ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

ChaosNet-Transformer: A Physics-Informed Deep Learning Framework for Predicting and Synchronizing High-Dimensional Chaotic Dynamical Systems

Wanayumini*, Sabir Ramadhan, Miftahul Jannah, Arya Agung Pratama, Tiara Rahmadhani Samosir, Novi Rahmawati

Faculty of Engineering, Informatics Engineering, Universitas Asahan, Sumatera Utara, Indonesia
Email: [1,*]wanayumini@yahoo.co.id, [2]ramasabir@gmail.com, [3]miftahhstudy@gmail.com
Correspondence Author Email: wanayumini@yahoo.co.id

Abstract−Predicting and synchronizing high-dimensional chaotic dynamical systems remains a fundamental challenge in artificial intelligence and nonlinear science because chaotic trajectories exhibit strong sensitivity to initial conditions and rapid error divergence. This study proposes ChaosNet-Transformer (CNT), a physics-informed deep learning framework that combines Lyapunov-regularized temporal attention, Reservoir-Enhanced Embedding, and a Koopman Linearization Module to improve chaotic time-series prediction while preserving the underlying dynamical structure. CNT is evaluated on six benchmark systems, namely Lorenz-63, Rössler, Mackey-Glass, Chen-Lu, Qi Hyperchaos, and the Kuramoto-Sivashinsky system. The proposed model is compared with Echo State Network (ESN), Next-Generation Reservoir Computing (NG-RC), FEDformer, and Chaos-Informed LSTM (CI-LSTM) using Valid Prediction Time (VPT) measured in Lyapunov Time. Experimental results show that CNT achieves the highest VPT across all benchmark systems and consistently outperforms the strongest baseline methods. The synchronization experiment further demonstrates that CNT produces lower synchronization error than ESN and FEDformer under comparable coupling conditions. Ablation results confirm that Reservoir-Enhanced Embedding, Lyapunov regularization, and Koopman-based latent linearization each contribute to the model performance. These findings indicate that integrating Transformer-based sequence modeling with physics-informed constraints can improve the prediction and synchronization of chaotic dynamical systems.

Keywords: Chaotic Systems; Deep Learning; Lyapunov Exponents; Koopman Operator; Reservoir Computing; Transformer Architecture; Time Series Prediction; Chaos Synchronization

1. INTRODUCTION

Chaos theory, pioneered by [1] and [2] describes the behavior of deterministic dynamical systems whose evolution is highly sensitive to initial conditions. Although chaotic systems are deterministic in nature, they exhibit trajectories that diverge exponentially over time, making long-term prediction extremely difficult and, in many cases, seemingly impossible (Lyapunov, 1992;Lin et al., 2023). This sensitivity is commonly characterized by the Lyapunov exponent, which quantifies the average rate of separation between neighboring trajectories in phase space. A positive maximum Lyapunov exponent (MLE > 0) is widely regarded as a fundamental indicator of chaotic behavior [5]. In high-dimensional dynamical systems, this problem becomes even more complex because small perturbations in the initial state can rapidly propagate across multiple interacting variables, resulting in significant deviations in the future evolution of the system.

Chaotic systems are commonly characterized by sensitive dependence on initial conditions, topological mixing, and dense periodic orbits [3]. The Lorenz system [1], which was originally derived from atmospheric convection, has become one of the canonical benchmarks in the study of nonlinear dynamics. Its strange attractor in $\mathbb{R}^3$ has a fractal dimension of approximately $D \approx 2.06$ and a maximum Lyapunov exponent of approximately $\lambda_1 \approx 0.905$. Other chaotic systems, such as the Rössler and Chen-Lu systems, extend this analysis to different attractor geometries, while the Qi hyperchaotic system extends the complexity into a four-dimensional setting with two positive Lyapunov exponents [6]. These characteristics show that chaotic systems are not only difficult to predict numerically, but also challenging to model because their geometric structure, stability properties, and long-term dynamical behavior must be preserved (Sun et al., 2023; Cheng et al., 2023).

The integration of deep learning and nonlinear dynamics has emerged as one of the frontiers of computational science over the past decade [9], [10], [11], [12]. Long Short-Term Memory (LSTM) networks [13] and reservoir computing [14] have demonstrated remarkable empirical performance in short- to medium-horizon chaotic prediction. Reservoir computing, particularly Echo State Networks (ESN) [14] and Liquid State Machines [15], pioneered the application of recurrent neural network structures to chaotic time-series modeling. [16] demonstrated the prediction of the Lorenz system up to 8× Lyapunov time, showing the ability of reservoir-based methods to capture temporal dependencies in chaotic dynamics. [17] established a rigorous benchmark comparing LSTM, data-driven methods, and reservoir computing approaches. Next-generation reservoir computing (NG-RC) [18] further simplified this paradigm by eliminating the reservoir entirely and using polynomial feature maps of the input with a linear readout.

