B. Granger Causality with Neural Networks models

For the calculation of NNGC, Tables II, III, IV, and V present the training parameters of the best-performing models for each neural network architecture. These models are selected based on having the lowest residual sum of squares (RSS) prediction errors when considering both $X$ and $Y$ for each causality direction analyzed.

TABLE I
LINEAR GRANGER CAUSALITY CALCULATION

| Chaotic Map | Connectivity | Lags | F-test | p-value | Chi2-test | p-value |
|---|---|---|---|---|---|---|
| Hénon | X→Y | 5 | - | - | - | - |
| | | 10 | - | - | - | - |
| | | 20 | - | - | - | - |
| | X←Y | 5 | 4.769 | 0.000 | 24.188 | 0.000 |
| | | 10 | 0.109 | 0.999 | 1.121 | 0.997 |
| | | 20 | 0.141 | 1.000 | 2.971 | 1.000 |
| Ikeda | X→Y | 5 | 663.144 | 0.000 | 3391.076 | 0.000 |
| | | 10 | 336.362 | 0.000 | 3514.225 | 0.000 |
| | | 20 | 161.772 | 0.000 | 3537.610 | 0.000 |
| | X←Y | 5 | 25.341 | 0.000 | 129.585 | 0.000 |
| | | 10 | 13.264 | 0.000 | 138.582 | 0.000 |
| | | 20 | 6.767 | 0.000 | 147.981 | 0.000 |
| Tinkerbell | X→Y | 5 | 285.000 | 0.000 | 1457.388 | 0.000 |
| | | 10 | 123.283 | 0.000 | 1288.030 | 0.000 |
| | | 20 | 63.229 | 0.000 | 1382.679 | 0.000 |
| | X←Y | 5 | 44.160 | 0.000 | 225.823 | 0.000 |
| | | 10 | 36.332 | 0.000 | 379.589 | 0.000 |
| | | 20 | 18.880 | 0.000 | 412.877 | 0.000 |
| Rulkov | X→Y | 5 | - | - | - | - |
| | | 10 | - | - | - | - |
| | | 20 | - | - | - | - |
| | X←Y | 5 | 1.504 | 0.186 | 7.631 | 0.178 |
| | | 10 | 0.664 | 0.757 | 6.811 | 0.743 |
| | | 20 | 0.325 | 0.997 | 6.823 | 0.997 |

Similarly, Table VI provides a summary of the training parameters for the best neural network architecture obtained for each chaotic map used, based on the lowest error for predicting the time series $X$ or $Y$, depending on the direction of causality being analyzed, using both $X$ and $Y$. Otherwise, a single run with the parameters yielding the best results using both $X$ and $Y$, based on Table VI, is summarized in a series

of visualizations that illustrate the performance of the models trained to compute nonlinear Granger causality for different chaotic maps. Each chaotic map is represented by a specific group of plots, which include predictions, training and validation errors.

In Figs. 1 and 2, it is possible to observe the graphs of the predicted values of the time series using two models: one based solely on $X$ or $Y$ and another based solely on $X$ and $Y$. Meanwhile, Figs. 3 and 4 illustrate the mean squared error (MSE) during training and validation for both models over 50 epochs. These graphs provide an overview of the models' ability to capture the dynamics of the chaotic maps.

Figs. 5 to 8 display histograms of the error from the forecasting applied to the Hénon, Ikeda, Tinkerbell, and Rulkov time series maps, based on 100 training results obtained using the architectures listed in Table VI and the specified training parameters for the predictions X→Y (from X to Y) and X←Y (from Y to X).

TABLE II
CONNECTIVITY RESULTS BETWEEN TIME SERIES X AND Y FOR THE HÉNON MAP AND CORRESPONDING MODEL PARAMETERS

| Model | Connectivity | Time series | Training parameters | | | Error |
|---|---|---|---|---|---|---|
| | | | Neurons | Lags | Batch size | |
| MLP | X→Y | Y | 100 | 5 | 16 | 4.781 ± 0.403 |
| | | YX | | | | 0.025 ± 0.010 |
| | X←Y | X | 100 | 5 | 16 | 4.775 ± 0.351 |
| | | XY | | | | 3.359 ± 0.355 |
| LSTM | X→Y | Y | 100 | 10 | 16 | 10.4210 ± 0.4898 |
| | | YX | | | | 0.004 ± 0.001 |
| | X←Y | X | 100 | 5 | 16 | 8.441 ± 0.569 |
| | | XY | | | | 0.393 ± 0.193 |
| GRU | X→Y | Y | 100 | 5 | 16 | 11.254 ± 0.865 |
| | | YX | | | | 0.005 ± 0.001 |
| | X←Y | X | 100 | 5 | 16 | 8.099 ± 0.954 |
| | | XY | | | | 0.803 ± 0.319 |

TABLE III
CONNECTIVITY RESULTS BETWEEN TIME SERIES X AND Y FOR THE IKEDA MAP AND CORRESPONDING MODEL PARAMETERS

| Model | Connectivity | Time series | Training parameters | | | Error |
|---|---|---|---|---|---|---|
| | | | Neurons | Lags | Batch size | |
| MLP | X→Y | Y | 50 | 20 | 16 | 5.032 ± 0.133 |
| | | YX | | | | 1.177 ± 0.183 |
| | X←Y | X | 100 | 20 | 16 | 4.492 ± 0.072 |
| | | XY | | | | 3.511 ± 0.097 |
| LSTM | X→Y | Y | 50 | 20 | 16 | 5.580 ± 0.052 |
| | | YX | | | | 1.207 ± 0.099 |
| | X←Y | X | 100 | 20 | 16 | 4.208 ± 0.130 |
| | | XY | | | | 3.630 ± 0.147 |
| GRU | X→Y | Y | 100 | 5 | 16 | 9.081 ± 0.035 |
| | | YX | | | | 1.209 ± 0.030 |
| | X←Y | X | 100 | 10 | 16 | 7.281± 0.065 |
| | | XY | | | | 6.044 ± 0.243 |

