Time-Varying Biological Time Series Prediction and Pattern Interpretation via Koopman Theory and Large Language Model

Yujie You , Yuzhu Ji , Feixiang Zhao , Ming Xiao , Le Zhang *

Posted Date: 24 March 2026

doi: 10.20944/preprints202603.1888.v1

Keywords: biological time-series prediction; Koopman embedding; large language model; frequency-domain information

 doi:10.20944/preprints202603.1888.v1

Article

Time-Varying Biological Time Series Prediction and Pattern Interpretation via Koopman Theory and Large Language Model

Yujie You [1], Yuzhu Ji [1], Feixiang Zhao [1], Ming Xiao [2] and Le Zhang [2,*]

[1] School of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin, 644000, China

[2] College of computer science, Sichuan University, Chengdu, 610065, China

* Correspondence: zhangle06@scu.edu.cn;

Abstract

Biological time series data characterizes the dynamic evolution of biological systems and plays a crucial role in genetic inheritance, disease diagnosis, and biological microenvironment. However, accurate prediction for biological time-series data remains challenging due to their pronounced time-varying, non-stationary, and noisy characteristics. Existing approaches often fail to capture latent distribution shifts and the coupled evolution of global and local patterns, limiting both predictive performance and interpretability. Thus, this study firstly proposes a time-varying neural network (TVNN) model that combine frequency-domain information with Koopman embedding theory. TVNN model Koopman transition matrices are used to model global dynamics and local time-varying behaviors for pattern extraction. Secondly, a time-varying pattern recognition large language model (TVPRLLM) is introduced to recognize and interpret the extracted time-varying patterns, enabling the discovery of their potential biological significance. Thirdly, we have developed such a biological time series predictive platform that can offer visualization, data analysis, and predictive services. Experimental results demonstrate that TVNN model outperforms existing mainstream methods in predicting on biological time-varying time series.

Keywords: biological time-series prediction; Koopman embedding; large language model; frequency-domain information

1. Introduction

Driven by the rapid development of high-throughput sequencing technologies, biological time-series data have grown rapidly, encompassing multi-omics molecular data, microbial environmental observations, and medical monitoring data. These heterogeneous data provide a dynamic and multidimensional characterization of biological systems across both temporal and spatial dimensions, thereby creating unprecedented opportunities for biomedical discovery, environmental microbiology research, and clinical auxiliary diagnosis [1,2]. In this context, biological time-series prediction uses past temporal data to predict future states in biological processes. This is of great significance not only for understanding the dynamic regulatory mechanisms of living systems and their interactions with the environment, but also for advancing precision medicine through increased diagnosis, prognosis assessment, and treatment planning [3].

However, biological systems are not static entities but inherently complex time-varying systems, whose molecular activities, physiological processes, biochemical reactions, and environmental interactions evolve continuously over time [4]. Such temporal variability is reflected across multiple biological layers, including gene regulatory activity, protein modification and environment interactions. As the intrinsic mechanisms of living systems and their extrinsic environments are both dynamically changing, biological time-series data often exhibit high dimensionality, nonlinear trends

 doi:10.20944/preprints202603.1888.v1

and seasonality, resulting in pronounced time-varying and non-stationary behavior [4]. Figure 1 illustrates a Lorenz system with different time-varying patterns, which as an idealized representation of chaotic gene regulatory dynamics [5]. From a biological perspective, these temporal shifts may be driven by disease progression, treatment responses, or ecological perturbations, which in turn induce changes in covariate patterns across time steps. Such changes in data distribution and statistical properties present substantial challenges for accurate biological time-series prediction.

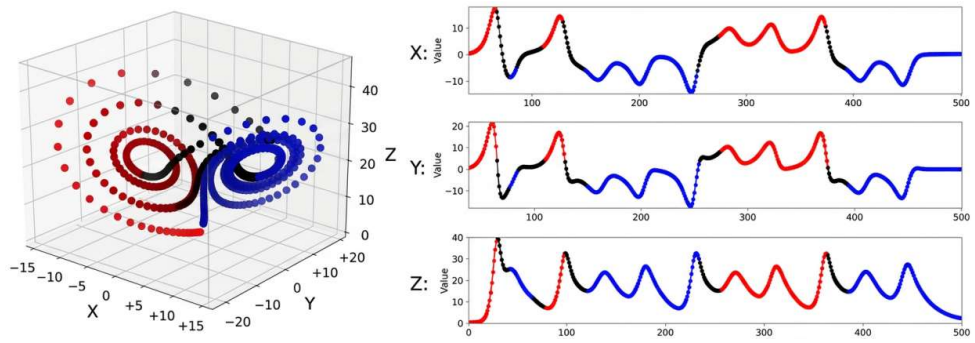

Figure 1. Lorenz system with different time-varying patterns and the temporal evolutions of its three state variables (X, Y, and Z).

The Lorenz system is used to model chaotic gene regulatory networks in biological time-varying systems. The red denotes a highly active dynamic regime, which may correspond to strong regulatory responses or pronounced gene-expression fluctuations; the black indicates a transitional regime reflecting switching between different regulatory states; and the blue represents a relatively stable or alternative oscillatory regime, which may correspond to weaker fluctuations or a balanced regulatory state.

Early studies on biological time-series prediction mainly focused on time-domain analysis [6–10], typically using classical statistical models such as AR Autoregression (AR) [9] and Autoregressive Integrated Moving Average (ARIMA) [10], to capture basic temporal dynamics. However, these methods rely on the assumption of stationarity and are therefore inadequate for biological systems, which are inherently time-varying and non-stationary. With the development of deep learning, the RevIN model [11] was proposed to remove and restore instance-specific mean and variance, thereby transforming the predictive task into modeling a relatively stationary sequence. Although such methods can increase predictive accuracy to some extent, they mainly focus on statistical components and remain limited in characterizing seasonal patterns, frequency structures, and deeper dynamic mechanisms, making them less suitable for complex time-varying biological systems.

As one of the principal methodologies for complex dynamics analysis, Koopman theory [12] offers a new perspective to predict time-varying biological time series. It embeds nonlinear dynamical systems into an infinite-dimensional linear space, where the evolution follows linear dynamics while preserving the nonlinear characteristics of the original system. To address temporal distribution shift, Liu et al. [13] proposed the Koopa forecaster, a Koopman-learning residual framework that hierarchically models the underlying dynamical system of complex time series. This design enables the model to progressively characterize different levels of temporal dynamics in non-stationary sequences, achieving strong performance in influenza and climate prediction.

Meanwhile, frequency-domain information is particularly well suited to characterize time-varying biological systems, because it can effectively separate periodic components from evolving components [14,15]. Such a decomposition helps us reveal the time-varying characteristics hidden in biological data. Therefore, how to integrate frequency-domain information into Koopman theory to construct a unified neural modeling framework for time-varying biological systems, so as to model

 doi:10.20944/preprints202603.1888.v1

the dynamical evolution of complex time-varying characteristics, extract biological time-varying patterns, and accurately predict future states, has become our first scientific problem.

Although biological time-series prediction models can predict future states, prediction alone is insufficient to fully reveal the time-varying behaviors of biological systems or to understand their underlying mechanisms. Wang et al. [13] used Koopman theory to uncover hidden time-varying patterns from biological time-series data, but their method mainly captures pattern features and provides limited biological interpretability. Recently, Large Language Models (LLMs), such as GPT-4 [16] and Qwen [17], have created new opportunities to enhance the biological interpretability of time-varying patterns, due to their strong pattern recognition and generalization abilities. Therefore, how to combine Koopman embedding theory with LLMs to build a time-varying pattern recognition model, identify patterns encoded in the Koopman transition matrix, and increase their biological interpretability has become our second scientific question.

To further support biological research practice, relevant methods are often transformed into platform-based tools to decrease usage barriers and increase their practical value [18,19]. For example, MathIOmica [20] enables in-depth analysis of multi-omics time-series data through multi-scale temporal trajectory modeling and supports visualization of spatiotemporal correlations, while MOVIS [21] provides embedding and interactive visualization for multimodal biological time-series data. However, these platforms mainly focus on multi-omics data analysis and provide limited support for predictive tasks across broader types of biological time-series data. In addition, most existing predictiv4 models are released as software packages [22], which often lack user-friendly interfaces and operational convenience. Therefore, how to build a biological time-series predictive platform with a friendly user experience has become our third scientific question.