Despite these achievements, existing architectures still face fundamental limitations. LSTM networks and reservoir computing approaches generally treat chaotic dynamics as arbitrary time series, thereby overlooking the rich geometric structure of strange attractors. In addition, most of these models lack a principled mechanism for encoding physical constraints such as energy conservation, Lyapunov stability, and the operator structure underlying nonlinear system evolution. This limitation is critical because chaotic prediction should not only minimize numerical error over short horizons, but also preserve the intrinsic dynamical properties of the original system. Without such physical and

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 120
This work is licensed under a Creative Commons Attribution 4.0 International License.

ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

geometric awareness, a model may produce low short-term error while gradually drifting away from the true attractor structure.

Physics-Informed Neural Networks (PINNs) [19] address part of this issue by embedding differential equation constraints as additional loss terms. Neural ordinary differential equations [20] model continuous-time dynamics using neural networks, while Hamiltonian neural networks [21] encode physical principles such as energy conservation into the learning process. However, these approaches generally require explicit knowledge of the governing equations. In many real-world systems, the governing equations may be partially unknown, difficult to derive analytically, or unavailable due to observational limitations. Therefore, a more flexible approach is needed, one that can incorporate physical awareness and nonlinear dynamical structure without relying entirely on explicitly known governing equations.

The Transformer architecture (Vaswani et al., 2017; Y. Li et al., 2024; Cao et al., 2024), with its global self-attention mechanism, offers a fundamentally different inductive bias and naturally captures long-range dependencies. Recent models, including Informer [25], FEDformer [26], and PatchTST [27], have successfully leveraged this property for general time-series forecasting. However, the application of Transformer-based architectures to chaotic systems remains structurally limited by the absence of physics-aware inductive biases specifically designed for nonlinear dynamics. In other words, although Transformers have strong representational capacity, they still require additional mechanisms to preserve Lyapunov stability, attractor geometry, and the operatorial structure of chaotic systems (Lin et al., 2024;Liu et al., 2024).

One theoretical foundation that can bridge nonlinear dynamics and linear representation is Koopman operator theory. The Koopman operator K lifts nonlinear dynamics into an infinite-dimensional linear space acting on observable functions (Mezic, 2005; Shi & Meng, 2022). Deep Koopman methods, such as DeepKoopman [32] and KoopmanAE [33], learn finite-dimensional approximations through autoencoder-based representations. This framework enables nonlinear dynamics to be represented in a latent space where the evolution can be approximated by a linear operator. However, the systematic integration of Koopman-based representations with Transformer attention and Lyapunov-aware learning remains underexplored, particularly for prediction and synchronization in high-dimensional chaotic dynamical systems.

Based on these limitations, this study proposes ChaosNet-Transformer (CNT), a physics-informed deep learning framework for the prediction and synchronization of high-dimensional chaotic dynamical systems. The purpose of this research is to develop a Transformer-based architecture that not only learns temporal patterns from chaotic data, but also incorporates physical and dynamical constraints into the learning process. The proposed solution integrates Lyapunov-regularized attention, a Koopman Linearization Module, and Reservoir-Enhanced Embedding within a unified architecture. This design is expected to improve prediction accuracy while maintaining the stability and structural consistency of the underlying chaotic attractor.

The main contribution of this study is the introduction of ChaosNet-Transformer as a Transformer-based architecture equipped with a Lyapunov-regularized attention mechanism. In addition, this study proposes the Koopman Linearization Module (KLM), which learns a finite-dimensional approximation of the infinite-dimensional Koopman operator. This research also designs a Reservoir-Enhanced Embedding (REE) layer that processes input sequences through a fixed echo-state reservoir to enrich nonlinear temporal representations. Finally, this study provides an extensive benchmark on six canonical chaotic systems with a complete ablation analysis to evaluate the contribution of each architectural component to prediction and synchronization performance.

2. RESEARCH METHODOLOGY

2.1 Basic Research Framework

This study is a computational experimental research that focuses on the development and evaluation of a physics-informed deep learning model for the prediction and synchronization of high-dimensional chaotic dynamical systems. The research objects are canonical chaotic dynamical systems used as benchmarks, while the research data are obtained from numerical simulations of those systems. Since this study is fully computational and does not involve human participants, there are no respondents in the conventional survey-based sense. The research setting refers to the computational environment used to generate chaotic time-series data, train the proposed model, evaluate multi-step prediction performance, and analyze the synchronization behavior of the learned dynamics.

The basic research framework begins with the generation of time-series data from chaotic dynamical systems. The generated data are then preprocessed and transformed into input-output sequences based on a predefined temporal window. These sequences are used to train the proposed ChaosNet-Transformer model. After the training process, the model is evaluated using multi-step prediction scenarios, Lyapunov-based stability analysis, Koopman operator consistency, and synchronization performance analysis. The main variables in this study include the system state $x(t)$, the noisy observation $y(t)$, the sampling interval Δt, the prediction horizon H, and the trainable parameters of the proposed model.