TABLE IV
CONNECTIVITY RESULTS BETWEEN TIME SERIES X AND Y FOR THE TINKERBELL MAP AND CORRESPONDING MODEL PARAMETERS

| Model | Connectivity | Time series | Training parameters | | | Error |
|---|---|---|---|---|---|---|
| | | | Neurons | Lags | Batch size | |
| MLP | X→Y | Y | 100 | 20 | 16 | 3.279 ± 0.108 |
| | | YX | | | | 0.810 ± 0.085 |
| | X←Y | X | 100 | 5 | 16 | 2.875 ± 0.098 |
| | | XY | | | | 0.880 ± 0.138 |
| LSTM | X→Y | Y | 100 | 5 | 16 | 3.637 ± 0.171 |
| | | YX | | | | 1.223 ± 0.053 |
| | X←Y | X | 100 | 10 | 16 | 4.334 ± 0.122 |
| | | XY | | | | 0.613 ± 0.044 |
| GRU | X→Y | Y | 100 | 10 | 16 | 2.615 ± 0.119 |
| | | YX | | | | 1.274 ± 0.035 |
| | X←Y | X | 100 | 5 | 16 | 7.536 ± 0.095 |
| | | XY | | | | 2.036 ± 0.526 |

TABLE V
CONNECTIVITY RESULTS BETWEEN TIME SERIES X AND Y FOR THE RULKOV MAP AND CORRESPONDING MODEL PARAMETERS

| Model | Connectivity | Time series | Training parameters | | | Error |
|---|---|---|---|---|---|---|
| | | | Neurons | Lags | Batch size | |
| MLP | X→Y | Y | 50 | 20 | 32 | 0.108 ± 0.016 |
| | | YX | | | | 0.047 ± 0.063 |
| | X←Y | X | 10 | 20 | 32 | 0.003 ± 0.003 |
| | | XY | | | | 0.592 ± 0.184 |
| LSTM | X→Y | Y | 100 | 5 | 32 | 0.010 ± 0.004 |
| | | YX | | | | 0.034 ± 0.016 |
| | X←Y | X | 10 | 20 | 16 | 0.217 ± 0.127 |
| | | XY | | | | 0.299 ± 0.402 |
| GRU | X→Y | Y | 100 | 10 | 32 | 0.089 ± 0.005 |
| | | YX | | | | 0.010 ± 0.004 |
| | X←Y | X | 10 | 20 | 32 | 3.962 ± 4.251 |
| | | XY | | | | 1.362 ± 1.435 |

## IV. DISCUSSION

The analysis of results integrates the findings from both the linear Granger causality analysis presented in Table I and the nonlinear neural network approaches summarized in Tables II through VI. By comparing these two perspectives, it was possible to have a more comprehensive understanding of causality in the four chaotic maps: Hénon, Ikeda, Tinkerbell, and Rulkov. In Table I, the linear GC tests, revealed diverse causal patterns among the four chaotic maps for lag values of 5, 10, and 20. For the Hénon map, significant causality is observed in the direction Y→X at lag 5 (F-test .p = 0.000, $\chi^2$-test p = 0.000), indicating Y as a useful predictor of X under a short memory, a finding supported by neural network models (MLP and GRU) that show reduced prediction error when Y is included. Models that incorporate both time series (XY) often achieve very low errors (on the order of 0.001 or 0.004), implying that there is a deeper, nonlinear interaction missed by the linear test. Thus, while the linear method reveals a partial influence at a specific lag, the neural networks capture a deeper structure of causal links across multiple settings.

In contrast, for Ikeda, the linear Granger test revealed a bidirectional causal connection at all tested lags (5, 10, and 20), with p-values consistently below 0.05. This strong interaction is confirmed in the neural network framework (Tables III and VI), where MLP, LSTM, and GRU models systematically yield improved prediction accuracy when both time series are included. The errors in columns labeled XY are noticeably lower than when only a single series is used to predict the other, reinforcing the idea that Ikeda's dynamics are highly interactive in both directions. In this scenario, neural networks and linear methods generally agree on the presence of causality, but the networks provide more flexibility in describing nonlinearity and achieving lower residual errors.

TABLE VI
SUMMARY OF THE BEST CONNECTIVITY RESULTS BETWEEN TIME SERIES X AND Y FOR EACH ANALYZED CHAOTIC MAP AND CORRESPONDING MODEL PARAMETERS

| Chaotic Map | Connectivity | Model | Training parameters | | | Error |
|---|---|---|---|---|---|---|
| | | | Neurons | Lags | Batch size | |
| Hénon | X→Y | LSTM | 100 | 10 | 16 | 0.004 ± 0.001 |
| | X←Y | LSTM | 100 | 5 | 16 | 0.393 ± 0.193 |
| Ikeda | X→Y | LSTM | 50 | 20 | 16 | 1.207 ± 0.099 |
| | X←Y | MLP | 100 | 20 | 16 | 3.511 ± 0.097 |
| Tinkerbell | X→Y | MLP | 100 | 20 | 16 | 0.810 ± 0.085 |
| | X←Y | LSTM | 100 | 5 | 16 | 0.612 ± 0.043 |
| Rulkov | X→Y | GRU | 100 | 10 | 32 | 0.010 ± 0.004 |
| | X←Y | LSTM | 10 | 20 | 16 | 0.299 ± 0.402 |

In a similar way, the Tinkerbell map demonstrated a robust two-way Granger causality by linear criteria (Table I), with p-values well below 0.05 across multiple lags. Neural techniques (Tables IV and VI) maintain this conclusion and additionally reduce the forecasting errors, especially with mid-range lags (5 or 10) and at least 50 to 100 hidden neurons. Although the linear results already demonstrated strong connections between X and Y, the neural approaches refined the credibility of these links by capturing complex global patterns that may not be addressed in purely linear frameworks. In practical terms, MLP, LSTM, and GRU perform similarly well, though specific configurations (e.g., 100 neurons, batch size 16) stand out in terms of error reduction.

