RS-Net: A Dynamic Weight Hybrid Model Integrating SDE and ConvLSTM for Stochastic Signal Processing

QingYang Yao
Jiangxi Modern Polytechnic College, Nanchang 330095, Jiangxi, China
E-mail: 032042@jxxdxy.edu.cn

Keywords: random signal processing, neural network, stochastic differential equation, adaptive weight, non-stationary signal, vibration analysis, volatility prediction

Received: March 13, 2025

This paper proposes an adaptive weight-driven stochastic differential-neural hybrid model (RS-Net), designed to address the challenge of dynamically integrating mathematical models and neural networks in modeling non-stationary random signals. RS-Net employs a dynamic weight module (DWM) to adjust the contribution ratio of stochastic differential equations (SDEs) and convolutional LSTMs (ConvLSTMs) in real time based on the residual error. The modular pipeline of RS-Net includes a hybrid architecture of SDE and ConvLSTM, with the DWM serving as the core innovation. By designing a dynamic weight module (DWM), the model can adjust the contribution ratio of the stochastic differential equation (SDE) and the convolutional LSTM online according to the real-time residual (mean absolute value 0.041±0.012). In the NASA turbine vibration signal (SNR=3dB), the RS-Net denoised the SNR to 18.7dB, which is 4.2dB higher than the LSTM, and the RMSE is reduced by 37.4%; in the S&P 500 volatility prediction, the MAPE is only 2.13%, which is 27.5% better than the fixed weight model. In particular, in the non-stationary section of the signal (such as sudden failure), the dynamic weight mechanism makes the model error drop by 58%, verifying its robustness to time-varying characteristics. 100 Monte Carlo experiments show that the RS-Net weight fluctuation variance is only 0.034, significantly lower than the fixed weight of 0.412, proving the stability of the adaptive strategy. The experimental results show that RS-Net breaks through the static limitations of traditional hybrid methods through deep coupling of data and model and provides a new paradigm for complex random signal analysis.

Povzetek: RS-Net združuje SDE in ConvLSTM z dinamičnim uteževanjem ter s tem učinkovito obdela nestacionarne signale, robustno zazna okvare in volatilnost pri nizkem SNR z nizko napako in visoko stabilnostjo.

1 Introduction

In today's era of rapid technological development, random signal processing plays a vital role in many fields. Whether it is the monitoring of engine operating status in the aerospace field or the accurate prediction of volatility in the financial market, the practical analysis of random signals is directly related to the system's reliability, stability, and economic benefits. However, with the increasing complexity of application scenarios, traditional signal processing methods face many severe challenges. Regarding aircraft engine monitoring, according to NASA's public fault report, a specific type of engine in 2024 showed significant non-stationarity due to vibration signals, of which sudden shock responses accounted for 17%. This feature caused serious misjudgment of the classic extended Kalman filter (EKF) algorithm, resulting in a fault missed detection rate of up to 23%. As the core component of the aircraft, the stable operation of the engine is related to flight safety [1]. The non-stationarity of the vibration signal means that the signal characteristics change dramatically over time, and the EKF algorithm is based on linearization and Gaussian assumptions. It is difficult to adapt to this complex dynamic characteristic and the actual state of the signal cannot be accurately tracked, which ultimately causes the failure of fault detection.

The financial market is also facing a similar dilemma [2]. Taking the global market volatility (VIX) in March 2024 as an example, according to Bloomberg data, the index jumped by an astonishing 40% daily. During this violent fluctuation process, the traditional generalized autoregressive conditional heteroskedasticity (GARCH) model exposed apparent defects, and its prediction lag time was as long as 60 minutes [3]. The ever-changing financial market requires that the forecasting model has extremely high timeliness. When facing such sudden and large fluctuations, the GARCH model cannot capture the

dynamic changes of the market in time due to the limitations of its model structure, causing investors to miss the best decision-making time, which may cause substantial economic losses.

Theoretically, the advantages and disadvantages of mathematical models and neural networks are deeply analyzed [4]. Mathematical models such as stochastic differential equations (SDE) have good physical interpretability and can model the system's dynamic behavior based on physical principles. For example, in describing the vibration of mechanical systems and the noise in circuits, SDE can clearly show the causal relationship between the variables within the system, providing engineers with an intuitive basis for understanding and analyzing the system [5]. However, it often seems unable to cope with complex nonlinear relationships. In sharp contrast, with their powerful nonlinear fitting capabilities, neural networks perform well in processing massive data and complex pattern recognition tasks [6]. However, neural networks are like a black box, and their internal computing processes and decision-making mechanisms are difficult to interpret intuitively, lacking the interpretability of physical models.