The conceptual framework of this research places the chaotic system as the source of nonlinear dynamical data, while ChaosNet-Transformer acts as the predictive model that learns temporal dependencies and latent dynamical structures from the observed sequences. The Reservoir-Enhanced Embedding component enriches the nonlinear

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 121
This work is licensed under a Creative Commons Attribution 4.0 International License.

ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

representation of the input sequence through a fixed echo-state reservoir. The Lyapunov-regularized attention mechanism encourages the model to maintain stable dynamical behavior during training. Meanwhile, the Koopman Linearization Module maps nonlinear dynamics into a latent representation where the system evolution can be approximated by a linear operator. Through this framework, the proposed model is expected to achieve accurate prediction while preserving the intrinsic dynamical characteristics of chaotic systems.

2.2 Problem Formulation

Let $x(t) \in \mathbb{R}^d$ denote the state of a chaotic dynamical system. The evolution of the system is governed by the following differential equation:

$$\frac{dx(t)}{dt} = F(x(t); \theta) \quad (1)$$

where $F: \mathbb{R}^d \to \mathbb{R}^d$ represents a nonlinear vector field parameterized by θ. In this study, the observed data are assumed to contain noise. Therefore, the noisy observation is defined as follows:

$$y(t) = x(t) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2 I) \quad (2)$$

where $y(t)$ denotes the observed state, $x(t)$ denotes the true system state, and ε represents Gaussian noise with zero mean and covariance $\sigma^2 I$. Given a sequence of N noisy observations sampled at a time interval Δt, the objective of this study is to learn a predictive model f_θ that minimizes multi-step prediction error across the prediction horizon H, while preserving the Lyapunov spectrum of the underlying attractor.

This formulation emphasizes that chaotic prediction should not be evaluated only by short-term numerical accuracy. A model must also maintain the dynamical structure of the original system so that the predicted trajectory does not rapidly drift away from the true attractor. Therefore, the training objective is designed to combine prediction loss, Lyapunov regularization, and Koopman-based consistency loss. This combination allows the model to learn both the observable temporal patterns and the latent physical structure of the chaotic system.

2.3 ChaosNet-Transformer Architecture

ChaosNet-Transformer is designed as a Transformer-based architecture enhanced with additional components that capture the nonlinear and physics-informed characteristics of chaotic systems. The architecture consists of Reservoir-Enhanced Embedding, Lyapunov-regularized multi-head attention, and the Koopman Linearization Module. These three components work together to produce temporal representations that are expressive, dynamically stable, and consistent with the structure of nonlinear systems.

The Reservoir-Enhanced Embedding (REE) layer computes a nonlinear expansion through a randomly initialized and fixed echo-state reservoir with $N_r = 512$ internal units. The reservoir state is updated using the following equation:

$$h(t) = \tanh\left(W_r h(t-1) + W_{in} y(t)\right) \quad (3)$$

where $h(t)$ denotes the reservoir state at time t, $W_r \in \mathbb{R}^{N_r \times N_r}$ is the reservoir weight matrix with spectral radius $\rho(W_r) = 0.9$, and $W_{in} \in \mathbb{R}^{N_r \times d}$ is the input coupling matrix. Both matrices are drawn from sparse Gaussian distributions with a density of 0.05. The combined embedding is then projected into the model dimension $D = 512$. This embedding mechanism enables the model to capture nonlinear temporal patterns before the sequence is processed by the Transformer encoder.

The core innovation of CNT lies in its Lyapunov-regularized multi-head attention mechanism. The standard scaled dot-product attention [22] is computed as follows:

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_k}}\right)V \quad (4)$$

where Q, K, and V represent the query, key, and value matrices, respectively, while D_k denotes the dimensionality of the key vectors. In this study, the training objective is augmented with a Lyapunov regularizer $\mathcal{L}_{LE}$ that penalizes attention weight matrices showing exponential divergence in their singular value spectrum. The total loss function is formulated as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{pred} + \alpha \mathcal{L}_{LE} + \beta \mathcal{L}_{Koopman} \quad (5)$$

where $\mathcal{L}_{pred}$ represents the prediction loss, $\mathcal{L}_{LE}$ denotes the Lyapunov regularization loss, and $\mathcal{L}_{Koopman}$ represents the Koopman consistency loss. The hyperparameters $\alpha = 0.01$ and $\beta = 0.1$ are determined through grid search. This mechanism encourages the attention process to remain sensitive to temporal dependencies while avoiding unstable divergence that may distort the learned chaotic dynamics.

The Koopman Linearization Module (KLM) approximates the Koopman operator $K: \Phi \to \Phi$, which acts on the space of observables. The encoder network $g_\phi: \mathbb{R}^D \to \mathbb{R}^M$, with $M = 128$, maps Transformer embeddings into Koopman eigenfunctions. In the Koopman latent space, the dynamics evolve as a linear map:

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 122
This work is licensed under a Creative Commons Attribution 4.0 International License.

ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

where $z(t)$ represents the Koopman latent representation and $\widehat{K}$ denotes the learned finite-dimensional approximation of the Koopman operator. The decoder $h_\psi: \mathbb{R}^M \to \mathbb{R}^D$ reconstructs the embedding from the Koopman latent representation. The Koopman loss is defined as follows:

$$\mathcal{L}_{Koopman} = \| z(t+1) - \widehat{K}z(t) \|^2 + \| h_\psi(z(t)) - e(t) \|^2 \quad (6)$$

where $e(t)$ denotes the output of the Transformer encoder. The first term enforces linear consistency in the Koopman latent space, while the second term ensures that the reconstructed embedding remains close to the Transformer representation. This loss enforces consistency between the nonlinear Transformer representation and the linear evolution in Koopman latent space.

Overall, the proposed research method combines data-driven learning, nonlinear temporal representation, Lyapunov-aware stability control, and Koopman-based linearization. This methodological design is intended to overcome the limitations of conventional deep learning models that treat chaotic systems only as generic time series. By explicitly incorporating dynamical stability and operator-based structure, ChaosNet-Transformer is expected to provide a more robust framework for predicting and synchronizing high-dimensional chaotic systems.

3. RESULTS AND DISCUSSION

3.1 Benchmark Chaotic Systems

The proposed ChaosNet-Transformer (CNT) was evaluated on six benchmark chaotic systems representing different dynamical characteristics, dimensional complexity, Lyapunov spectra, and attractor geometries. The selected systems include Lorenz-63, Rössler, Mackey-Glass, Chen-Lu, Qi Hyperchaos, and the KS-PDE. These systems were chosen because they are widely used in nonlinear dynamics, chaotic prediction, and synchronization studies, and they provide a diverse testing environment for evaluating the robustness of the proposed architecture. The benchmark systems used in this study are presented in Table 1.

Table 1. Benchmark Chaotic Systems and Lyapunov Spectrum

System	Dim.	λ_1	λ_2	Fractal Dim.
Lorenz-63	3	0.906	≈0	2.06
Rössler	3	0.071	≈0	2.01
Mackey-Glass	∞(23)	0.006	—	≈2.1
Chen-Lu	3	2.012	≈0	2.08
Qi Hyperchaos	4	1.387	0.421	3.24
KS-PDE	64	0.094	…	≈4.8

As shown in Table 1, the selected benchmark systems provide different levels of prediction difficulty. Lorenz-63, Rössler, and Chen-Lu are three-dimensional chaotic systems with different attractor geometries and Lyapunov characteristics. Mackey-Glass represents a delay differential system with an effectively infinite-dimensional structure, making it useful for evaluating the model's ability to capture delayed nonlinear dependencies. Qi Hyperchaos introduces a higher level of instability because it has two positive Lyapunov exponents, indicating that divergence occurs in more than one direction of the phase space. The Kuramoto-Sivashinsky partial differential equation (KS-PDE) represents a high-dimensional spatiotemporal chaotic system, making it one of the most challenging benchmarks in this study.

The diversity of these systems is important because a model that performs well only on low-dimensional systems may not necessarily generalize to hyperchaotic or spatiotemporal chaotic systems. Therefore, the use of these six benchmark systems allows the evaluation to cover several important aspects, including short-term prediction accuracy, long-term trajectory stability, robustness to attractor geometry, sensitivity to Lyapunov instability, and scalability to higher-dimensional dynamics. This evaluation setting is aligned with the objective of this study, namely to develop a physics-informed deep learning architecture capable of predicting and synchronizing high-dimensional chaotic dynamical systems.

3.2 Comparison of Methods Based on Valid Prediction Time

The main quantitative evaluation was conducted using Valid Prediction Time (VPT), measured in Lyapunov Time (LT). VPT indicates how long a model can produce valid predictions before the predicted trajectory significantly diverges from the reference trajectory. Since chaotic systems naturally amplify small errors over time, VPT is more informative than one-step prediction error alone. A higher VPT indicates that the model is able to maintain predictive accuracy for a longer period relative to the intrinsic instability of the system.

Table 2 presents the VPT results across the six benchmark chaotic systems. The proposed CNT model was compared with four baseline methods, namely Echo State Network (ESN), Next-Generation Reservoir Computing (NG-RC), FEDformer, and Chaos-Informed LSTM (CI-LSTM). CI-LSTM refers to an LSTM-based baseline that incorporates chaos-related temporal features to improve chaotic time-series prediction. These baselines were selected

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 123
This work is licensed under a Creative Commons Attribution 4.0 International License.

ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

because they represent different families of sequence modeling approaches, including reservoir-based models, Transformer-based time-series forecasting, and recurrent neural networks enhanced with chaos-related information.