Conversely, the Rulkov map presented a different scenario, according to Table I, no significant causality emerges at any lag for either direction, indicating that a purely linear model fails to detect meaningful influence. However, the neural network-based results for Rulkov (Tables V and VI) show that both MLP and recurrent architectures identify lower prediction errors when trained with data from both X and Y. In some cases, errors drop to 0.010 or 1.362 in X→Y and X←Y predictions with the GRU architecture, hinting at an underlying nonlinear dependency that escapes linear detection. These improvements strongly suggest that neural methods, particularly those capable of capturing complex temporal behavior, are essential for unveiling subtle couplings in chaotic systems like Rulkov.

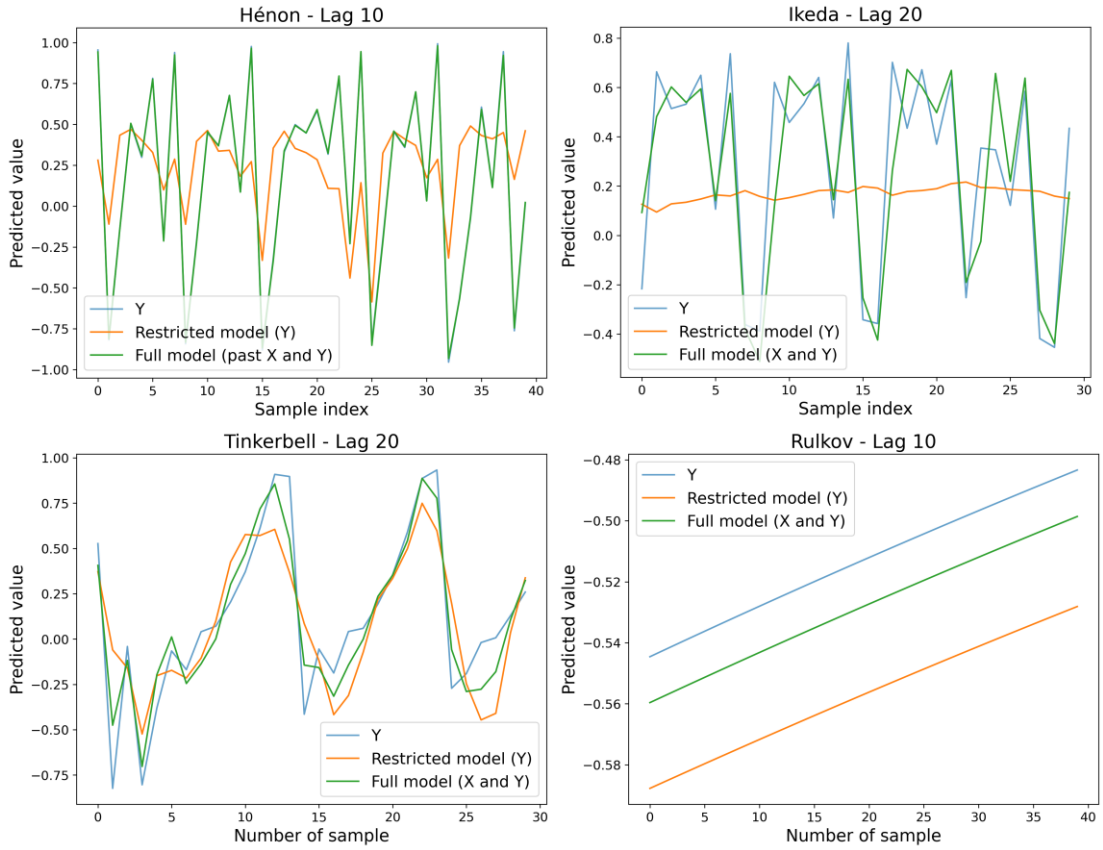


Fig. 1. Predicted values of the time series for four chaotic maps: Hénon (top-left, lag = 10), Ikeda (top-right, lag = 20), Tinkerbell (bottom-left, lag = 20), and Rulkov (bottom-right, lag = 10). Each panel compares the actual values of $Y$ (blue line) with predictions from two models: the restricted model, based solely on Y (orange line), and the full model, based on both X and Y (green line).