To address the above three scientific problems, we propose a time-varying neural network (TVNN). By combining frequency-domain information with Koopman embedding theory, TVNN effectively extracts the non-stationary representations of biological time series data and conducts non-linear dynamic inference on time-varying biological systems, thereby realizing highly efficient prediction of their time series data. Secondly, by incorporating a large language model fine-tuned collaboratively with LoRA[23–25], a time-varying pattern recognition large language model (TVPRLLM) is constructed to identify the time-varying patterns of these biological systems and unearth their potential biological semantics. Finally, we developed a biological time-series prediction platform to provide users with efficient prediction services. Here, we carried out a comprehensive evaluation for our TVNN model on six biological time series datasets and demonstrated that our proposed method outperforms the existing mainstream methods in time series prediction for time-varying biological systems.

2. Materials and Methods

2.1. Experimental Datasets

We evaluated the proposed model on six biological time series datasets: Proteomics, Gene, Solar, EMG, Climate, and ILI. These datasets span multiple biological scales, including molecular expression, gene regulation, physiological state monitoring, environmental drivers of biological activity, and population-level disease dynamics, thus providing a broad benchmark for assessing the model's ability to capture biologically meaningful temporal patterns.

Proteomics: Proteomic time-series data exhibit complex time-varying behavior and often present oscillatory expression patterns associated with intracellular regulation and protein-protein interactions. To mimic the unstable oscillatory behavior observed in protein expression [26,27], we adopt the nonlinear pendulum system [28]:

$$\begin{cases} \frac{d^2\theta}{dt^2} + \frac{g}{l} sin\theta = 0 \\ \theta(t_0) = \theta_0 \end{cases} \quad (1)$$

 doi:10.20944/preprints202603.1888.v1

Here l is the pendulum length, g is the gravitational acceleration and θ_0 is the initial condition. A greater θ_0 implies stronger nonlinear interactions among proteins. To account for biological variability and measurement noise commonly encountered in real proteomic data [5], Gaussian white noise [29] is added:

$$\tilde{x}_i^t = x_i^t + \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad (2)$$

Here, $\tilde{x}_i^t$ represents the noisy ith dimensional data at time step t. The noise term ε_i follows a normal distribution $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, where σindicates the noise intensity. In this experiment, we set $\theta_0 = 2.4$ and $\sigma = 0.1$ to generate a sequence of 2000 steps , and project it into a 32-dimensional space through random orthogonal transformation [30].

Gene: Gene regulatory networks are highly nonlinear and chaotic systems that reflect the dynamic interplay of transcriptional regulation and feedback control [5]. To model these characteristics, we generate data using the Lorenz system [31]:

$$\begin{cases} x^{t+1} = x^t + h\big(\eta(y^t - z^t)\big) \\ y^{t+1} = z^t + h(x^t(\rho - z^t) - y^t) \\ z^{t+1} = z^t + h(x^t y^t - \beta z^t) \end{cases} \quad (3)$$

Here η, ρ and β are constants, and h controls the interaction strength. Greater h values correspond to stronger nonlinear coupling among genes.

Solar: Solar radiation is a key environmental factor affecting photosynthesis, growth, and metabolic rhythms in biological systems. The dataset [32] was collected from 137 photovoltaic power plants in Alabama at 10-minute intervals. Predicting solar radiation patterns is beneficial for understanding ecosystem functioning and the environmental determinants of biological activity.

EMG: Muscle fatigue is a time-varying physiological process involving changes in neuromuscular activation and energy metabolism. The EMG dataset [33] records signals from subjects sustaining 20% of their maximum grip strength until fatigue. Predicting the temporal and spectral features of these signals helps us reveal fatigue-related physiological mechanisms and supports rehabilitation-oriented biomedical analysis.

Climate: Climate dynamics strongly affect ecological stability, species adaptation, and biological responses to environmental stress. This dataset [34] contains 21 meteorological variables recorded every 10 minutes throughout 2020. Accurate climate prediction can help us quantify the effects of environmental change on biological systems and ecosystem resilience.

ILI: Influenza-like illness exhibits pronounced temporal variability driven by viral mutation, host immunity, and seasonal transmission patterns. This dataset [34] records weekly ILI cases reported by the U.S. CDC from 2002 to 2021. Modeling its temporal evolution is valuable for outbreak prediction, epidemiological surveillance, and public health intervention planning.

2.2. Problem Setting

Let $St(t) \in \mathcal{H}$ denote the underlying state of a biological system, where $\mathcal{H}$ is a finite-dimensional vector space, and let $Z(t) \in \mathbb{R}^D$denote its observed time-series measurement. For the observed biological time-series data $Z = (Z_1, \ldots, Z_t, \ldots Z_M) \in \mathbb{R}^{D \times M}$, D and M represent the variable dimension and the number of observed time series steps, respectively. The biological system is described by:

$$\frac{\partial \mathrm{St(t)}}{\partial \mathrm{t}} = \mathrm{A(St(t))} \quad (4)$$

$$Z_{t+1} = B(Z_t) \quad (5)$$

Here A is the governing equation of the system and B characterizes the evolution of the observed data. Our first objective is to learn a mapping function F that predicts the future L steps from M historical observation. Formally, given observed biological time series data Z, the goal is to estimate $\hat{Z}^p \in \mathbb{R}^{D \times L}$:

$$\hat{Z}^p = (\hat{Z}^p_{M+1}, \hat{Z}^p_{M+2}, \ldots, \hat{Z}^p_{M+L}) = F(Z_1, Z_2, \ldots, Z_t, \ldots, Z_M) \quad (6)$$

Our second objective is to learn another mapping function G for latent biological semantic discovery. Specifically, the M historical observations are divided into S segments of equal length

 doi:10.20944/preprints202603.1888.v1

M/S. For a segment beginning at time step t, Gtakes $(Z_t, Z_{t+1}, \ldots, Z_{t+M/S-1})$ as input and output the latent semantic state $state_t$ of the biological system:

$$state_t = G(Z_t, Z_{t+1}, \ldots, Z_{t+M/S-1}) \quad (7)$$

2.3. Time-Varying Neural Network

Illustrated by Figure 2, to accomplish the task objectives described in Section 2.1, we develop a time-varying neural network (TVNN) for biological time-series prediction to realize the mapping function F. This section is composed of four parts: data segmentation and decomposition, global prediction module, time-varying predictive module, and combination module.

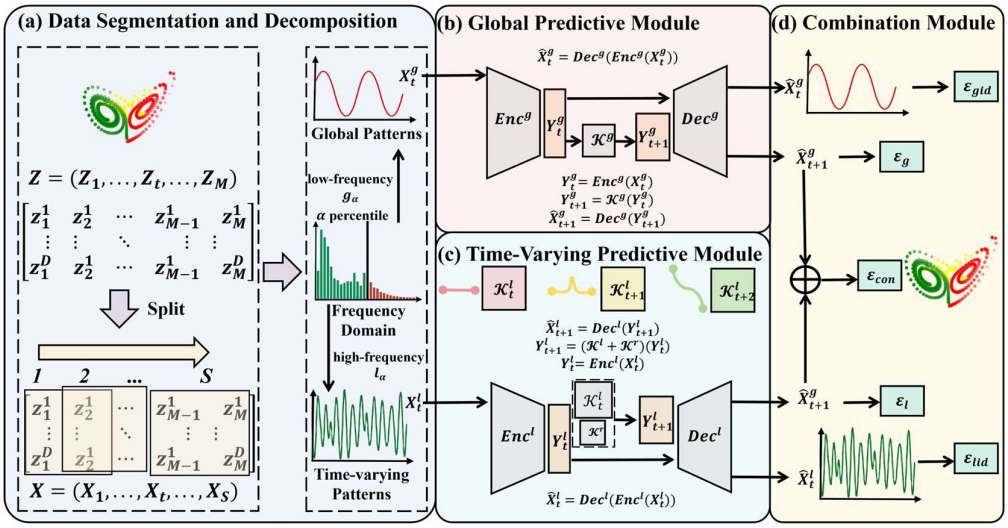

Figure 2. The time-varying neural network is composed of four parts: (a) data segmentation and decomposition; (b) global prediction module; (c) time-varying predictive module and (d) combination module.