Table 2. Valid Prediction Time in Lyapunov Time

System	ESN	NG-RC	FEDformer	CI-LSTM	CNT (Ours)
Lorenz-63	6.8±0.3	8.1±0.2	7.3±0.2	7.8±0.3	11.2±0.2
Rössler	9.1±0.4	10.3±0.3	10.1±0.3	10.6±0.2	14.7±0.3
Mackey-Glass	12.4±0.6	14.1±0.4	13.8±0.4	14.5±0.3	19.3±0.4
Chen-Lu	4.2±0.3	5.1±0.2	5.0±0.2	5.3±0.2	7.6±0.2
Qi Hyperchaos	3.1±0.2	3.9±0.2	3.8±0.2	4.1±0.2	6.2±0.2
KS-PDE	5.6±0.4	6.8±0.3	7.1±0.3	7.4±0.3	9.8±0.3

The results in Table 2 show that CNT achieved the highest VPT on all six benchmark systems. On the Lorenz-63 system, CNT obtained a VPT of 11.2 LT, outperforming NG-RC with 8.1 LT and FEDformer with 7.3 LT. This result indicates that the proposed architecture is able to extend the prediction horizon substantially compared with both reservoir-based and Transformer-based baselines. On the Rössler system, CNT achieved 14.7 LT, while the best baseline, CI-LSTM, reached 10.6 LT. This improvement demonstrates that CNT is not only effective on strongly chaotic systems such as Lorenz-63, but also on systems with slower divergence rates such as Rössler.

For the Mackey-Glass system, CNT achieved the highest VPT of 19.3 LT. This result is particularly important because Mackey-Glass contains delayed nonlinear dynamics, which are often difficult to capture using conventional sequence models. The improvement suggests that the combination of Reservoir-Enhanced Embedding and Transformer attention helps CNT represent both short-term and delayed dependencies more effectively. In the Chen-Lu system, CNT achieved 7.6 LT, compared with 5.3 LT from CI-LSTM and 5.1 LT from NG-RC. This confirms the ability of CNT to handle different attractor structures within three-dimensional chaotic systems.

The Qi Hyperchaos system produced the lowest VPT values across all methods because it has two positive Lyapunov exponents, causing faster divergence in multiple directions of phase space. Nevertheless, CNT still achieved 6.2 LT, outperforming CI-LSTM with 4.1 LT and NG-RC with 3.9 LT. This result shows that the proposed model remains robust under hyperchaotic conditions. For the KS-PDE, CNT obtained 9.8 LT, while CI-LSTM achieved 7.4 LT and FEDformer achieved 7.1 LT. Although the improvement remains significant, the margin is smaller compared with lower-dimensional systems, indicating that spatiotemporal chaos still presents a major challenge for deep learning-based prediction models.

Overall, CNT consistently achieved higher VPT than the strongest baseline across all benchmark systems. This result suggests that the proposed architecture does not merely overfit to a specific chaotic benchmark, but benefits from the physics-informed components embedded in the model, particularly Lyapunov-regularized attention and Koopman-based linearization.

3.3 Chaos Synchronization Results

In addition to prediction performance, this study also evaluated the ability of CNT to support chaos synchronization. Synchronization is an important property in chaotic systems because it measures whether a response system can follow the dynamics of a driver system under certain coupling conditions. In this experiment, two identical Lorenz-63 systems were coupled with coupling strength $c \in [0.001, 10.0]$. The predicted trajectory generated by CNT was used as the driver trajectory to drive the response system. The synchronization error was then measured after the response system evolved under the influence of the predicted driver.

The results show that CNT achieved a synchronization error below 0.05 when the coupling strength was greater than 0.12. At the same coupling strength, ESN produced a synchronization error of 0.41, while FEDformer produced an error of 0.28. This result indicates that CNT preserved the synchronization manifold of the attractor more effectively than the baseline models. The lower synchronization error suggests that the predicted trajectory generated by CNT did not merely approximate the numerical values of the reference system, but also retained the underlying dynamical structure required to drive another chaotic system into synchronization.

This finding is important because chaos synchronization is more demanding than ordinary trajectory prediction. A model may produce acceptable short-term prediction error but still fail to preserve the phase-space structure needed for synchronization. The ability of CNT to achieve low synchronization error indicates that the proposed architecture captures deeper dynamical information beyond surface-level temporal patterns. The Koopman Linearization Module plays an important role in this result because it encourages the nonlinear dynamics to be represented in a latent space where the evolution can be approximated linearly. This latent linearization helps maintain a more stable relationship between the predicted driver trajectory and the response system.

The synchronization result also supports the relevance of Lyapunov-regularized attention. Since synchronization depends on the stability of trajectory evolution, a model that allows uncontrolled divergence in its internal representation may produce predictions that quickly leave the synchronization manifold. By penalizing unstable divergence in the attention structure, CNT improves the consistency of the predicted dynamics. Therefore, the synchronization experiment provides additional evidence that the proposed model is not only accurate in terms of VPT, but also dynamically meaningful.