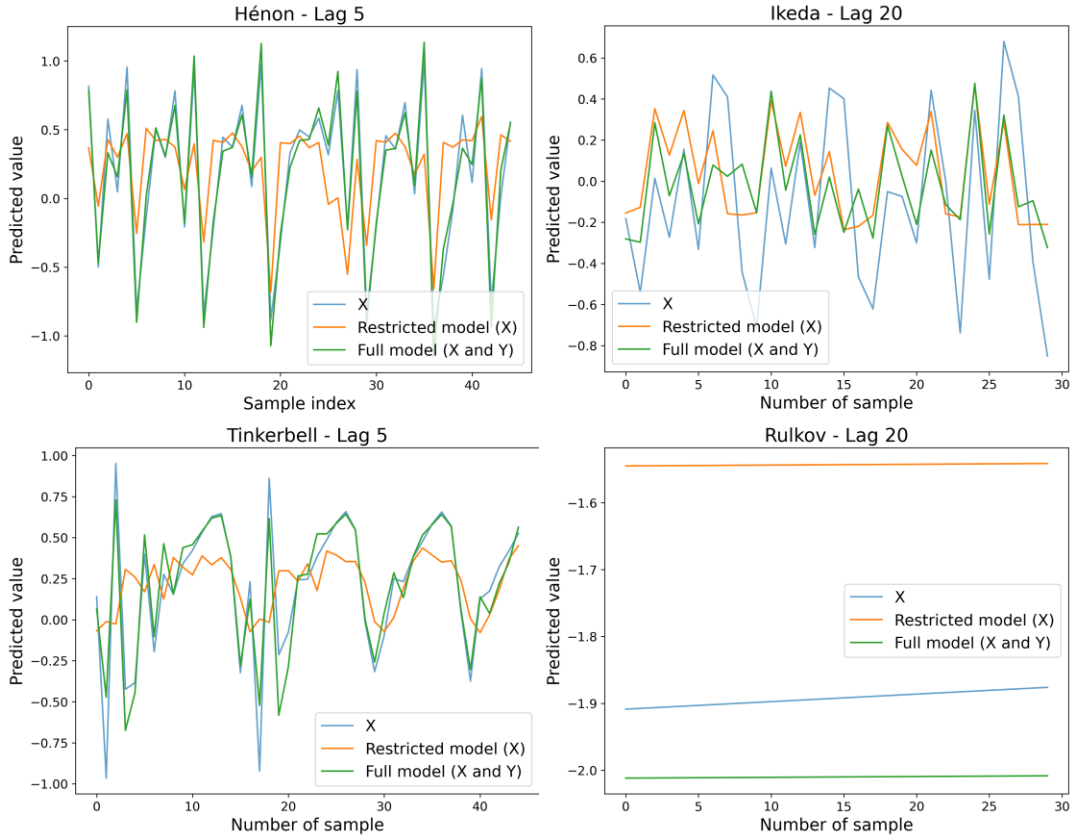


Fig. 2. Predicted values of the time series for four chaotic maps: Hénon (top-left, lag = 5), Ikeda (top-right, lag = 20), Tinkerbell (bottom-left, lag = 5), and Rulkov (bottom-right, lag = 20). Each panel compares the actual values of $X$ (blue line) with predictions from two models: the restricted model, based solely on $X$ (orange line), and the full model, based on both $X$ and $Y$ (green line).

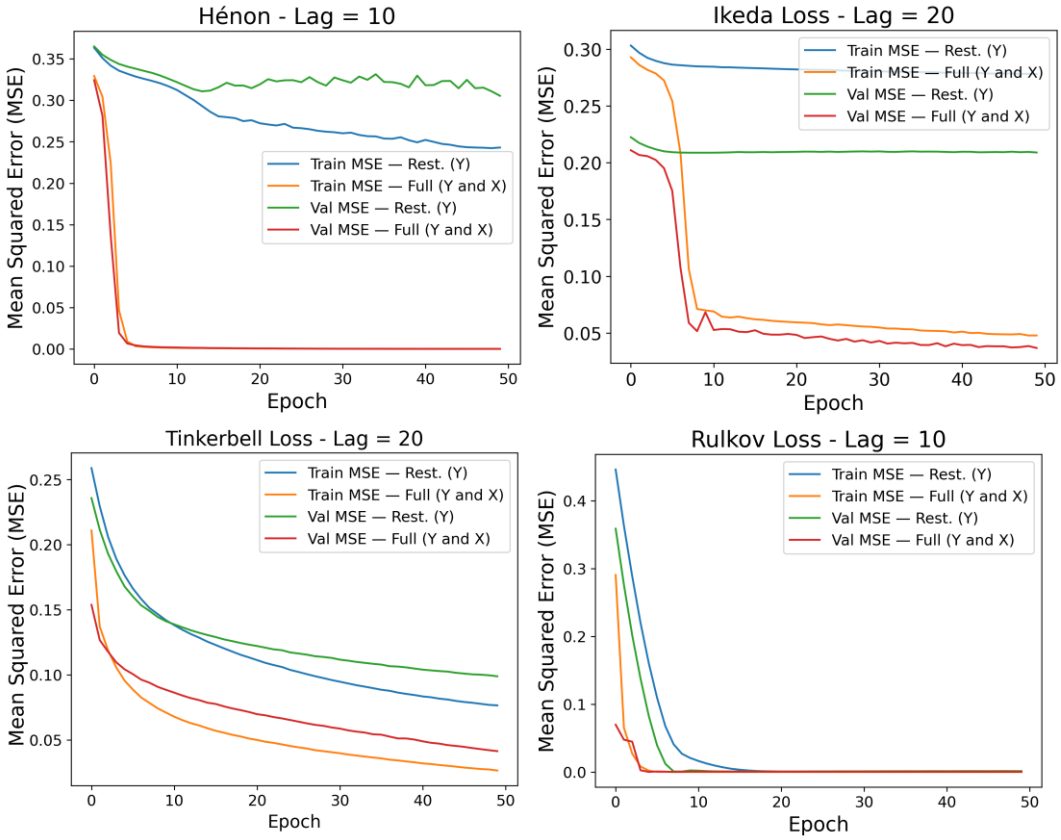


Fig. 3. MSE loss curves for chaotic maps: Hénon (top-left), Ikeda (top-right), Tinkerbell (bottom-left), and Rulkov (bottom-right). Each plot shows the evolution of the training and validation Mean Squared Error (MSE) for two models: the restricted model, based only on Y, and the full model, based on both Y and X.