2.3.1. Data Segmentation and Decomposition

Lin et al. [35] demonstrated that segment-wise iteration achieves better predictive performance than point-wise iteration for time-series prediction. Accordingly, given the observed biological time-series data $Z = (Z_1, \ldots, Z_t, \ldots Z_M) \in \mathbb{R}^{D \times M}$, where $Z_t = (z_1, \ldots., z_D) \in \mathbb{R}^D$, we divide the data into S segments of equal length M/S. The segmented sequence is denoted as $X = (X_1, \ldots, X_t, \ldots X_s)$, where $X_t = (Z_{(t-1)M/S+1}, Z_{(t-1)M/S+2}, \ldots., Z_{tM/S}) \in \mathbb{R}^{D \times (M/S)}$.

For biological time-series data, low-frequency global patterns characterize the organism's fundamental physiological states or long-term trends, whereas high-frequency time-varying patterns reflect transient physiological responses or stimulus-induced fluctuations. Motivated by Liu et al. [13], who verified the effectiveness of Fourier filter decomposition for time-series prediction, we employ Fourier filtering [14,15] to decompose each segment X_t in the frequency domain. As a result, each segment is separated into low-frequency global patterns X_t^g and high-frequency time-varying patterns X_t^l , formulated as:

$$X_t^g = \mathcal{F}^{-1}(g_\alpha(\mathcal{F}(X_t))) \quad (8)$$

$$X_t^l = \mathcal{F}^{-1}(l_\alpha(\mathcal{F}(X_t))) \quad (9)$$

$$X_t = X_t^g + X_t^l \quad (10)$$

Specifically, a Fast Fourier Transform (FFT) is applied to each segment X_t. Components with amplitudes above the α percentile are preserved as low-frequency global components, while those below the α percentile are regarded as high-frequency time-varying components. Here, g and l denote the global and time-varying filters, respectively; α is a hyperparameter; $\mathcal{F}$ denotes the Fourier Transform, and $\mathcal{F}^{-1}$ denotes the Inverse Fourier Transform.

 doi:10.20944/preprints202603.1888.v1

2.3.2. Global Predictive Module

In biological systems, low-frequency global patterns characterize changes and responses that occur over relatively long-time scales. These patterns are associated with macroscopic processes. Reconstructing such low-frequency global patterns can help us reveal the intrinsic laws governing the long-term evolution of biological systems.

This study employs an encoder-decoder architecture to project the observed data onto a high-dimensional measurement space. A linear Koopman operator is then applied to model the evolution of the global patterns and infer the trajectory of the low-frequency global dynamics:

$$\hat{X}_{t+1}^{g} = Dec^{g}(K^{g}(Enc^{g}(X_t^{g}))) \quad (11)$$

Here, $Enc^{g}: \mathbb{R}^{D\times(M/S)} \to \mathbb{R}^{D\times d}$ denotes the projection encoder, $K^{g}: \mathbb{R}^{D\times d} \to \mathbb{R}^{D\times d}$ denotes the Koopman transition operator, and $Dec^{g}: \mathbb{R}^{D\times d} \to \mathbb{R}^{D\times(M/S)}$ denotes the projection decoder for the global patterns. Both $\mathrm{Enc^g}$and $\mathrm{Dec^g}$ are implemented as feed forward neural networks.

2.3.3. Time-Varying Predictive Module

High-frequency time-varying patterns characterize the changes and responses of biological systems over short time scales. These patterns involve microscopic and mesoscopic processes, such as individual physiological activities and interactions among organisms. Reconstructing high-frequency time-varying patterns helps us reveal the local temporal variation trends of biological systems over extended periods.

This study employs an encoder,$Enc^{l}: \mathbb{R}^{D\times(M/S)} \to \mathbb{R}^{D\times d}$, to project the observed data onto a high-dimensional measurement space, yielding Y_t^{l}. Here, the Enc^{l} is implemented as a feed forward neural network.

$$Y_t^{l} = Enc^{l}(X_t^{l}) \quad (12)$$

Subsequently, for each high-dimensional representation Y_t^{l}, the time-varying predictive module employs the Dynamic Mode Decomposition (DMD) algorithm [36] to compute the Koopman operator K_t^{l} (Eq. 13). The operator K_t^{l} is then used to drive the evolution of high-frequency time-varying trajectories (Eq. 14).

$$K_t^{l} = Y_{t+1}^{l} {Y_t^{l}}^{+} \quad (13)$$

$$\hat{Y}_{t+1}^{l} = K_t^{l}(Y_t^{l}) \quad (14)$$

Here, ${Y_t^{l}}^{+}$denotes the generalized inverse of Y_t^{l}, satisfying $Y_t^{l}{Y_t^{l}}^{+} = id$, where id represents the identity matrix.

In addition, since residual errors are inevitable in reconstructing time-varying patterns, a learnable Koopman residual matrix $\mathrm{K^r}$ is introduced to dynamically adjust and compensate for reconstruction errors (Eq. 15). This mechanism accounts for both system modeling errors and real-world disturbances:

$$\hat{Y}_{t+1}^{r} = K^{r}(Y_t^{l}) \quad (15)$$

Finally, the decoder $Dec^{l}: \mathbb{R}^{D\times d} \to \mathbb{R}^{D\times(M/S)}$ is used to map the high-dimensional representations back to the original measurement space, thereby reconstructing $\hat{\mathrm{X}}_{\mathrm{t}}^{\mathrm{l}}$ (Eq. 16). Meanwhile, to ensure that the decoder can also reconstruct the states evolved by the Koopman operator and thus improve the generalization ability of the predictive model, Dec^{l} further projects the evolved representation $\hat{Y}_{t+1}^{l}$back to the measurement space to reconstruct $\hat{X}_{t+1}^{l}$ (Eq. 17). Here, Dec^{l} is also implemented as a feed-forward neural network.

$$\hat{X}_t^{l} = Dec^{l}(Y_t^{l}) \quad (16)$$

$$\hat{X}_{t+1}^{l} = Dec^{l}(\hat{Y}_{t+1}^{l} + \hat{Y}_{t+1}^{r}) = Dec^{l}((K_t^{l} + K^{r})Enc^{l}(X_t^{l})) \quad (17)$$

2.3.4. Combination Module

The final prediction is obtained by combining the outputs of the global predictive module and the time-varying predictive module. Specifically, the evolution of the observed data is expressed as:

$$\hat{X}_{t+1} = \hat{X}_{t+1}^{g} + \hat{X}_{t+1}^{l} \quad (18)$$

 doi:10.20944/preprints202603.1888.v1

Following the segmentation scheme described in Section 2.2.1, the predictive estimate $\hat{Z}^p$ is rearranged into segmented form $\hat{X}^p = (\hat{X}^p_{S+1}, \ldots, \hat{X}^p_{S+t}, \ldots, \hat{X}^p_{S+LS/M})$, where $\hat{X}^p_{S+t} = (\hat{Z}^p_{M+Mt/S+1}, \ldots, \hat{Z}^p_{M+M(t+1)/S})$. Under this partitioning scheme, each segment predicted segment $\hat{X}^p_{S+t}$ is computed as:

$$\hat{X}^g_{S+t} = Dec^g(K^{g^{(t)}}(Enc^g(X^g_S))) \quad (19)$$

$$\hat{X}^l_{S+t} = Dec^l((K^l_{S-1} + K^r)^{(t)}(Enc^l(X^l_S))) \quad (20)$$

$$\hat{X}^p_{S+t} = \hat{X}^g_{S+t} + \hat{X}^l_{S+t} \quad (21)$$

Because TVNN contains trainable encoders and decoders in both prediction modules, together with the global Koopman transition operator K^g, the entire framework can be trained end-to-end. The training objective consists of five loss terms: global reconstruction loss ε_{gid}, global prediction loss ε_g, time-varying reconstruction loss ε_{lid}, time-varying prediction loss ε_l, and combination loss ε_{con}. The overall loss ε is defined as:

$$\varepsilon = \lambda_{gid}\varepsilon_{gid} + \lambda_g\varepsilon_g + \lambda_{lid}\varepsilon_{lid} + \lambda_l\varepsilon_l + \lambda_{con}\varepsilon_{con} \quad (22)$$

Here, λ_{gid}, λ_g, λ_{lid}, λ_l, and λ_{con} denote the weighting coefficients of the corresponding loss terms.