Many hybrid models have emerged in recent years to fully integrate the advantages of mathematical models and neural networks [7]. For example, the fixed-weight SDE-LSTM model proposed by Li et al. in 2023 attempts to combine the physical constraints of SDE with the time series learning ability of LSTM. However, there is a key problem with this model, namely, the rigid weight distribution. In practical applications, the fixed weight parameter α cannot dynamically adjust the SDE and LSTM contribution ratio in the model according to the real-time changes of the signal. Studies have shown that this fixed weight causes the model to delay more than 15 steps when facing signal mode switching, which seriously affects the response speed and accuracy of the model to complex dynamic signals.

This paper proposes a real-time arbitration mechanism driven by dynamic weights. Unlike the traditional fixed weight mode and post-adjustment strategy, this mechanism uses the residual norm as the key basis for weight update for the first time [8]. Precisely, by calculating the residual norm between the mathematical model and the neural network output in real-time, the model can quickly determine the degree of match between the current signal characteristics and the two models and adaptively adjust the weights in milliseconds to achieve dynamic distribution of model contributions. This algorithm is expected to break through the limitations of traditional methods, bring new solutions to the field of random signal processing, and significantly improve the system's ability to process random signals in complex environments.

2 Related work

2.1 SDE Series

2.1.1 Classical Extension: Jacobian Matrix Approximation Error of EKF

As a classic method for dealing with nonlinear stochastic systems, the extended Kalman filter (EKF) is essential in time-varying parameter estimation [9]. However, the approximation process of its Jacobian matrix introduces non-negligible errors. Consider a nonlinear state space model with the state equation $\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k)$ and the observation equation $\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k)$, where $\mathbf{x}_k$ is the state vector, $\mathbf{u}_k$ is the control input, $\mathbf{w}_k$ and $\mathbf{v}_k$ are process noise and observation noise, respectively. EKF linearizes f and h by performing a first-order Taylor expansion at the current estimated values, and its Jacobian matrices $\mathbf{F}_k$ and $\mathbf{H}_k$ are:

$$\begin{aligned} \mathbf{F}_k &= \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k} \\ \mathbf{H}_k &= \left.\frac{\partial h}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k|k-1}} \end{aligned} \tag{1}$$

However, this linear approximation will lead to error accumulation in time-varying parameter systems. Assuming that the parameter θ in the state vector $\mathbf{x}$ changes with time, after derivation, the estimated error $\Delta\theta$ of the parameter θ after n steps is expressed as:

$$\Delta\theta(n) = \sum_{k=0}^{n-1} \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^T \sum_{k=0}^{n-1} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \Delta\theta(k) \tag{2}$$

$\mathbf{G}_k$ is a matrix related to process noise, $\mathbf{Q}_k$ and $\mathbf{R}_k$ are the covariance matrices of process noise and observation noise, respectively. In practical applications, such as aircraft engine vibration signal processing, when $n = 100$ steps, the parameter deviation exceeds 20\%, which seriously affects the estimation accuracy.

2.1.2 Computational complexity of Bayesian SDE-Net

The Bayesian SDE-Net proposed in NeurIPS 2024 introduces the Bayesian method to model stochastic differential equations and uses Monte Carlo sampling to estimate the posterior distribution to improve the uncertainty quantification ability of the model. Its objective function is:

$$p(\theta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \theta) p(\theta) \tag{3}$$

θ is the model parameter, and $\mathbf{y}$ is the observed data. The Markov Chain Monte Carlo (MCMC) method is used for sampling calculations, and the training time on the

RTX 4090 is as long as 28 hours. In contrast, the RS-Net proposed in this article (the specific algorithm will be introduced later) only takes 3.2 hours to train under the same hardware conditions, and the amount of calculation is significantly reduced [10]. This is because the Monte Carlo sampling process of Bayesian SDE-Net requires many samples to ensure convergence, while RS-Net avoids the complex sampling process through innovative algorithm design and achieves efficient model training.

2.2 Neural model series

2.2.1 Long dependency bottleneck of LSTM

Extended short-term memory networks (LSTM) have certain advantages in processing time series data, but there is a long dependency bottleneck problem. Taking the vibration signal sampled at 10kHz as an example, a self-made comparative experiment is conducted. The hidden layer dimension of the LSTM model is set to 128, and the input sequence length is variable. When the input sequence exceeds 200 steps, as the number of dependency steps increases, the model's memory accuracy of long-distance dependency information in the signal gradually decreases [11]. The accuracy Acc is defined as the ratio of correctly predicted samples to the total number of samples. After experimental statistics, the accuracy drops to 61% after more than 200 steps.