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 124
This work is licensed under a Creative Commons Attribution 4.0 International License.

ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

3.4 Ablation Study

An ablation study was conducted to evaluate the contribution of each major architectural component in CNT. The tested configurations include the full CNT model, CNT without Reservoir-Enhanced Embedding, CNT without the Koopman Linearization Module, CNT without Lyapunov regularization, and CNT without both REE and KLM. The results are shown in Table 3.

Table 3. Ablation Study Results

Configuration	Lorenz VPT	Average VPT	Δ vs. Full
Full CNT (Ours)	11.2±0.2	11.5±0.2	—
w/o REE	9.3±0.3	9.4±0.3	−18.3%
w/o KLM	8.9±0.3	8.9±0.3	−22.1%
w/o ℒLE	9.6±0.2	9.8±0.2	−14.6%
w/o REE & KLM	7.1±0.4	7.3±0.4	−36.5%

Table 3 shows that removing any major component reduced the performance of CNT. Removing REE reduced the average VPT by 18.3%, indicating that the reservoir-based nonlinear preprocessing contributes significantly to the model's ability to extract transient chaotic features. This result confirms that self-attention alone may not be sufficient to capture all nonlinear structures present in chaotic sequences, especially when the context window is limited.

Removing KLM caused the largest single-component performance reduction, with an average VPT decrease of 22.1%. This result suggests that Koopman-based latent linearization is one of the most important components of CNT. By mapping nonlinear dynamics into a latent space with approximately linear evolution, KLM provides a more structured representation of the system. Without this component, the Transformer must learn nonlinear evolution directly from the observed sequence, which may be less stable and less efficient.

Removing the Lyapunov regularizer reduced the average VPT by 14.6%. Although this decrease is smaller than the removal of KLM, it remains substantial. This result indicates that Lyapunov-aware learning helps control the divergence behavior of the model during training. Since chaotic systems naturally exhibit exponential divergence, incorporating Lyapunov regularization helps the model distinguish between physically meaningful divergence and unstable model-induced divergence.

The largest performance drop occurred when both REE and KLM were removed, resulting in a 36.5% reduction in average VPT. This confirms that the components of CNT are complementary. REE improves nonlinear feature extraction at the input representation level, while KLM provides operator-based structure in the latent space. When both components are removed, the model loses both its nonlinear preprocessing capability and its structured dynamical representation, resulting in a substantial decline in prediction performance.

3.5 Discussion

The experimental results demonstrate that integrating physics-informed inductive biases into a Transformer-based architecture produces substantial and consistent improvements over strong baseline models. Compared with ESN and NG-RC, CNT benefits from global self-attention, allowing it to capture long-range temporal dependencies more effectively. Compared with FEDformer, CNT provides additional dynamical awareness through Lyapunov regularization and Koopman linearization. Compared with CI-LSTM, CNT offers a more flexible representation because it combines attention-based sequence modeling with explicit mechanisms for nonlinear feature expansion and latent operator approximation.

The results also show that prediction performance in chaotic systems cannot be explained solely by architectural complexity. FEDformer and CI-LSTM are already strong sequence models, but they do not explicitly preserve the physical and geometric structure of chaotic systems. CNT improves upon these methods because it incorporates mechanisms that are directly related to the nature of chaotic dynamics. The Lyapunov-regularized attention mechanism addresses the problem of exponential divergence, while the Koopman Linearization Module addresses the problem of representing nonlinear dynamics in a more tractable latent space. Meanwhile, Reservoir-Enhanced Embedding improves the model's sensitivity to nonlinear transient patterns before the data are processed by the Transformer encoder.

The comparison with prior approaches also highlights the novelty of the proposed framework. Reservoir computing has long been effective for chaotic prediction because it creates high-dimensional nonlinear representations from temporal inputs. However, classical reservoir methods generally rely on fixed random dynamics and linear readout, which may limit their flexibility. Transformer-based forecasting models, on the other hand, are powerful for general time-series prediction but are not specifically designed to preserve attractor geometry or Lyapunov stability. CNT combines the strengths of both directions by using a fixed reservoir for nonlinear expansion and a trainable Transformer backbone for global temporal representation. The addition of Koopman-based latent linearization further distinguishes CNT from ordinary attention-based models.

The synchronization experiment strengthens this interpretation. CNT achieved much lower synchronization error than ESN and FEDformer at the same coupling strength, suggesting that the predicted trajectories generated by CNT are dynamically more consistent with the original attractor. This is an important result because synchronization

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 125
This work is licensed under a Creative Commons Attribution 4.0 International License.

ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

depends not only on pointwise prediction accuracy, but also on whether the predicted trajectory preserves the structural relationship between the driver and response systems. The result indicates that CNT learns representations that remain closer to the true synchronization manifold.

Nevertheless, this study has several limitations. The Koopman Linearization Module adds approximately 23% computational overhead during training. This additional cost may become significant when the model is applied to very large-scale spatiotemporal systems. In addition, CNT's prediction quality decreases more rapidly on high-dimensional spatiotemporal systems, especially in the Kuramoto-Sivashinsky benchmark. This limitation suggests that future work should explore hierarchical Koopman representations or multi-scale latent operators to improve scalability. Furthermore, theoretical guarantees for VPT improvement remain probabilistic because the fundamental predictability horizon of chaotic systems is bounded by Lyapunov time. Therefore, CNT should be understood as a model that extends the practical prediction horizon, not as a method that eliminates the intrinsic unpredictability of chaos.

Overall, the findings support the hypothesis that combining modern sequence modeling with physics-informed dynamical constraints can improve both prediction and synchronization performance in chaotic systems. The proposed CNT framework contributes to the literature by providing a principled bridge between nonlinear dynamics theory, Koopman operator learning, reservoir-based nonlinear representation, and Transformer-based temporal modeling. This combination opens opportunities for developing more reliable artificial intelligence systems for complex physical phenomena, including climate dynamics, fluid turbulence, biological oscillations, and sensor-based monitoring systems.

4. CONCLUSION

This study introduced ChaosNet-Transformer (CNT), a physics-informed deep learning architecture for predicting and synchronizing high-dimensional chaotic dynamical systems. The proposed framework integrates three complementary components: Reservoir-Enhanced Embedding, Lyapunov-regularized attention, and the Koopman Linearization Module. Experiments on six canonical chaotic benchmarks show that CNT achieves the highest Valid Prediction Time across all evaluated systems and consistently outperforms the strongest baseline models. The synchronization experiment also indicates that CNT preserves the synchronization behavior of the Lorenz-63 attractor more effectively than ESN and FEDformer, as reflected by lower synchronization error under comparable coupling conditions. These findings suggest that chaotic prediction can be improved by combining Transformer-based sequence modeling with physics-aware constraints and operator-based latent representations. Nevertheless, this study has several limitations, including the additional computational overhead introduced by the Koopman Linearization Module and reduced prediction performance on high-dimensional spatiotemporal benchmarks. Future research should explore online adaptation for time-varying attractors, extensions to stochastic chaotic systems, hierarchical Koopman representations, and lightweight CNT implementations for real-time sensor monitoring on IoT devices.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for supporting this research through the Fundamental Research Grant No. 2025/C3-FRP/RT-01-15 and the Research and Community Service Program for the 2025 Fiscal Year under Decree Number 0070/C3/AL.04/2025. The authors also thank the computing facilities of Universitas Asahan and the collaborative partners from STMIK Tunas Bangsa Pematangsiantar for their infrastructure support during the implementation of this research.

REFERENCES

[1] E. N. Lorenz, "Deterministic Nonperiodic Flow," *J. Atmos. Sci.*, vol. 20, no. 2, pp. 130–141, 1963, doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
[2] D. Ruelle and F. Takens, "On the Nature of Turbulence," *Commun. Math. Phys.*, vol. 20, pp. 167–192, 1971.
[3] A. M. Lyapunov, *The General Problem of the Stability of Motion*. London: Taylor & Francis, 1992.
[4] H. Lin *et al.*, "A Review of Chaotic Systems Based on Memristive Hopfield Neural Networks," *Mathematics*, vol. 11, no. 6, p. 1369, 2023, doi: 10.3390/math11061369.
[5] A. M. García-García, J. J. M. Verbaarschot, and J. Zheng, "Lyapunov Exponent as a Signature of Dissipative Many-Body Quantum Chaos," *Physical Review D*, vol. 110, p. 086010, 2024, doi: 10.1103/PhysRevD.110.086010.
[6] Q. Lai, X.-W. Zhao, F. Liu, and L. Wang, "Advances in Chaotification and Chaos-Based Applications," *Front. Phys.*, vol. 10, p. 996825, 2022.
[7] Y. Sun, L. Zhang, and M. Yao, "Chaotic Time Series Prediction of Nonlinear Systems Based on Various Neural Network Models," *Chaos Solitons Fractals*, vol. 175, p. 113971, 2023, doi: 10.1016/j.chaos.2023.113971.
[8] H. Cheng, H. Li, Q. Dai, and J. Yang, "A Deep Reinforcement Learning Method to Control Chaos Synchronization Between Two Identical Chaotic Systems," *Chaos Solitons Fractals*, vol. 174, p. 113809, 2023, doi: 10.1016/j.chaos.2023.113809.

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 126
This work is licensed under a Creative Commons Attribution 4.0 International License.