The global reconstruction loss ε_{gid} is defined as the Mean Squared Error (MSE)[37] between the true low-frequency global component and its reconstruction:

$$\varepsilon_{gid} = \left\|\hat{X}^g_t - X^g_t\right\|_{MSE} \quad (23)$$

The global prediction loss ε_g measures the MSE of the one-step-ahead global prediction:

$$\varepsilon_g = \frac{1}{M}\sum_{t=1}^{M}\left\|\hat{X}^g_{t+1} - X^g_{t+1}\right\|_{MSE} \quad (24)$$

The time-varying reconstruction loss ε_{lid} is defined as the MSE between the true high-frequency time-varying component and its reconstruction:

$$\varepsilon_{lid} = \left\|\hat{X}^l_t - X^l_t\right\|_{MSE} \quad (25)$$

The time-varying prediction loss ε_l measures the MSE of the one-step-ahead time-varying prediction:

$$\varepsilon_l = \frac{1}{M}\sum_{t=1}^{M}\left\|\hat{X}^l_{t+1} - X^l_{t+1}\right\|_{MSE} \quad (26)$$

Finally, the combination loss evaluates the discrepancy between the combined prediction and the true observation:

$$\varepsilon_{con} = \frac{1}{M}\sum_{t=1}^{M}\left\|\hat{X}^g_{t+1} + \hat{X}^l_{t+1} - X_{t+1}\right\|_{MSE} \quad (27)$$

2.4. Time-Varying Pattern Recognition Large Language Model

Liu et al. [13] demonstrated that the eigenvalues of the Koopman operator can characterize the amplitude of dynamical evolution. Motivated by this finding, we propose a time-varying pattern recognition large language model, termed TVPRLLM, for pattern prediction in time-varying biological systems. Specifically, the model takes the Koopman matrix, which encodes the evolutionary characteristics of biological time-series data, as input and learns its features to infer the corresponding dynamic patterns. As illustrated in Figure 2, TVPRLLM consists of three stages: prompt engineering, modeling & fine-tuning, and prediction.

Prompt engineering: Prompts serve as task instructions for LLMs. Among various prompt engineering strategies, role-playing has become a widely adopted technique. By embedding a specific expert persona into the prompt, role-playing guides the LLM to perform the target task from a predefined professional perspective.

 doi:10.20944/preprints202603.1888.v1

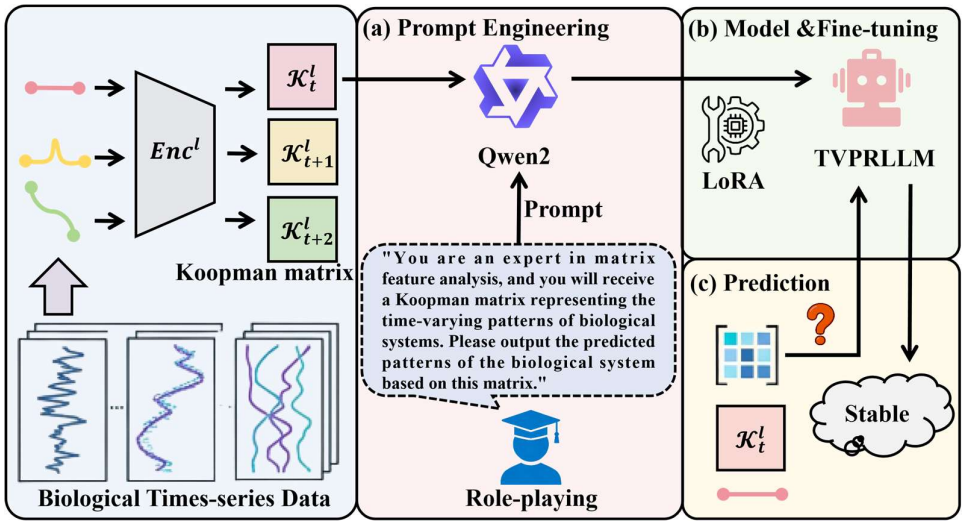

Figure 3. The TVPRLLM workflow, which consists of three steps: prompt engineering, modeling & fine-tuning, and prediction.

Modeling & fine-tuning: TVPRLLM is built upon the Qwen2 LLM [17] and LoRA [38]. More specifically, TVPRLLM employs the pre-trained Transformer-based LLM Qwen2 as its base classifier. The model is then fine-tuned on the time-varying patterns extracted from the observed data, while LoRA is used to specialize its knowledge and increase the prediction of biological meanings from these time-varying patterns. LoRA is a parameter-efficient fine-tuning approach that constrains updates of the weight matrices to a low-rank space [38]. The effectiveness of TVPRLLM lies in its ability to leverage the extensive prior knowledge encoded in the pre-trained Qwen2 model.

For the input data K_t^l, if the Koopman matrix K_t^l is relatively small in scale, the entire matrix is directly used as input; otherwise, its eigenvalues are used instead. Since large language models can be fine-tuned for non-linguistic tasks without modifying their architecture or loss functions [39], the default cross-entropy loss [40] is adopted to fine-tune TVPRLLM. For each fine-tuning sample, the training template is defined as follows:

Prompt: "You are an expert in matrix feature analysis, and you will receive a Koopman matrix representing the time-varying patterns of biological systems. Please output the predicted patterns of the biological system based on this matrix."

Input: "K_t^l"

Output: "Stable"

Prediction: After fine-tuning TVPRLLM, the model outputs binary classification results, such as "Increase" or "Reduce", and "Chaos" or "Stable", to represent the potential meaning state of the biological system.

3. Results

3.1. Experimental Details

Baselines: Six representative biological time series predicting methods were selected for performance comparison with our proposed TVNN: Vector Autoregressive Integrated Moving Average (VARIMA) [41], Support Vector Regression (SVR) with radial basis function kernel [42], Recurrent Neural Network (RNN) [43], Koopman Autoencoder (KAE) [44], Embedding, Koopman, and Autoencoder-based multi-omics Time series Prediction model (EKATP) [5], and Koopman Predictor (Koopa)[13].

Evaluation metrics: To quantitatively assess the prediction performance of different models, the Root Mean Square Error (RMSE) [45] was used as the evaluation metric:

 doi:10.20944/preprints202603.1888.v1

$$RMSE = \sqrt[2]{\frac{1}{L}\sum_{i=M+1}^{m+L}\left\|\hat{Z}_i - Z_i\right\|^2} \quad (28)$$

where Z_i denotes the true value, $\hat{Z}_i$ denotes the predicted value, and L is the prediction horizon.

Input data: For all datasets, the input variables were normalized using Min-Max normalization [46] prior to model training.

3.2. Experimental Results and Analysis on TVNN

To address our first scientific question, we compared the proposed TVNN with six representative methods: the widely used VARIMA, SVR, RNN, and KAE, as well as two recently developed models, EKATP and Koopa[13]. Subsequently, we investigated how the Fourier filter and the time-varying degree of biological systems affect the prediction performance of TVNN model. Finally, the effectiveness of the Fourier frequency domain decomposition was demonstrated by ablation experiments.

3.2.1. Prediction of Biological Time-Series Data

Table 1 summarizes the quantitative comparison results of multivariate prediction. The first column lists six different datasets, including Proteomics, Gene, Solar, EMG, Climate, and ILI. Then, Table 1 lists the performance of the evaluation metric RMSE for each model at different predictive time steps. The best average results are highlighted in bold. The winning counts show how many times each model achieves the best metric.

Table 1. The prediction results of prediction models on six time series datasets.