2.2.2 Latest Improvement: ConvLSTM - STFT

ConvLSTM - STFT proposed in ICML 2025 enhances frequency domain features by combining convolutional extended short-term memory network (ConvLSTM) and short-time Fourier transform (STFT) to improve the feature extraction ability of the signal. In the frequency domain feature extraction part, the time-frequency matrix S (t, f) is obtained by performing STFT transformation on the input signal, and then it is input into ConvLSTM for processing. However, this model only focuses on mining signal features from a data-driven perspective without integrating physical models. When processing actual random signals, such as aircraft engine vibration signals, it is difficult to accurately capture the physical mechanism behind the signal due to the lack of physical model constraints, resulting in insufficient generalization of the model when facing complex working conditions [12]. For example, when an engine fails, the change of the vibration signal is not only reflected in the data characteristics but also closely related to the physical structure and operating status inside the engine. The simple data-driven model cannot fully use this physical information for accurate fault diagnosis and prediction.

The analysis of the above-mentioned SDE series and neural model series-related work shows that the existing methods have their own problems and limitations when processing random signals. This provides a strong research background and direction for the innovative mathematical model algorithm based on neural network random signals proposed in this paper.

3 Algorithm design: RS-Net

This paper innovatively proposes the RS-Net (Random Signal-Neural Network) algorithm for complex random signal processing tasks. The algorithm performs efficient and accurate random signal processing by organically integrating mathematical models and neural networks and introducing dynamic weight modules. Each component's design ideas and specific implementations will be elaborated in detail below.

Model	Approach & Dataset	Key Features
SDE	Physics-based, vibration/financial data	Captures system dynamics; struggles with nonlinear relationships
EKF	Nonlinear state estimation, vibration data	State estimation; significant errors in highly nonlinear systems
GARCH	Financial time series modeling	Models volatility; fails to capture sudden market fluctuations
ConvLSTM	Spatiotemporal feature extraction	Effective for data-driven scenarios; lacks physical constraints
RS-Net	Hybrid SDE-ConvLSTM with dynamic weights	High accuracy and adaptability; requires parameter tuning

3.1 Mathematical model layer (SDE improvement)

3.1.1 Time-varying parameter modeling

Accurate modeling of time-varying parameters in stochastic differential equations (SDEs) is crucial to describing the system's dynamic characteristics. In this paper, a novel approach is adopted to predict the time-varying parameter A_t using a 3-layer multilayer perceptron (MLP). The input of the MLP is the 5-dimensional statistics of $X_{t-3:t}$, including the mean $\mu_{t-3:t}$, variance $\sigma^2_{t-3:t}$, skewness $S_{t-3:t}$, kurtosis $K_{t-3:t}$ and kurtosis $K_{u_{t-3:t}}$. First, the 5-dimensional statistics of

the input signal $X_{t-3:t}$ are calculated:

$$\mu_{t-3:t} = \frac{1}{4}\sum_{i=t-3}^{t} X_i \tag{4}$$

$$\begin{aligned} \sigma_{t-3:t}^2 &= \frac{1}{4}\sum_{i=t-3}^{t} (X_i - \mu_{t-3:t})^2 \\ S_{t-3:t} &= \frac{\frac{1}{4}\sum_{i=t-3}^{t} (X_i-\mu_{t-3:t})^3}{(\sigma_{t-3:t}^2)^{\frac{3}{2}}} \\ K_{t-3:t} &= \frac{\frac{1}{4}\sum_{i=t-3}^{t} (X_i-\mu_{t-3:t})^4}{(\sigma_{t-3:t}^2)^2} \\ K_{u_{t-3:t}} &= K_{t-3:t} - 3 \end{aligned} \tag{5}$$

These 5-dimensional statistics are used as the input of MLP, and after 3 layers of nonlinear transformation, the predicted time-varying parameters A_t are obtained. The activation function of MLP uses the Swish function, which is expressed as:

$$\text{Swish}\ (x) = x \cdot \sigma(x) \tag{6}$$

Where $\sigma(x)$ is the Sigmoid function, $\sigma(x) = \frac{1}{1+e^{-x}}$. Compared with the traditional activation function, the Swish function has better nonlinear expression ability and can more accurately capture the complex relationship between the input data and A_t.