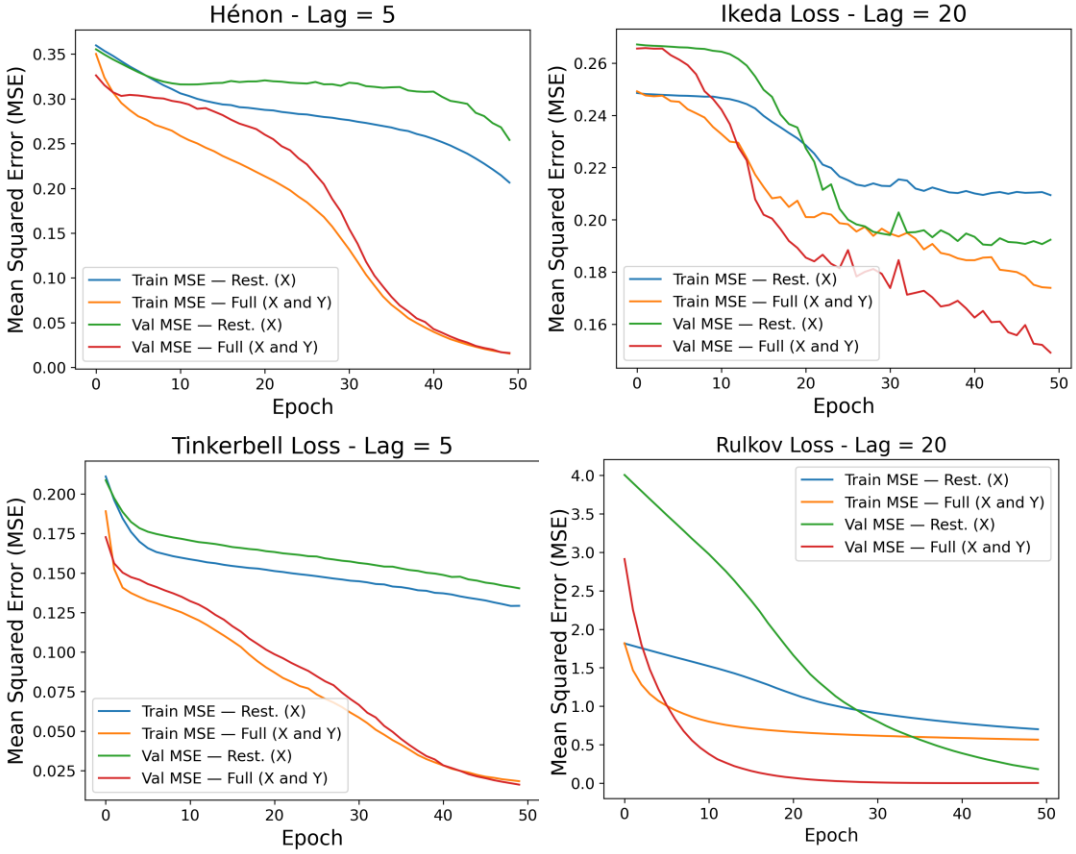


Fig. 4. MSE loss curves for chaotic maps: Hénon (top-left), Ikeda (top-right), Tinkerbell (bottom-left), and Rulkov (bottom-right). Each plot shows the training and validation Mean Squared Error (MSE) for two models: the restricted model, based only on X, and the full model, based on both X and Y.

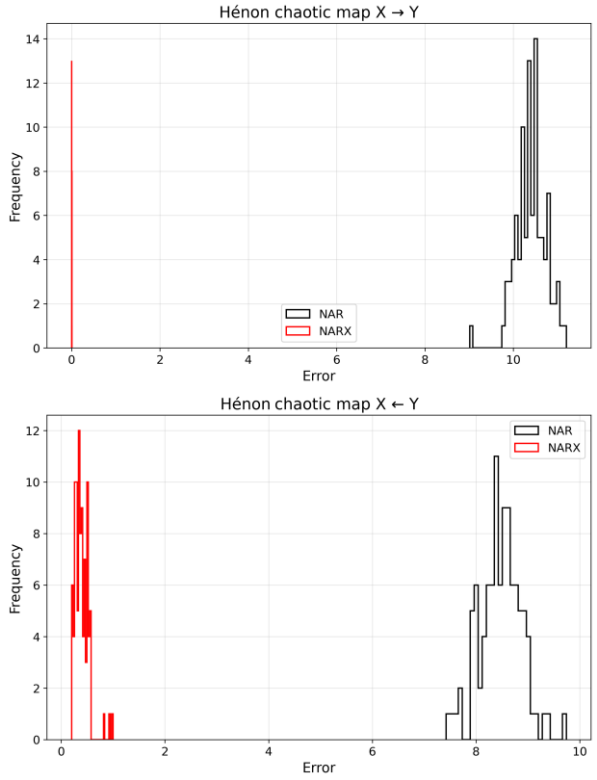


Fig. 5. Comparison of the errors when NAR and NARX models were applied to the Hénon time series map. Histogram corresponds to the errors for trainings of the model with characteristics in Table VI.

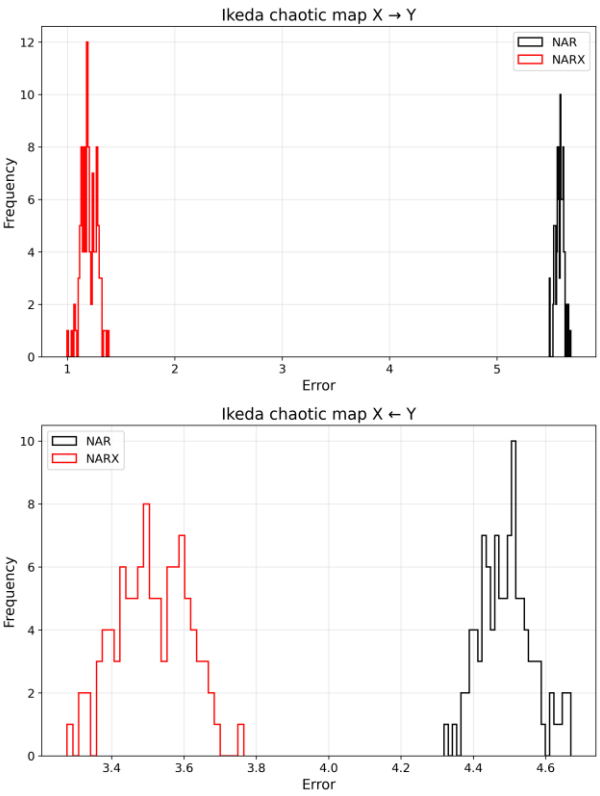


Fig. 6. Comparison of the errors when NAR and NARX models were applied to the Ikeda time series map. Histogram corresponds to the errors for trainings of the model with characteristics in Table VI.

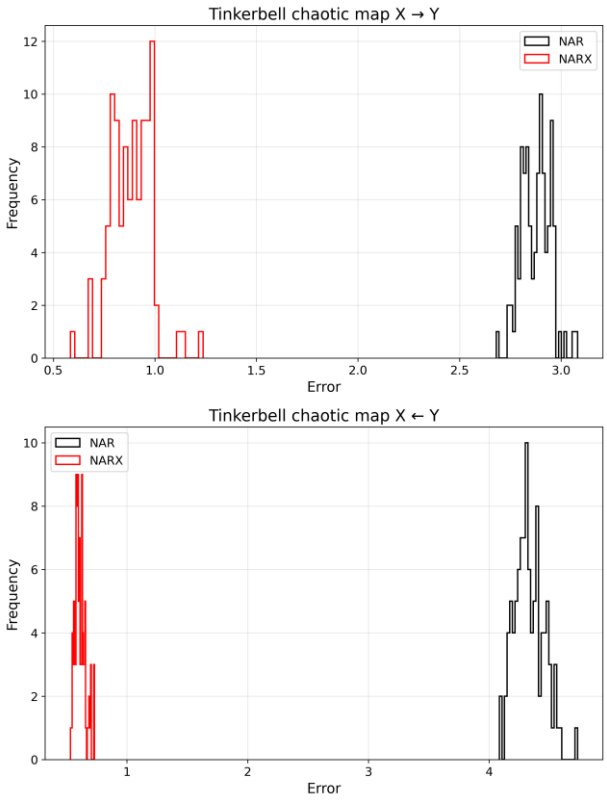


Fig. 7. Comparison of the errors when NAR and NARX models were applied to the Tinkerbell time series map. Histogram corresponds to the errors for trainings of the model with characteristics in Table VI.