Datasets	Steps	TVNN	VARIMA	SVR	RNN	KAE	EKATP	Koopa
Proteomics	16	0.0740	0.0862	0.0776	0.0867	0.0765	0.0631	0.0672
	32	0.0752	0.0793	0.4385	0.0901	0.0797	0.0765	0.0772
	48	0.0749	0.0758	0.3580	0.0928	0.0835	0.0817	0.0719
Gene	8	0.1898	0.2478	3.070	0.3396	0.5979	0.2412	0.1901
	16	0.1596	0.3874	2.171	0.3805	0.7811	0.2627	0.1843
	24	0.1980	0.3637	1.773	0.4558	0.9572	0.3594	0.2411
Solar	32	0.6468	0.8987	1.6531	0.5371	1.6192	0.8328	0.4415
	64	0.4831	0.9489	1.4152	0.4776	1.8598	0.7687	0.5753
	96	0.4465	1.1246	1.4549	0.4448	1.8247	0.6623	0.5422
EMG	16	0.2950	0.4218	0.3452	0.4413	0.4329	0.4378	0.3024
	32	0.3383	0.4334	0.2441	0.4527	0.4432	0.4489	0.3241
	48	0.3036	0.4523	0.1993	0.4128	0.4024	0.4087	0.2819
Climate	16	0.0107	0.0112	1.0037	0.0134	0.0113	0.0096	0.0981
	32	0.0125	0.0133	0.7420	0.0157	0.0185	0.0159	0.1132
	48	0.0131	0.0221	0.6058	0.0190	0.0162	0.0195	0.1258
ILI	8	0.1249	0.6304	0.7457	0.1471	0.1375	0.1382	1.1291

 doi:10.20944/preprints202603.1888.v1

	12	0.1232	0.6253	0.5273	0.1712	0.1519	0.1225	1.3238
	16	0.1398	0.6412	0.3729	0.2387	0.1385	0.1493	1.4397
Winning counts		9	0	2	1	1	3	2

The following observations can be drawn from Table 1:

(1) The winning counts reported in the last row of Table 1 show that TVNN achieves the best overall performance in the multivariate prediction task among all compared methods.

(2) In biological time series prediction, TVNN consistently outperforms the Koopman-theory-based prediction models KAE, EKATP, and Koopa[13], under most experimental settings.

(3) In contrast, the overall prediction performance of VARIMA and SVR is inferior to that of the Koopman-theory-based models. Their results are generally higher and less competitive across the evaluated datasets and prediction horizons.

3.2.2. Fourier Frequency Domain Decomposition

Table 2 summarizes the effects of the Fourier filter parameter α and the time-varying degree of the biological system for the predictive performance of TVNN. The first column lists the values of the time-varying parameter θ, the second column reports the Fourier filter parameter α, and the remaining columns present the corresponding RMSE statistics of prediction results, including the minimum (Min), maximum (Max), average (Avg), and variance (Var). The experiments were conducted on the proteomics dataset with the noise level fixed at $\sigma = 0$. To increase the time-varying degree of the data, θ_0 was set to $[0.8, 1.6, 2.4]$. For each value of θ_0, α was set to $[10\%, 20\%, 30\%]$. The model was trained using the first 800-time steps and then used to predict the subsequent 96-time steps. Each combination of θ_0 and α was evaluated under five different random seeds.

Table 2. The impact of Fourier filter parameter α and time-varying degree θ on the predictive performance on TVNN model.

θ	**α**	**RMSE**			
		Min	Max	Avg	Var
0.8	10%	0.0852	0.1791	0.1098	1.557×10^{-3}
	20%	0.0417	0.1279	0.0821	1.860×10^{-3}
	30%	0.0444	0.0613	0.0481	5.462×10^{-5}
1.6	10%	0.0501	0.1203	0.0815	7.546×10^{-4}
	20%	0.0475	0.5367	0.1557	4.544×10^{-2}
	30%	0.0308	0.1020	0.0552	7.989×10^{-4}
2.4	10%	0.0012	0.0081	0.0033	7.692×10^{-6}
	20%	0.0015	0.0035	0.0023	7.939×10^{-7}
	30%	0.0014	0.0054	0.0031	3.092×10^{-6}

The results in Table 2 indicate that both Fourier filter parameter α and time-varying degree θ have a substantial impact on the predictive performance of TVNN, as follows:

 doi:10.20944/preprints202603.1888.v1

(1) When the system exhibits a relatively low degree of time variation ($\theta = 0.8$), the average RMSE decreases as α increases. In contrast, when the system becomes highly time-varying ($\theta = 2.4$), the average RMSE first decreases and then increases with increasing α.

(2) For a fixed α, as the time-varying degree θ increases from 0.8 to 2.4, the minimum, maximum, and average RMSE values all show an overall downward trend.

(3) As θ increases from 0.8 to 2.4, the variance of RMSE generally decreases.

Overall, the results in Table 2 demonstrate that TVNN achieves not only greater predictive accuracy but also better stability and robustness under stronger time-varying conditions, which further confirms the effectiveness of the proposed model to predict time-varying biological systems.

3.2.3. Ablation Experiment

To verify the effectiveness of the Fourier filter-based frequency-domain decomposition, we further conducted ablation experiments. Specifically, while keeping all other settings unchanged, we constructed both TVNN^{g} using only the global prediction module and TVNN^{l} using only the time-varying prediction module. Each model was then trained under five different random seeds. Finally, the average values of the predictive metric RMSE for TVNN, TVNN^{g}, and TVNN^{l} were summarized in Table 3, where the best average results are highlighted in bold.

Table 3. The ablation experiments on TVNN model.

Datasets	**TVNN**		**TVNN^{g}**		**TVNN^{l}**	
	RMSE	P-val	RMSE	P-val	RMSE	P-val
Proteomics (16 Steps)	**6.431×10^{-2}**$\pm7.045\times10$	—	$6.876\times10^{-2}\pm1.941\times10$	3.054×10^{-1}	$7.641\times10^{-2}\pm3.197\times10$	1.171×10^{-1}
Proteomics (48 Steps)	**6.723×10^{-2}**$\pm5.823\times10$	—	$6.920\times10^{-2}\pm1.378\times10$	3.680×10^{-1}	$7.747\times10^{-2}\pm1.748\times10$	**4.876×10^{-2}**
Gene (16 Steps)	**1.695×10^{-1}**$\pm3.229\times10$	—	$1.795\times10^{-1}\pm1.273\times10$	1.119×10^{-1}	$3.556\times10^{-1}\pm1.759\times10$	**1.590×10^{-2}**
Climate (16 Steps)	**9.963×10^{-3}**$\pm4.405\times10$	—	$1.078\times10^{-2}\pm2.823\times10$	1.590×10^{-1}	$1.217\times10^{-2}\pm4.978\times10$	**2.428×10^{-2}**
Climate (32 Steps)	**1.173×10^{-2}**$\pm3.007\times10$	—	$1.247\times10^{-2}\pm8.692\times10$	1.692×10^{-1}	$1.589\times10^{-2}\pm7.777\times10$	**1.800×10^{-2}**

Table 3 shows that TVNN achieves the best overall performance among the three models, indicating the benefit of jointly modeling global and time-varying dynamics in biological time-series prediction. To further examine the contribution of each component, we conducted T-test [47–52] between TVNN and its two ablated variants, TVNN^{g} and TVNN^{l}. Compared with TVNN, TVNN^{g} yields worse average predictive performance, but the differences are not statistically significant ($p > 0.05$). In contrast, TVNN^{l} underperforms TVNN, and the differences are statistically significant on

 doi:10.20944/preprints202603.1888.v1

most datasets ($p < 0.05$). These results suggest that both modules are necessary for accurate forecasting: the global predictive module captures substantial long-term and stable information of the system, while the time-varying predictive module characterizes local dynamic variations. Their joint modeling enables TVNN to preserve global structure while adapting to temporal changes, thereby achieving the best predictive performance.

3.3. Experimental Results and Analysis on TVPRLLM

To address the second scientific question, we first analyze the characteristics of biological time-varying patterns represented by the Koopman transition matrices. After that, we further investigate the potential biological significance underlying these time-varying patterns with the aid of the TVPRLLM model.

3.3.1. Biological Time-Varying Patterns

Figure 4 shows the heatmaps of the time-varying Koopman matrices ($\mathcal{K}^1$) at different periods on the influenza disease dataset. In the heatmaps, the red, blue and white indicate positive matrix values, negative matrix values, and zero values, respectively. The darker the color, the greater the magnitude of the corresponding matrix value.