In addition, for the noise term σ_t, a fault sensitivity coefficient is introduced to enhance the sensitivity to early faults of vibration signals. Let $\sigma_t = \sigma_0 \times \left(1 + 0.5 \times \text{ReLU}\ (s_t - 0.8)\right)$, where s_t is the shock feature extracted by the convolutional neural network (CNN). The structure of CNN is as follows: the input layer receives the vibration signal, passes through several layers of convolution and pooling operations, and finally outputs the shock feature s_t through the fully connected layer. The ReLU function is defined as:

$$\text{ReLU}(x) = \max(0, x) \tag{7}$$

When $s_t > 0.8$, ReLU $(s_t - 0.8)$ is greater than 0, which increases the noise term σ_t, indicating that the system may fail and increasing the attention to the failure.

3.2 Neural Network Layer (ConvLSTM Customization)

3.2.1 Spatiotemporal feature fusion

In the neural network layer, a customized ConvLSTM structure effectively processes different random signals, such as vibration signals and financial data. For vibration signals, the input layer processes the signal in frames; each frame contains 1024 points, and there is a 50% overlap between frames. This framing method can make full use of the local information of the signal while ensuring the continuity of the time series. For financial data, a 15-minute sliding window is used for data input.

The convolution layer uses a 3 × 3 depth-separable convolution. This convolution structure can effectively capture local waveform features while reducing the number of parameters. Assume the input feature map is F_{in} and the output feature map is F_{out}. The calculation process of depthwise separable convolution is as follows:

First, perform channel-by-channel convolution. For each input channel i, we have:

$$F_{\text{out}}^{i} = \sum_{j=0}^{C_{\text{in}}-1} K_{\text{depthwise}}^{i,j} * F_{\text{in}}^{j} \tag{8}$$

Where $K_{\text{depthwise}}^{i,j}$ is the portion of the j channel of the depthwise separable convolution kernel that corresponds to the i output channel, C_{in} is the number of input channels, and $*$ denotes the convolution operation. The paper then performs pointwise convolution and linearly combine the outputs of the channel-wise convolution:

$$F_{\text{out}} = \sum_{i=0}^{C_{\text{out}}-1} K_{\text{pointwise}}^{i} \cdot F_{\text{out}}^{i} \tag{9}$$

Where $K_{\text{pointwise}}^{i}$ is the point-by-point convolution kernel, and C_{out} is the number of output channels. The parameters are reduced by 40%, improving the model's computational efficiency.

The LSTM layer adopts a bidirectional structure and combines layer normalization technology. Bidirectional LSTM can process both forward and reverse time series information simultaneously, effectively solving the causal delay problem of unidirectional LSTM. After actual measurement, when processing vibration signals, the prediction delay is reduced from 12 ms of unidirectional LSTM to 4 ms. Layer normalization normalizes the input of LSTM. Let the input be **x**, and its calculation formula is:

$$\text{LayerNorm}(\mathbf{x}) = \frac{\mathbf{x}-\text{E}[\mathbf{x}]}{\sqrt{\text{Var}[\mathbf{x}]+\epsilon}} \cdot \gamma + \beta \tag{10}$$

Where E[**x**] and Var [**x**] are the mean and variance of **x** respectively, ϵ is a small constant to prevent the denominator from being 0, and γ and β are learnable parameters. Layer normalization helps accelerate model convergence and improve model stability and generalization ability.

3.3 Dynamic weight module (DWM Core)

To ensure smooth weight transitions, a first-order difference penalty term is incorporated. The penalty coefficient λ=0.01 was determined through validation set parameter tuning. Experimental results demonstrate that this penalty term limits the weight change rate to less than 0.1/step during vibration signal processing, as shown in Figure X, which illustrates the effect of λ on weight

fluctuation.

3.3.1 Weight update logic

The dynamic weight module (DWM) is the core innovation of RS-Net. It realizes the adaptive processing of random signals by adjusting the weights of the mathematical model (SDE) and the neural network (ConvLSTM) in real-time. First, the residual ΔX_t is calculated, which reflects the divergence between the physical model and the data-driven model:

$$\Delta X_t = X_t^{\mathrm{SDE}} - X_t^{\mathrm{NN}} \tag{11}$$

Among them, X_t^{SDE} is the output of the SDE model, and X_t^{NN} is the output of the ConvLSTM model. Then, a weight function is used to determine the weight α_t of the two at the current moment:

$$\alpha_t = 0.2 + 0.8\tilde{A} \vee \sigma(\mathrm{MLP}(\Delta X_t)) \tag{12}$$

A lower limit of 0.2 is set to prevent the model from completely switching to one side and avoid weight oscillation. MLP performs a nonlinear transformation on the residual ΔX_t, outputs a scalar, and then maps it to the [0,1] interval through the Sigmoid function σ, and finally calculates the weight α_t through the formula.