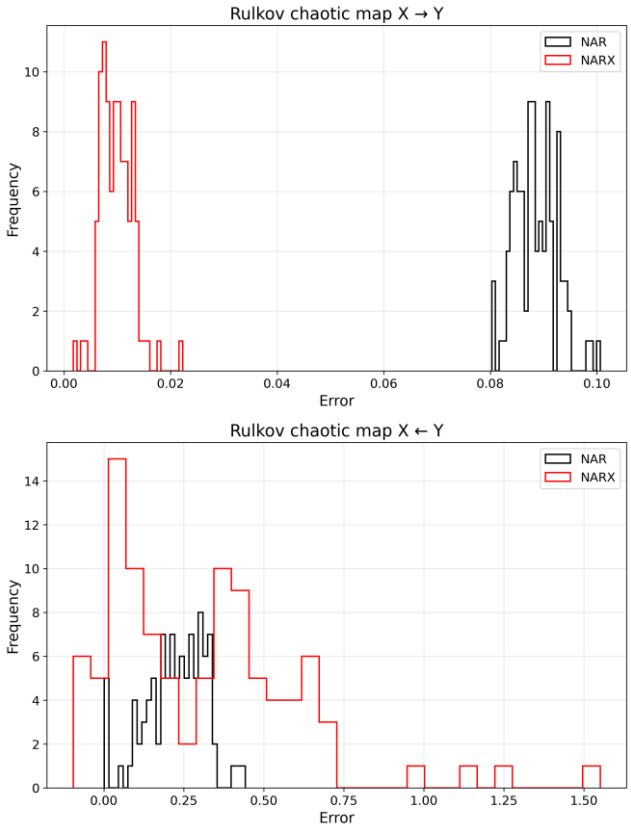


Fig. 8. Comparison of the errors when NAR and NARX models were applied to the Rulkov time series map. Histogram corresponds to the errors for trainings of the model with characteristics in Table VI.

The consistency of these findings is further supported by additional analyses, which indicate significant information flow in the Ikeda and Tinkerbell maps at all tested lags, partial flow for Hénon only at lag 5, and no significant flow for Rulkov. These observations are consistent with the results reported in Table I and corroborated by the neural network findings. Although the linear and nonlinear methods differ in their modeling approaches, both converge on the conclusion that the Ikeda and Tinkerbell systems exhibit robust bivariate dependence across multiple lag lengths, while Hénon shows a strong effect only at shorter lags from Y to X, and Rulkov does not exhibit consistent flow under the tested conditions, except for the GRU neural network architecture, which exhibited a significant error drop when both X and Y time series were included.

From the analysis summarized in Table VI, several broader themes emerge regarding the performance of the best models across all chaotic maps. Notably, there is no universal architecture that consistently outperforms others across all chaotic systems. The MLP frequently serves as an excellent baseline for univariate forecasting of X, while LSTM models often excel when Y is included. The performance of GRU models was comparable to LSTM approach, but appeared more sensitive to hyperparameter tuning. Optimal lag with 20-lag settings often performing best for the Hénon and Ikeda maps, while Tinkerbell and Rulkov maps sometimes required fewer lags, reflecting the unique complexities of each chaotic system. Furthermore, the inclusion of Y generally improved forecasting results, as indicated by lower residual sum of squares (RSS) compared to univariate modeling of X alone. This suggests that cross-series information, which may capture nonlinear Granger causality, is valuable, although the magnitude of improvement varied, most dramatically in the Hénon map, moderately in the Ikeda map, and still beneficial, but smaller, in the Tinkerbell and Rulkov maps.

From the analysis summarized in Table VI, several broader themes emerge regarding the performance of the best models across all chaotic maps. Notably, there is no universal architecture that consistently outperforms others across every system examined. For the Hénon map, LSTM networks with around 100 neurons stand out for both X→Y and X←Y predictions, though the optimal lag counts (10 versus 5) and batch sizes (16 versus 32) vary slightly. Meanwhile, Ikeda appears to benefit mostly from LSTM in the X→Y configuration and MLP in the X←Y case, although in both instances a reasonably large number of neurons and 20 lags often yield the lowest errors. On the other hand, Tinkerbell exhibits its best performance using an MLP for X→Y and an LSTM for X←Y, each leveraging about 100 neurons and either 5 or 20 lags. Lastly, Rulkov, which tends to confound purely linear methods, is best captured by a GRU model (100 neurons, 10 lags, batch size 32) for X→Y and an LSTM with a more modest neuron count (10) for X←Y predictions. Taken together, these findings underscore that fine-tuning model architectures, neuron counts, lags, and batch sizes in response to each map's unique nonlinear characteristics is essential for uncovering causal relationships in complex time series. Furthermore, the inclusion of both X and Y generally improved forecasting results, as indicated by lower residual sum of squares (RSS) compared to univariate modeling of X or Y alone. This suggests that cross-series information, which may capture nonlinear Granger causality, is valuable, although the magnitude of improvement varied, most dramatically in the Hénon map, moderately in the Ikeda map, and still beneficial, but smaller, in the Tinkerbell and Rulkov maps.

These performance variations also highlight the importance of architectural design choices when modeling different dynamical systems. The implementation of single hidden layer architectures for the MLP, LSTM, and GRU models represented a deliberate trade-off, balancing model simplicity against modeling power. This decision prioritized reduced overfitting risk and computational efficiency, aligning with our initial aim for 'low error and low complexity.' While deeper architectures can undoubtedly capture more intricate hierarchical patterns, our results suggest that a single-layer configuration provided an effective framework for the causality tests in this study, successfully balancing predictive performance with manageable model complexity. This approach serves as a robust baseline, and future investigations could explore the potential benefits of more complex, multi-layered architectures should the need for enhanced modeling capacity or broader generalization arise.