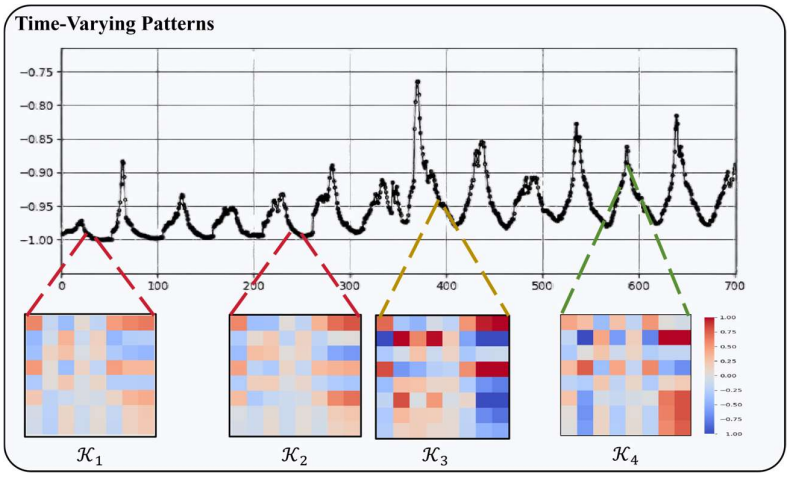

Figure 4. Time-varying Koopman matrix heatmaps on ILI dataset.

It can be observed that the time-varying Koopman matrices ($\mathcal{K}^1$) exhibit distinctly different temporal patterns across different stages. Specifically, the heatmaps of ($\mathcal{K}_1$) and ($\mathcal{K}_2$) show strong overall similarity, with relatively small values that are more uniformly distributed. In contrast, the heatmap of ($\mathcal{K}_3$) contains more extreme big and small values. Likewise, the heatmap of ($\mathcal{K}_4$) includes many extreme values and shows a certain complementary trend with ($\mathcal{K}_3$).

This phenomenon can be explained by the fact that the sequences corresponding to ($\mathcal{K}_1$) and ($\mathcal{K}_2$) both display similar stable trends. As a result, the color variations in the heatmaps of ($\mathcal{K}_1$) and ($\mathcal{K}_2$) are relatively smooth, mainly consisting of light blue and light red, with only minor differences in intensity. In contrast, ($\mathcal{K}_3$) corresponds to a sharply decreasing sequence, so its heatmap exhibits more pronounced color changes. The dark red and dark blue regions indicate values with great absolute magnitudes. Likewise, the sequence represented by ($\mathcal{K}_4$) shows a steep upward trend, and its heatmap displays pronounced color variations with many dark red and dark blue areas. Furthermore, it can be observed that the heatmap of ($\mathcal{K}_3$) corresponding to a decreasing pattern and the heatmap of ($\mathcal{K}_4$) corresponding to an increasing pattern exhibit opposite sign relationships in the regions with large absolute values, as reflected by the reversed red–blue color distribution. Therefore, the time-varying Koopman matrices in the ILI dataset can, to some extent, reflect whether the disease incidence is in a decreasing or increasing phase during a specific period.

 doi:10.20944/preprints202603.1888.v1

3.3.2. Interpretation of Biological Time-Varying Patterns

To further explore the information contained in the time-varying Koopman matrices and investigate the states of biological time-varying systems at different stages, this study systematically evaluates the effects of different Koopman matrix combinations on the accuracy of predicting the current state of biological time-varying systems based on the proposed TVPRLLM model. Specifically, we examine the performance of the time-varying Koopman matrix ($\mathcal{K}^l$), the combination of the global Koopman matrix and the time-varying Koopman matrix ($\mathcal{K}^g$+$\mathcal{K}^l$) in the state predictive task. F1-Score and AUPRC [53] are adopted as the evaluation metrics.

Table 4 presents the predictive results of different Koopman matrix combinations to identify the current states of biological time-varying systems on the TVPRLLM model, with the best average results highlighted in bold. The results show that, in both proteomics and gene domains, using only the time-varying Koopman matrix ($\mathcal{K}^l$) achieves the best predictive performance on both F1-Score and AUPRC. In contrast, after introducing the global Koopman matrix ($\mathcal{K}^g$), the predictive performance decreases under the ($\mathcal{K}^g + \mathcal{K}^l$) settings.

Table 4. The predictive results of different Koopman matrices on TVPRLLM model.

Datasets	$\mathcal{K}^l$				$\mathcal{K}^g$+$\mathcal{K}^l$			
	F1-Score		AUPRC		F1-Score		AUPRC	
	Ave±Var	P-val	Ave±Var	P-val	Ave±Var	P-val	Ave±Var	P-val
Proteomics	0.969±1.2e^{-5}	—	0.943±4.5e^{-5}	—	0.163±0.003	**4.23e^{-5}**	0.223±4.8e^{-4}	**2.19e^{-6}**
Gene	**0.781**±0.001	—	**0.659**±0.003	—	0.577±0.016	**0.041**	0.479±0.008	**0.040**

To further evaluate the statistical reliability of these differences, we conducted t-tests between the ($\mathcal{K}^l$) and ($\mathcal{K}^g + \mathcal{K}^l$) setting. The test results show that the predictive performance of TVPRLLM using only the time-varying Koopman matrix ($\mathcal{K}^l$) is significantly better than that under the ($\mathcal{K}^g + \mathcal{K}^l$) setting. This indicates that the time-varying Koopman matrix captures more discriminative dynamic patterns for biological state identification, whereas the introduction of the global Koopman matrix may partially obscure such time-varying information. Therefore, the LoRA fine-tuned TVPRLLM model can effectively identify the time-varying patterns of biological systems encoded by Koopman matrices.

3.4. The Biological Time Series Data Predictive Platform

To answer our third scientific question, we developed an online predictive platform based on TVNN model with three key functions: performance analysis, biological time-Series data prediction, and data download. Figure 5(a) visually presents the predicted values and variation trends of proteomic time-series data, thereby helping users evaluate the accuracy and reliability of the TVNN model predictions. Figure 5(b) shows the predictive results generated by the TVNN model based on the user-input biological time series. Figure 5(c) presents the data download module, where users can click “Download” to obtain the data.

 doi:10.20944/preprints202603.1888.v1

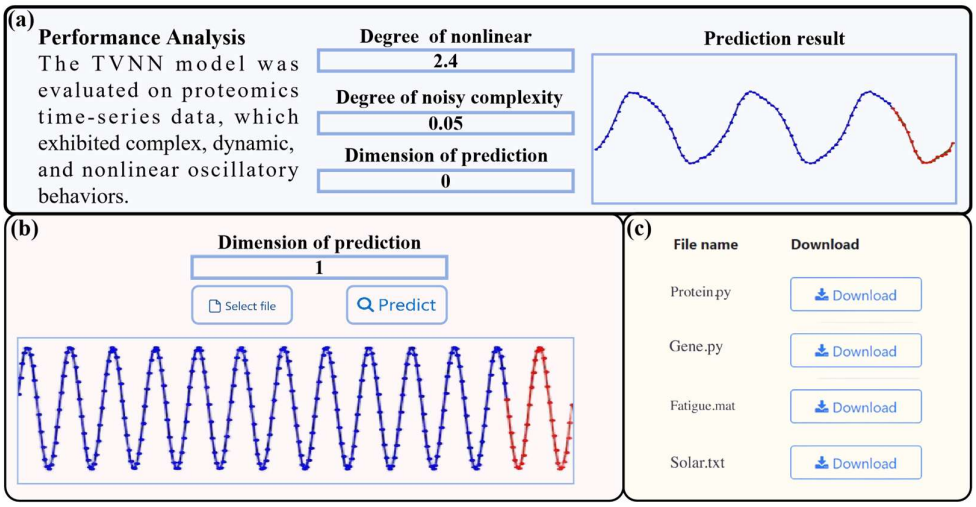

Figure 5. The biological time series data predictive platform. (a) The performance analysis module; (b) The biological time-series data prediction module; (c) The data download module.

4. Discussion

To address the challenges of biological time-series prediction, this study proposes a TVNN model that combines frequency-domain information with Koopman embedding theory. Furthermore, we introduced a TVPRLLM model to interpret the biological significance of the hidden time-varying patterns in biological systems.