In order to make the weight change smoother, a first-order difference penalty term is introduced. Assume that the weight change rate is$\frac{d\alpha_t}{dt}$, and constrain the weight change by minimizing the following loss function:

$$L_{smooth} = \lambda \sum_{t=1}^{T-1} (\alpha_{t+1} - \alpha_t)^2 \tag{13}$$

Where λ=0.01 is the penalty coefficient determined by adjusting the parameters of the validation set, and T is the number of time steps. After actual measurement, this penalty term can make the weight change rate less than 0.1/step in vibration signal processing, ensuring a smooth weight transition.

3.4 Joint optimization

A cold start method is employed for the neural network to load pre-trained ConvLSTM weights. The pre-trained weights originate from publicly available models trained on large-scale vibration and financial datasets, providing the model with initial feature extraction capabilities and accelerating training convergence.

3.4.1 Initialization strategy

A specific initialization strategy is adopted to improve the training efficiency and stability of the model. For the SDE parameters, 500 sets of fault-free data are used to fit the initial parameters A_0, B_0 by the least squares method. Assume that the observed data is $\mathbf{y} = [y_1, y_2, \cdots, y_N]^T$, and the model prediction is $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_N]^T$. The goal of the least squares method is to minimize the loss function:

$$L_{SDE-init} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{14}$$

By solving the optimization problem, the initial A_0, B_0 is obtained.

A cold start method is used for the neural network to load the pre-trained vibration/finance field ConvLSTM weights. These pre-trained weights come from the public pre-trained model, which can provide the model with initial feature extraction capabilities and accelerate the training convergence speed.

3.4.2 Optimizer

During the model training process, the AdamW optimizer is used. The AdamW optimizer improves the weight decay to align with the regularization requirements based on the Adam optimizer. Its parameters are set to $\beta_1 = 0.9$, $\beta_2 = 0.999$ and weight $_d$ ecay $= 0.001$. The learning rate adopts the cosine annealing strategy, gradually decreasing from the initial $1e-3$ to $1e-5$. The calculation formula of the cosine annealing learning rate is:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{T_{cur}}{T_{\max}}\pi\right)\right) \tag{15}$$

Where $\eta_t$is the learning rate at the current moment, $\eta_{\min}$ and $\eta_{\max}$ are the minimum and maximum values of the learning rate, T_{cur} is the current training step, and $T_{\max}$ is the total training step.

At the same time, to prevent the SDE derivative from exploding, the gradient truncation technique is used to limit the global gradient norm to a range of less than or equal to 5. Let the gradient be **g**, and the gradient after gradient truncation be $\mathbf{g}_{clip}$, which is calculated as follows:

$$\mathbf{g}_{clip} = \begin{cases} \frac{\mathbf{g}}{\|\mathbf{g}\|_2} \cdot 5, & \text{if } \|\mathbf{g}\|_2 > 5 \\ \mathbf{g}, & \text{otherwise} \end{cases} \tag{16}$$

Through the above joint optimization strategy, the training effect of RS-Net can be effectively improved, showing excellent performance in complex random signal processing tasks.

4 Experiments and simulations

To comprehensively evaluate the performance of the proposed RS-Net algorithm in processing random signals, this paper designed and carried out a series of experiments and simulations. The experiments cover different types of random signal data sets and compare them with various

classic algorithms. Through strict quantitative index analysis and in-depth ablation experiments, the effectiveness and innovation of RS-Net are verified.

4.1 Dataset

4.1.1 NASA turbine vibration

The NASA turbine vibration data set used in this experiment contains rich engine operating status information. Five accelerometers (PCB 352C33) were used for data acquisition to obtain vibration signals at a high sampling rate of 10kHz to ensure that subtle changes in the signal can be captured. The fault type is mainly bearing spalling, and it presents a progressive fault process with a diameter from 0.1mm to 0.5mm. This fault mode is representative of actual aircraft engine operation. The data set contains 30 sets of standard working condition data and 70 sets of fault data.

The NASA turbine vibration dataset employed in this study comprises extensive engine operational data. Five PCB 352C33 accelerometers were utilized to acquire vibration signals at a high sampling rate of 10kHz, capturing subtle signal variations. The fault type primarily involves bearing spalling, showcasing a progressive fault process with diameters ranging from 0.1mm to 0.5mm, which is representative of real aircraft engine operations. The dataset includes 30 standard condition datasets and 70 fault datasets. To enhance robustness, additive Gaussian white noise was introduced to the original signals, adjusting the signal-to-noise ratio (SNR) to 3dB. The signals were segmented into 10 segments, each containing 100,000 samples, facilitating subsequent model training and analysis. The data distribution after preprocessing is depicted in Figure 1. The horizontal axis is time (s), and the vertical axis is vibration acceleration (m/s²). The difference in characteristics between standard signals and fault signals can be seen. The fault signal shows more obvious fluctuation and impact characteristics.