ADA Journal of Information System Research
ISSN 3025-9568 (media online)
Volume 3, No 3, June 2026 Page: 120−127
DOI doi.org/10.64366/adajisr.v3i3.584
https://journals.adaresearch.or.id/adajisr

[9] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data," *Phys. Rev. Lett.*, vol. 120, no. 2, p. 024102, 2018.
[10] P. R. Vlachas *et al.*, "Backpropagation Algorithms and Reservoir Computing in Recurrent Neural Networks for Forecasting of Complex Spatiotemporal Dynamics," *Neural Networks*, vol. 126, pp. 191–217, 2020.
[11] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, "Reservoir Observers: Model-Free Inference of Unmeasured Variables in Chaotic Systems," *Chaos*, vol. 27, no. 4, p. 041102, 2017.
[12] D. Hartama, Wanayumini, and I. S. Damanik, "Optimizing Disaster Response: A Systematic Review of Time-Dependent Cumulative Vehicle Routing in Humanitarian Logistics," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 3, 2024, doi: 10.26555/jiteki.v10i3.29686.
[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[14] H. Jaeger and H. Haas, "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication," *Science (1979).*, vol. 304, no. 5667, pp. 78–80, 2004.
[15] W. Maass, T. Natschlager, and H. Markram, "Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations," *Neural Comput.*, vol. 14, no. 11, pp. 2531–2560, 2002.
[16] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, "Using Machine Learning to Replicate Chaotic Attractors and Calculate Lyapunov Exponents from Data," *Chaos*, vol. 27, no. 12, p. 121102, 2017.
[17] P. R. Vlachas, "Data-Driven Forecasting of High-Dimensional Chaotic Systems with Long Short-Term Memory Networks," *Proceedings of the Royal Society A*, vol. 474, no. 2213, p. 20170844, 2018.
[18] D. Gauthier, E. Bollt, A. Griffith, and W. Barbosa, "Next Generation Reservoir Computing," *Nat. Commun.*, vol. 12, no. 1, p. 5564, 2021.
[19] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear PDEs," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.
[20] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural Ordinary Differential Equations," in *Proceedings of NeurIPS*, 2018, pp. 6571–6583.
[21] S. Greydanus, M. Dzamba, and J. Yosinski, "Hamiltonian Neural Networks," in *Proceedings of NeurIPS*, 2019, pp. 15379–15389.
[22] A. Vaswani *et al.*, "Attention Is All You Need," in *Proceedings of NeurIPS*, 2017, pp. 5998–6008.
[23] Y. Li, J. Cao, Y. Xu, L. Zhu, and Z. Y. Dong, "Deep Learning Based on Transformer Architecture for Power System Short-Term Voltage Stability Assessment with Class Imbalance," *Renewable and Sustainable Energy Reviews*, vol. 189, p. 113913, 2024, doi: 10.1016/j.rser.2023.113913.
[24] K. Cao, T. Zhang, and J. Huang, "Advanced Hybrid LSTM-Transformer Architecture for Real-Time Multi-Task Prediction in Engineering Systems," *Sci. Rep.*, vol. 14, p. 4890, 2024, doi: 10.1038/s41598-024-55483-x.
[25] H. Zhou *et al.*, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," in *Proceedings of AAAI*, 2021, pp. 11106–11115.
[26] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting," in *Proceedings of ICML*, 2022, pp. 27268–27286.
[27] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers," in *Proceedings of ICLR*, 2023.
[28] H. Lin, C. Wang, and Y. Sun, "A Universal Variable Extension Method for Designing Multiscroll/Wing Chaotic Systems," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 7, pp. 7806–7818, 2024, doi: 10.1109/TIE.2023.3299020.
[29] X. Liu, L. Zhao, and J. Jin, "A Noise-Tolerant Fuzzy-Type Zeroing Neural Network for Robust Synchronization of Chaotic Systems," *Concurr. Comput.*, vol. 36, no. 22, pp. e8218–e8218, 2024, doi: 10.1002/cpe.8218.
[30] I. Mezic, "Spectral Properties of Dynamical Systems, Model Reduction and Decompositions," *Nonlinear Dyn.*, vol. 41, pp. 309–325, 2005.
[31] H. Shi and M. Q.-H. Meng, "Deep Koopman Operator With Control for Nonlinear Systems," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 7700–7707, 2022, doi: 10.1109/LRA.2022.3184036.
[32] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep Learning for Universal Linear Embeddings of Nonlinear Dynamics," *Nat. Commun.*, vol. 9, p. 4950, 2018.
[33] Q. Li, F. Dietrich, E. M. Bollt, and I. G. Kevrekidis, "Extended Dynamic Mode Decomposition with Dictionary Learning," *Chaos*, vol. 27, no. 10, p. 103111, 2017.

Copyright © 2026 The Author(s). Published by *ADA Research Center*. Page 127
This work is licensed under a Creative Commons Attribution 4.0 International License.