Specifically, TVNN is developed to effectively extract the non-stationary characteristics of biological time-series data and model the nonlinear dynamical evolution of time-varying biological systems, thereby enabling accurate prediction of temporal shift data. In addition, a large language model collaboratively fine-tuned with LoRA is employed to construct TVPRLLM, which is used to identify time-varying patterns in biological systems and uncover their potential biological semantics. Comprehensive experiments conducted on multiple biological time-series datasets systematically validate the effectiveness of the proposed methods through model comparison, stability analysis, ablation studies, and time-varying pattern recognition experiments.

In terms of the overall model performance, Table 1 shows that TVNN achieves the superior overall performance, indicating its stronger modeling capability to predict time-varying biological systems. The fact that TVNN outperforms KAE, EKATP, and Koopa demonstrates that the collaborative design of the global prediction module and the time-varying prediction module can more effectively capture both global patterns and local time-varying dynamics in biological time series data. The weaker performance of VARIMA and SVR suggests that traditional statistical and regression-based methods are not better than Koopman theory-based methods in handling complex nonlinear and strongly time-varying biological systems.

Table 2 reveals the role of Fourier-domain decomposition in TVNN. When the system has a low degree of time variation, its evolution is dominated mainly by the Koopman operator associated with low-frequency global patterns. Therefore, increasing α appropriately increases the ability of TVNN to extract global patterns, leading to an improved predictive performance. By contrast, when the system is highly time-varying, its evolution is governed more by the Koopman operator associated with high-frequency time-varying patterns. In this case, moderately increasing α facilitates a more effective separation of global and time-varying components, thereby improving predictive accuracy. However, when α becomes excessively large, the proportion of retained time-varying components decreases, weakening the ability of TVNN to characterize temporal dynamics and ultimately degrading predictive performance.

 doi:10.20944/preprints202603.1888.v1

The ablation experiment results listed in Table 3 further confirm that the dynamics of time-varying biological systems depend on both global long-term trends and local time-varying changes. The global predictive module mainly models the overall evolutionary trend of the system, whereas the time-varying predictive module captures local dynamics and nonstationary characteristics. Because biological time-series data are typically highly time-varying, local changes have a more direct impact on prediction errors; therefore, removing the time-varying predictive module leads to a greater statistically significant performance drop. By contrast, when only the global predictive module is retained, the model can still capture the main evolutionary trend, so the performance degradation is relatively limited. Overall, the joint modeling of these two modules is essential for accurate biological time-series forecasting and further validates the effectiveness of the frequency-domain decomposition strategy.

Figure 4 shows that the time-varying Koopman matrix ($\mathcal{K}^l$) can effectively characterize the time-varying patterns of biological systems across different periods, as reflected by the variations in the heatmaps. To further explore the information contained in the time-varying Koopman matrices, Table 4 indicates that the proposed time-varying predictive module can more effectively capture the dynamic characteristics of biological time-varying systems, thereby increasing the predictive accuracy of the TVNN model. In addition, Table 4 indicates that the global Koopman matrix ($\mathcal{K}^g$) may obscure the discriminative features of time-varying patterns, as global information is insufficient to characterize the temporal variability of biological systems.

Overall, these findings indicate that our proposed framework not only improves predictive performance for biological time-series data in time-varying systems, but also provides an effective way to discover and interpret hidden dynamic patterns with potential biological significance.

5. Conclusion

In conclusion, the proposed TVNN effectively addresses biological time-series prediction in time-varying systems by combining frequency-domain decomposition with Koopman embedding theory. Experimental results demonstrate that TVNN can successfully disentangle global evolutionary patterns and local time-varying dynamics, leading to superior predictive performance, robustness, and consistency across benchmark and real biological datasets. Furthermore, TVPRLLM provides an interpretable way to analyze the biological significance of the dynamic patterns captured by TVNN. The developed platform also offers a visual and user-friendly tool for biological time-series analysis and prediction, further facilitating the application of the proposed methods in systems biology.

Despite these promising results, there still remain several challenges, including the non-stationarity and heterogeneity of biological data, as well as the need for stronger interpretability and generalization across diverse datasets. Thus, our future work will focus on integrating richer biological prior knowledge, extending the framework to multimodal data, and further increasing model robustness and practical applicability in real-world biomedical research.

Author Contributions: Methodology, Y.Y., L.Z. and Y.J.; investigation, Y.Y., Y.J. and F.Z.; resources, Y.Y., M.X. and Y.J.; writing—original draft preparation, Y.Y., and Y.J.; writing—review and editing, Y.Y. and L.Z.; supervision, M.X.; funding acquisition, F.Z. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from National Natural Science Foundation of China [62372316]; Noncommunicable Chronic Diseases-National Science and Technology Major Project [2024ZD0532900]; Sichuan Science and Technology Program key project [2025YFHZ0066]; Sichuan Science and Technology Program (2025NSFSC2088).

Informed Consent Statement: Not applicable.

 doi:10.20944/preprints202603.1888.v1

Data Availability Statement: The datasets and codes are publicly available at https://github.com/347251369/Time-Varying-Biological-Time-Series-Prediction. The platform can be available at: http://www.combio-lezhang.online/QCETBBTSPP/home/.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang L, Liu G, Kong M, Li T, Wu D, Zhou X, Yang C, Xia L, Yang Z, Chen L: Revealing dynamic regulations and the related key proteins of myeloma-initiating cells by integrating experimental data into a systems biological model. Bioinformatics 2021, 37(11):1554-1561.
2. Xiao M, Liu G, Xie J, Dai Z, Wei Z, Ren Z, Yu J, Zhang L: 2019nCoVAS: Developing the Web Service for Epidemic Transmission Prediction, Genome Analysis, and Psychological Stress Assessment for 2019-nCoV. IEEE/ACM Trans Comput Biol Bioinform 2021, 18(4):1250-1261.
3. Cao Y, Geddes TA, Yang JYH, Yang P: Ensemble deep learning in bioinformatics. Nature Machine Intelligence 2020, 2(9):500-508.
4. Li W, Yang X, Liu W, Xia Y, Bian J: DDG-DA: Data Distribution Generation for Predictable Concept Drift Adaptation. In.; 2022: 4092-4100.
5. Liu S, You Y, Tong Z, Zhang L: Developing an Embedding, Koopman and Autoencoder Technologies-Based Multi-Omics Time Series Predictive Model (EKATP) for Systems Biology research. Front Genet 2021, 12:761629.
6. Wu W, Song L, Yang Y, Wang J, Liu H, Zhang L: Exploring the dynamics and interplay of human papillomavirus and cervical tumorigenesis by integrating biological data into a mathematical model. BMC Bioinformatics 2020, 21(Suppl 7):152.
7. Ma X, Zhang Y, Wang Y: Performance evaluation of kernel functions based on grid search for support vector regression. In: 2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). 2015: 283-288.
8. Zhang J, Bai X, Shan G: Comparison of Two Different Kernel Functions of Support Vector Regression for Tracking Tumor Motion. In: Proceedings of the 2019 6th International Conference on Bioinformatics Research and Applications. 2019: 131-134.
9. Masarotto G: Bootstrap prediction intervals for autoregressions. International Journal of Forecasting 1990, 6(2):229-239.
10. Box GE, Pierce DA: Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American statistical Association 1970, 65(332):1509-1526.
11. Kim T, Kim J, Tae Y, Park C, Choi J-H, Choo J: Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution Shift. In: International Conference On Learning Representations. 2022.
12. Koopman BO: Hamiltonian Systems and Transformation in Hilbert Space. Proc Natl Acad Sci U S A 1931, 17(5):315-318.
13. Liu Y, Li C, Wang J, Long M: Koopa: Learning Non-stationary Time Series Dynamics with Koopman Predictors. In: Conference and Workshop on Neural Information Processing Systems. 2023.
14. Zhou T, Ma Z, Wang X, Wen Q, Sun L, Yao T, Yin W, Jin R: FiLM: Frequency improved Legendre Memory Model for Long-term Time Series Forecasting. In: Conference and Workshop on Neural Information Processing Systems. 2022.
15. Yi K, Zhang Q, Fan W, Wang S, Wang P, He H, An N, Lian D, Cao L, Niu Z: Frequency-domain MLPs are More Effective Learners in Time Series Forecasting. In: Conference and Workshop on Neural Information Processing Systems. 2023.
16. Jiang Z, Cheng D, Qin Z, Gao J, Lao Q, Ismoilovich AB, Gayrat U, Elyorbek Y, Habibullo B, Tang D et al.: TV-SAM: Increasing Zero-Shot Segmentation Performance on Multimodal Medical Images Using GPT-4 Generated Descriptive Prompts Without Human Annotation. Big Data Mining and Analytics 2024, 7(4):1199-1211.
17. Yang A, Yang B, Hui B, Zheng B, Yu B, Zhou C, Li C, Li C, Liu D, Huang F et al.: Qwen2 Technical Report. CoRR 2024, abs/2407.10671.