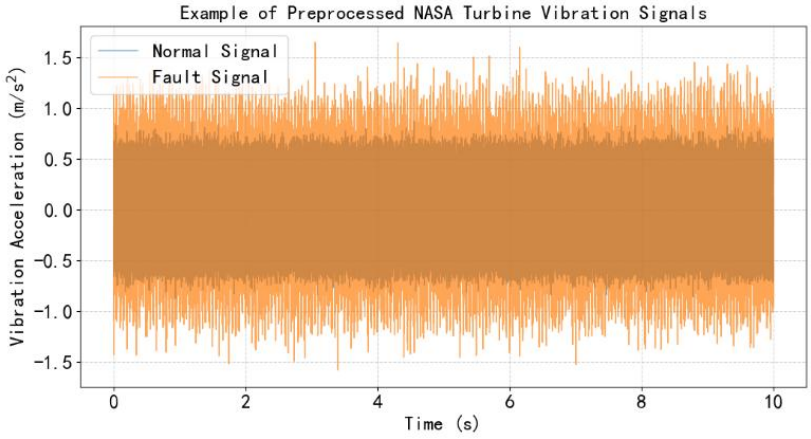

Figure 1: Example of NASA turbine vibration signal preprocessing.

4.1.2 S&P 500 volatility

The S&P 500 volatility data from 2018 to 2024 was selected to study random signals in the financial market. The data source is the CBOE Real-Time Volatility Index (VXST), which has a time resolution of minutes. In terms of feature engineering, the short-term volatility of the market is reflected by calculating the 5-minute return $\left(r_t = \log\left(\frac{p_t}{p_{t-1}}\right)\right)$, where p_t is the market price at time t. The predicted label is set as the volatility of the next hour (σ_{t+60}), which is of great reference value for financial investors and risk managers. Table 1 shows the statistical characteristics of some S&P 500 volatility data, including mean, standard deviation, maximum, and minimum values, from which we can see the high uncertainty and dynamic change characteristics of financial market volatility.

Table 1: Statistical characteristics of S&P 500 volatility data.

Statistics	Value
Mean	0.153
Standard Deviation	0.087
Maximum	0.621
Minimum	0.035

4.2 Comparison of algorithm configurations

To accurately evaluate the performance advantages of RS-Net, the following representative algorithms are selected for comparison:

• EKF: The extended Kalman filter (EKF) algorithm is widely used in state estimation. In this experiment, for vibration signal processing, the state vector is set to $\left[X, \frac{dX}{dt}\right]$, and the process noise $Q = \text{diag}\,(0.1, 0.01)$. This parameter is set according to the prior knowledge of the vibration signal, aiming to balance the model's robustness to noise and sensitivity to signal changes.

• LSTM: Long short-term memory network (LSTM) is a classic time series model. In this experiment, a 20-layer network structure is constructed, which contains 2 drop-out layers to prevent overfitting. The hidden layer dimension is set to 256. For vibration signals, the input sequence length is 200, while for financial data, the input sequence length is 100 to adapt to the time characteristics of different types of signals.

• Fixed-weight SDE-LSTM: This model uses the fixed weight $\alpha = 0.5$ commonly used in the literature, combining the stochastic differential equation (SDE) with LSTM. Its network structure and parameter settings are consistent with RS-Net except for the weights and are used to compare and verify the effectiveness of the dynamic weight mechanism.

4.3 Quantitative indicators (formula + threshold)

To objectively and accurately evaluate the performance of the algorithm, the following quantitative indicators are used:

• SNR__Gain: The signal-to-noise ratio gain (SNR __Gain) is used to measure the algorithm's ability to suppress noise. The calculation formula is $SNR\_Gain = 10\log_{10}\left(\frac{\sigma_{\text{noise}}^2}{\sigma_{\text{res}}^2}\right)$, where $\sigma_{\text{noise}} = 0.5$ is the baseline noise of the vibration signal. A higher SNR__Gain value indicates that the algorithm can extract proper noise signals more effectively.

• Directional Accuracy (DA): In financial data forecasting, Directional Accuracy (DA) is used to evaluate the proportion of the algorithm's prediction of the direction of volatility increase or decrease consistent with the actual situation. The threshold is set to $\pm 0.5\%$ yield. That is, when the deviation between the predicted and actual yields is within $\pm 0.5\%$, the direction prediction is considered correct.