In this context, these findings also have important implications for real-world biomedical applications of NNGC. Previous work analyzing brain–heart interactions in sleep-related disorders, such as obstructive sleep apnea, has demonstrated that Granger causality–based approaches can reveal meaningful directional relationships between EEG and ECG-derived signals, both before and after therapeutic intervention [5]. This work shows that changes in functional connectivity are closely related to physiological regulation mechanisms and clinical outcomes. From this perspective, the results presented here emphasize that NNGC estimates are highly sensitive to architectural and training parameters such as the number of hidden neurons and the selected number of lags, which can substantially influence the inferred connectivity patterns. This sensitivity is particularly critical in biomedical contexts, where suboptimal parameter choices may lead to misleading interpretations of physiological interactions and, consequently, incorrect clinical conclusions.

Moreover, the same study applying NNGC to EEG and heart rate variability signals in patients with obstructive sleep apnea reported that these methods are capable of detecting treatment-related changes in brain–heart connectivity [5]. However, such applications inherently depend on specific architectural and training choices that are often fixed a priori and not systematically analyzed. The present work complements this prior contribution by offering a controlled evaluation of how neural network parameters affect NNGC estimation, thereby strengthening the methodological foundations for future case studies based on real-world biomedical signals. In this sense, the insights presented in this work can directly guide the design and interpretation of NNGC analyses beyond synthetic settings.

Beyond architectural considerations, it is important to emphasize the broader relevance of Granger Causality (GC) in the analysis of real-world, high-dimensional datasets. GC remains a valuable tool for identifying directional relationships in temporal data and has been widely applied in domains such as EEG [20], finance [2], and climate science [21]. For example, it helps reveal functional brain connectivity from EEG recordings, uncovers lead-lag dynamics in stock markets, and detects causal interactions between geophysical variables. However, applying GC in these contexts also presents challenges, including high noise levels, non-stationary behavior, and partial observability of the underlying

systems. These factors can lead to unreliable or spurious causal inferences if not properly addressed. To mitigate such issues, researchers often incorporate robust preprocessing steps or extend the traditional GC framework using nonlinear methods or data-driven models, such as neural network-based GC approaches, as explored in this study.

Another important consideration arises from the trends observed in Figs. 3 and 4, it can be inferred that increasing the number of epochs would likely lead to a further reduction in prediction error. This is consistent with the general behavior of neural networks, where additional training epochs allow the model to better optimize its parameters and minimize residual errors. However, in this study, the number of epochs was fixed to a specific value due to the high computational cost associated with training the models, especially given the number of configurations tested. While this constraint ensured a manageable runtime, it leaves open the possibility that further improvements in accuracy could be achieved with extended training. Exploring this trade-off between computational time and error reduction in future work could provide valuable insights into the optimal training strategies for nonlinear Granger causality detection.

In addition, the results summarized in Tables II to VI highlight the strong relationship between model configuration, computational cost, and prediction error in NNGC analyses. In this study, the computational burden is primarily driven by the three hyperparameters explicitly explored: the number of hidden neurons, the number of lags, and the batch size. Increasing the number of hidden neurons and lags generally increases the computational cost, as it enlarges both the number of trainable parameters and the dimensionality of the input space. However, the results indicate that the reduction in prediction error reported in the last column of Tables II to VI is not monotonic and strongly depends on the underlying chaotic system and the selected neural network architecture. In several cases, the lowest error values were obtained with moderate numbers of neurons or lags rather than with the most complex configurations. While other training parameters, such as the number of epochs and the learning rate, were fixed in this work to ensure tractable experimentation, they remain relevant factors that can further influence convergence behavior and computational cost. Overall, these observations suggest that an optimal convergence regime for NNGC models is system-dependent and requires a careful balance between model complexity, training effort, and error minimization, rather than a uniform increase in hyperparameter values.

Furthermore, these findings highlight the need for further exploration of neural network models and a deeper investigation into various parameters, as these details are often underreported in the existing literature. By systematically varying model architectures and hyperparameters, researchers can uncover additional insights into the dynamics of chaotic systems and enhance the understanding of causal relationships. This ongoing exploration is crucial for refining predictive models and improving their applicability to real-world scenarios, where the complexities of nonlinear interactions are prevalent.

## V. CONCLUSIONS

Neural networks are highly sensitive to tuning parameters such as the number of neurons, lag length, and batch size, which critically affect their ability to learn complex dynamics in nonlinear Granger causality (NNGC). The results indicate that each chaotic system requires a specific approach, as forecasting accuracy and the reliability of NNGC depend on balancing network complexity with data characteristics. As a promising extension, exploring architectures such as Transformers could further improve time-series modeling by reducing sensitivity to lag selection. Through self-attention mechanisms, Transformers can capture long-range temporal dependencies without explicitly defining large lag windows, potentially improving robustness in modeling nonlinear dynamics and extending data-driven Granger causality methods.

Furthermore, future work should address the sensitivity of NNGC to modeling choices in multivariate Granger causality settings. In multivariate systems, causal inference must be formulated using conditional Granger causality, where interactions are assessed while conditioning on the remaining variables to account for indirect effects and common drivers. Extending NNGC to this framework requires a careful analysis of how hyperparameters, embedding strategies, and network complexity influence the estimation of conditional causal interactions. This issue is particularly relevant in physiological applications, where brain–heart dynamics are inherently multivariate and causal relationships arise at the network level rather than from isolated pairwise interactions.

## REFERENCES

[1] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Jul. 1969, doi: 10.2307/1912791.

[2] R. P. Maradana *et al*., "Innovation and economic growth in European Economic Area countries: The Granger causality approach," *IIMB Manag. Rev.*, vol. 31, no. 3, pp. 268–282, 2019, doi: 10.1016/j.iimb.2019.03.002.