 doi:10.20944/preprints202603.1888.v1

18. Zhang Y, Yang C, Wang J, Wang L, Zhao Y, Sun L, Sun W, Zhu Y, Li J, Wu S: BioLadder: A bioinformatic platform primarily focused on proteomic data analysis. Imeta 2024, 3(4):e215.
19. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J: The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 2016, 13(9):731-740.
20. Mias GI, Yusufaly T, Roushangar R, Brooks LR, Singh VV, Christou C: MathIOmica: An Integrative Platform for Dynamic Omics. Sci Rep 2016, 6(1):37237.
21. Anzel A, Heider D, Hattab G: MOVIS: A multi-omics software solution for multi-modal time-series clustering, embedding, and visualizing tasks. Comput Struct Biotechnol J 2022, 20:1044-1055.
22. Chen P, Liu R, Aihara K, Chen L: Autoreservoir computing for multistep ahead prediction based on the spatiotemporal information transformation. Nat Commun 2020, 11(1):4568.
23. You YJ, Tan K, Jiang ZK, Zhang L: Developing a Predictive Platform for Antimicrobial Resistance Based on a Large Language Model and Quantum Computing. Engineering-Prc 2025, 48:174-184.
24. You Y, Zhang L: AnnoAgent: a language agent for single-cell automatic annotation. Briefings in Bioinformatics 2025, 26(Supplement_1):i25-i26.
25. Gao J, Lao Q, Kang Q, Liu P, Du C, Li K, Zhang L: Boosting Your Context by Dual Similarity Checkup for In-Context Learning Medical Image Segmentation. IEEE Trans Med Imaging 2025, 44(1):310-319.
26. Bertalan T, Dietrich F, Mezić I, Kevrekidis IG: On learning Hamiltonian systems from data. Chaos: An Interdisciplinary Journal of Nonlinear Science 2019, 29(12):121107.
27. Greydanus S, Dzamba M, Yosinski J: Hamiltonian Neural Networks. In: Conference and Workshop on Neural Information Processing Systems. 2019: 15353-15363.
28. Smale S: Differential equations, dynamical systems, and linear algebra, vol. 60: Academic press; 1974.
29. Gökçe A: A Mathematical Modeling Approach to Analyse the Effect of Additional Food in a Predator-Prey Interactions with a White Gaussian Noise in Prey's Growth Rate. International Journal of Applied and Computational Mathematics 2022, 8(1):21.
30. Axtmann G, Rist U: Scalability of OpenFOAM with large eddy simulations and DNS on high-performance systems. In: High Performance Computing in Science and Engineering' 16: Transactions of the High Performance Computing Center, Stuttgart (HLRS) 2016: 2016: Springer; 2016: 413-424.
31. En L: Predictability: A problem partly solved. ECMWF Seminar Proceedings I 1995, 1.
32. Lai G, Chang W-C, Yang Y, Liu H: Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In: The 41st International ACM SIGIR Conference. 2018: 95-104.
33. Ou J, Li N, He H, He J, Zhang L, Jiang N: Detecting muscle fatigue among community-dwelling senior adults with shape features of the probability density function of sEMG. Journal of NeuroEngineering and Rehabilitation 2024, 21(1):196.
34. Wu H, Xu J, Wang J, Long M: Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In: Conference and Workshop on Neural Information Processing Systems. 2021: 22419-22430.
35. Lin S, Lin W, Wu W, Zhao F, Mo R, Zhang H: SegRNN: Segment Recurrent Neural Network for Long-Term Time Series Forecasting. CoRR 2023, abs/2308.11200.
36. Schmid PJ: Dynamic mode decomposition of numerical and experimental data. Journal of Fluid Mechanics 2010, 656:5-28.
37. Li B, Xiao M, Zeng R, Zhang L: Developing a multiomics data-based mathematical model to predict colorectal cancer recurrence and metastasis. Bmc Med Inform Decis 2025, 25(Suppl2):188.
38. Ma F, Xiao M, Zhu L, Jiang W, Jiang J, Zhang P-F, Li K, Yue M, Zhang L: An integrated platform for Brucella with knowledge graph technology: From genomic analysis to epidemiological projection. In: Frontiers in genetics. vol. 13; 2022: 981633.
39. Dinh T, Zeng Y, Zhang R, Lin Z, Gira M, Rajput S, Sohn J-y, Papailiopoulos DS, Lee K: LIFT: Language-Interfaced Fine-Tuning for Non-language Machine Learning Tasks. In: Conference and Workshop on Neural Information Processing Systems. vol. 35; 2022: 11763-11784.
40. Kline DM, Berardi VL: Revisiting squared-error and cross-entropy functions for training neural network classifiers. Neural Computing & Applications 2005, 14(4):310-318.

 doi:10.20944/preprints202603.1888.v1

41. Ridenhour BJ, Brooker SL, Williams JE, Van Leuven JT, Miller AW, Dearing MD, Remien CH: Modeling time-series data from microbial communities. The ISME Journal 2017, 11(11):2526-2537.
42. Ma X, Zhang Y, Wang Y: Performance evaluation of kernel functions based on grid search for support vector regression. In: 7th International Conference on Cybernetics and Intelligent Systems, CIS 2015, and IEEE Conference on Robotics, Automation and Mechatronics. 2015: RAM:283-288.
43. Jiang J, Lai Y-C: Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius. Physical Review Research 2019, 1(3):033056.
44. Azencot O, Erichson NB, Lin V, Mahoney MW: Forecasting Sequential Data Using Consistent Koopman Autoencoders. In.; 2020: 475-485.
45. Hyndman RJ, Koehler AB: Another look at measures of forecast accuracy. International Journal of Forecasting 2006, 22(4):679-688.
46. Gopal Krishna Patro S, Sahu KK: Normalization: A Preprocessing Stage. In.; 2015: arXiv:1503.06462.
47. Li B, Yu YC, Ma FB, Dong ZY, Zhang L: Developing a quantum computing model for sequence annotation of interferon protein. Computational and Structural Biotechnology Journal 2025, 30:134-143.
48. Jiang ZK, Dai W, Wei Q, Qin ZY, Wei R, Li MY, Chen XL, Huo Y, Liu JY, Li K et al.: Diffusion Model-Based Multi-Channel EEG Representation and Forecasting for Early Epileptic Seizure Warning. Interdisciplinary Sciences-Computational Life Sciences 2025: .
49. Huang H, Yang Y, Zhang QJ, Yang YF, Xiong ZQ, Mao SQ, Song TT, Wang YL, Liu ZQ, Bu H et al.: alveolar macrophages accelerate the progression of precancerous atypical adenomatous hyperplasia by promoting the angiogenic function regulated by fatty acid metabolism. Elife 2025, 13.
50. Zhang L, Song W, Zhu T, Liu Y, Chen W, Cao Y: ConvNeXt-MHC: improving MHC-peptide affinity prediction by structure-derived degenerate coding and the ConvNeXt model. Brief Bioinform 2024, 25(3): .
51. Xiao M, Wei R, Yu J, Gao C, Yang F, Zhang L: CpG Island Definition and Methylation Mapping of the T2T-YAO Genome. Genomics Proteomics Bioinformatics 2024, 22(2).
52. Gao J, Lao Q, Liu P, Yi H, Kang Q, Jiang Z, Wu X, Li K, Chen Y, Zhang L: Anatomically Guided Cross-Domain Repair and Screening for Ultrasound Fetal Biometry. IEEE Journal of Biomedical and Health Informatics 2023, 27(10):4914-4925.
53. Li B, Xiao X, Zhang C, Xiao M, Zhang L: DGHNN: a deep graph and hypergraph neural network for pan-cancer related gene prediction. Bioinformatics 2025, 41(7).