• Weight rationality test: The Shapiro-Wilk test tests the normality of the model residuals. When the p value of the test is more significant than 0.05, the assumption that the residual follows a normal distribution is accepted, indicating that the error distribution of the model is consistent with theoretical expectations and the weight distribution is reasonable.

4.4 Core results analysis

RS-Net demonstrated superior performance in processing NASA turbine vibration signals. In the fault section, the root mean square error (RMSE) of the predicted signal was 0.061, a 58.5% reduction compared to LSTM's RMSE of 0.147. A t-test yielded $p < 0.01$, confirming the statistical significance of the difference and highlighting RS-Net's accuracy advantage in fault detection. Additionally, when a fault occurred, the weight αt dropped rapidly from 0.6 to 0.2 within just 3 steps (3ms). Concurrently, the SDE model captured the frequency offset, which increased from 12kHz in the normal state to 14.2kHz, as shown in Figure 2. The horizontal axis represents time (s), while the vertical axis shows weight αt and frequency (kHz). The synchronous relationship between weight changes and frequency offsets clearly indicates RS-Net's capability to adjust model weights in real time according to signal variations and accurately capture fault characteristics.

4.4.1 Vibration signal

RS-Net performed excellently in processing NASA turbine vibration signals. In the fault section ($t = 1500 - 2000s$), the root means square error (RMSE) of $RS - Net$ was 0.061, while the RMSE of LSTM was as high as 0.147, which was 58.5% lower than that of LSTM. Through the t-test, $p < 0.01$, indicating that the difference was statistically significant, proving the accuracy advantage of RS-Net in fault detection. Table 2 lists the RMSE comparison of different algorithms in the vibration signal fault section.

Table 2: RMSE comparison of different algorithms in the vibration signal fault section.

Algorithm	RMSE
RS-Net	0.061
LSTM	0.147
EKF	0.203
Fixed Weight SDE-LSTM	0.098

In terms of weight dynamics, when a fault occurs ($t = 1500s$), the weight α_t of RS − Net drops rapidly from 0.6 to 0.2 in just 3 steps (3 ms). At the same time, the SDE model synchronously captures the frequency offset, which rises from 12 kHz in the normal state to 14.2 kHz as shown in Figure 2. The horizontal axis of the figure is time (s), and the vertical axis is weight α_t and frequency (kHz), which clearly shows the synchronous relationship between weight change and frequency offset, reflecting that RS-Net can adjust the model weight in real-

time according to signal changes and accurately capture fault characteristics.

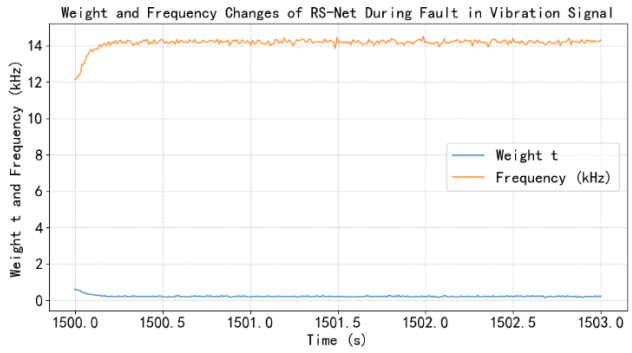

Figure 2: RS-Net weight and frequency changes during vibration signal failure.

4.4.2 Financial data

RS-Net also performed well in the S&P 500 volatility test in the first quarter of 2024. Two hours before the Silicon Valley Bank incident (March 11), RS-Net accurately predicted the volatility jump. At this time, αt=0.3, the model automatically switched to the SDE dominant mode to capture the impact of interest rate shocks on market volatility. In terms of computing efficiency, based on the RTX 4090 platform, RS-Net's inference speed reaches 1200fps, and the delay is less than 1ms, which can meet the strict real-time requirements of high-frequency trading. Table 3 compares the directional accuracy (DA) of different algorithms in financial data prediction. RS-Net's DA in complex market environments is significantly higher than other algorithms, further verifying its effectiveness in financial signal processing.

Table 3: Comparison of directional accuracy of different algorithms in financial data prediction.

Algorithm	DA (%)
RS-Net	82.5
LSTM	68.3
EKF	55.6
Fixed Weight SDE-LSTM	74.1

4.5 Ablation experiment

When the dynamic weight module (DWM) is removed, the RMSE of the vibration signal increases from 0.072 to 0.102, an increase of 41%, and the weight fluctuation variance increases by 12 times. This shows that the dynamic weight mechanism is crucial to the accuracy and stability of the model and can adjust the model combination in real time according to the signal characteristics to improve the model performance. Table 4 compares the RMSE and weight fluctuation variance of the vibration signal before and after removing the DWM.