[3] S. Hu *et al.*, "More discussions for granger causality and new causality measures," *Cogn. Neurodyn.*, vol. 6, no. 1, pp. 33–42, Feb. 2012, doi: 10.1007/s11571-011-9175-8.

[4] M. Rosoł, M. Młyńczak, and G. Cybulski, "Granger causality test with nonlinear neural-network-based methods: Python package and simulation study," *Comput. Methods Programs Biomed.*, vol. 216, p. 106669, 2022, doi: 10.1016/j.cmpb.2022.106669.

[5] A. D. Orjuela-Cañón, A. Cerquera, J. A. Freund, G. Juliá-Serdá, and A. G. Ravelo-García, "Sleep apnea: Tracking effects of a first session of CPAP therapy by means of Granger causality," *Comput. Methods Programs Biomed.*, vol. 187, p. 105235, Apr. 2020, doi: 10.1016/j.cmpb.2019.105235.

[6] N. Nicolaou and T. G. Constandinou, "A nonlinear causality estimator based on non-parametric multiplicative regression," *Front. Neuroinformatics*, vol. 10, p. 19, 2016, doi: 10.3389/fninf.2016.00019.

[7] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Phys. Rev. Lett.*, vol. 100, no. 14, Apr. 2008, doi: 10.1103/physrevlett.100.144103.

[8] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, Jul. 2000, doi: 10.1103/physrevlett.85.461.

[9] A. Tank, I. Covert, N. Foti, A. Shojaie, and E. B. Fox, "Neural granger causality," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4267–4279, Aug. 2022, doi: 10.1109/TPAMI.2021.3065601.

[10] A. Montalto, L. Faes, and D. Marinazzo, "MuTE: a MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy," *PloS One*, vol. 9, no. 10, p. e109462, 2014, doi: 10.1371/journal.pone.0109462.

[11] M. Hénon, "A Two-dimensional Mapping with a Strange Attractor," in *The Theory of Chaotic Attractors*, B. R. Hunt, J. A. Kennedy, T. Sauer, and J. A. Yorke, Eds. New York, NY: Springer New York, 1976, pp. 94–102. doi: 10.1007/978-0-387-21830-4_8.

[12] K. Ikeda, H. Daido, and O. Akimoto, "Optical turbulence: chaotic behavior of transmitted light from a ring cavity," *Phys. Rev. Lett.*, vol. 45, no. 9, pp. 709–712, Sep. 1980, doi: 10.1103/physrevlett.45.709.

[13] A. Attanasio and U. Triacca, "Detecting human influence on climate using neural networks based Granger causality," *Theor. Appl. Climatol.*, vol. 103, no. 1–2, pp. 103–107, Jan. 2011, doi: 10.1007/s00704-010-0285-8.

[14] A. Montalto, S. Stramaglia, L. Faes, G. Tessitore, R. Prevete, and D. Marinazzo, "Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality," *Neural Netw.*, vol. 71, pp. 159–171, 2015, doi: 10.1016/j.neunet.2015.08.003.

[15] N. Ancona, D. Marinazzo, and S. Stramaglia, "Radial basis function approach to nonlinear Granger causality of time series," *Phys. Rev. E*, vol. 70, no. 5, Nov. 2004, doi: 10.1103/physreve.70.056221.

[16] Y. Wang *et al.*, "Estimating brain connectivity with varying-length time lags using a recurrent neural network," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1953–1963, 2018, doi: 10.1109/TBME.2018.2842769.

[17] B. J. Koch, T. Sainburg, P. Geraldo Bastías, S. Jiang, Y. Sun, and J. G. Foster, "A Primer on Deep Learning for Causal Inference," *Sociol. Methods Res.*, vol. 54, no. 2, pp. 397–447, May 2025, doi: 10.1177/00491241241234866.

[18] J.-S. Hwang, S.-S. Lee, J.-W. Gil, and C.-K. Lee, "Determination of Optimal Batch Size of Deep Learning Models with Time Series Data," *Sustainability*, vol. 16, no. 14, p. 5936, Jul. 2024, doi: 10.3390/su16145936.

[19] Josef Perktold *et al.*, *statsmodels/statsmodels: Release 0.14.2*. (Apr. 17, 2024). Zenodo. doi: 10.5281/ZENODO.593847

[20] G. Deshpande and X. Hu, "Investigating effective brain connectivity from fMRI data: past findings and current issues with reference to Granger causality analysis," *Brain Connect.*, vol. 2, no. 5, pp. 235–245, 2012, doi: 10.1089/brain.2012.0091.

[21] S. Tripaldi, S. Scippacercola, A. Mangiacapra, and Z. Petrillo, "Granger causality analysis of geophysical, geodetic and geochemical observations during volcanic unrest: A case study in the campi flegrei caldera (Italy)," *Geosciences*, vol. 10, no. 5, p. 185, 2020, doi: 10.3390/geosciences10050185.

**Diego N. Chacón Wilches** received the B.Sc. degree in Biomedical Engineering from Universidad del Rosario in collaboration with Universidad Escuela Colombiana de Ingeniería Julio Garavito, Bogotá, Colombia, in 2024. He is currently pursuing the M.Sc. degree in Data Science at Universidad Escuela Colombiana de Ingeniería Julio Garavito, which he began in 2025. As part of his academic formation, he completed an international exchange program in Applied Computer Science at Ruhr University Bochum, Germany, where he deepened his knowledge in deep learning and data science. His experience focuses on the application of data science and artificial intelligence in biomedical and interdisciplinary contexts. His research interests include artificial intelligence, neural networks, and biomedical engineering.

**Alvaro D. Orjuela-Cañón** (StM'00–M'06–SM'17) received a B.Sc. in Electronic Engineering from Universidad Distrital Francisco José de Caldas in 2006, and M.Sc. and Ph.D. degrees from the Federal University of Rio de Janeiro (UFRJ), Brazil, in 2009 and 2015, respectively. He is currently part of the Biomedical Engineering program at the School of Medicine and Health Sciences, Universidad del Rosario, Bogotá. His research interests include biomedical signal processing, computational intelligence in healthcare, and alternative energy sources. He has been an IEEE member for almost 25 years and is an active member of the IEEE Computational Intelligence Society.