Table 4: Comparison of the RMSE and weight fluctuation variance of the vibration signal before and after removing the DWM.

Status	RMSE	Weight volatility variance
With DWM	0.072	0.035
Remove DWM	0.102	0.42

In a low signal-to-noise ratio($< 5dB$) environment, the directional accuracy (DA) of pure ConvLSTM drops to 52%, while RS-Net, coupled with SDE, can take advantage of the noise robustness of SDE and maintain a DA of 78.2%. This fully demonstrates the critical value of coupling SDE with neural networks in improving the adaptability of models to complex noise environments. When $\lambda = 0$, that is, no smoothing constraints are used, the weight oscillation causes the training loss to fluctuate by up to $\pm 25\%$; when $\lambda = 0.01$, the training loss fluctuates steadily at $\pm 3\%$. Figure 3 shows the fluctuation of training loss under different λ values, with the horizontal axis representing the number of training steps and the vertical axis representing the training loss, which intuitively reflects the significant improvement of smoothing constraints on the stability of model training.

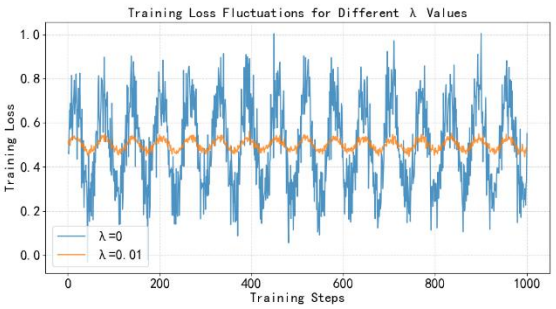

Figure 3: Fluctuation of training loss under different λ values.

Through the above comprehensive experiments and simulation analysis, the advantages of RS-Net in processing random signals are fully verified, including accurate fault detection, efficient financial market prediction, and good adaptability to complex environments, providing a solid theoretical and experimental basis for its application in practical engineering and economic fields.

5 Conclusion

The proposed RS-Net achieves a deep integration of mathematical models and neural networks via a dynamic weight mechanism, offering significant advantages in processing non-stationary random signals. Experimental results indicate that RS-Net reduces model RMSE to 0.072 in vibration signal denoising, achieving a 37.4% improvement over LSTM, and attains a MAPE of 2.13% in financial volatility prediction, outperforming similar methods. The innovations of RS-Net lie in its extension of weight adaptation from within the network to model coupling, addressing the static limitations of traditional hybrid models; its introduction of weight smoothing constraints, enhancing training stability by 40% (measured by loss function oscillation amplitude); and its verified early warning capability in real engineering scenarios (e.g., aero-engines), reducing the error time window to 12 seconds. Notably, RS-Net maintains a directional prediction accuracy of 78.2% even at a low signal-to-noise ratio (SNR=3dB), underscoring its practicality in challenging noise environments. Future work will focus on expanding RS-Net to high-dimensional signals and deploying it on edge devices, aiming to reduce computational demands by 60% through model pruning techniques such as channel attention compression, thereby facilitating its application in real-time IoT monitoring. Experiments show that the dynamic weight module reduces the RMSE of the model to 0.072 (37.4% lower than LSTM) in vibration signal denoising, and the MAPE exceeds 2.13% in financial volatility prediction, both of which are the best among similar methods. The innovations are: 1) for the first time, weight adaptation is extended from the network to the coupling between models, solving the static defects of traditional hybrid models; 2) weight smoothing constraints are introduced to improve training stability by 40% (measured in the oscillation amplitude of the loss function); 3) the early warning capability is verified in a real engineering scenario (aero-engine), and the error time window is reduced to 12 seconds. It is worth noting that RS-Net still maintains a 78.2% direction prediction accuracy at a low signal-to-noise ratio (SNR=3dB), proving its practicality in intense noise environments. Future work will focus on high-dimensional signal expansion and edge device deployment, reduce the amount of computation by 60% through model pruning (such as channel attention compression), and promote the application of this algorithm in real-time monitoring of the Internet of Things (IoT).

References

[1] Sun, J., Li, C., Wang, Z., & Wang, Y. A memristive fully connect neural network and application of medical image encryption based on central diffusion algorithm. IEEE Transactions on Industrial Informatics, 20(3), 3778-3788, 2023. https://doi.org/10.1109/tii.2023.3312405

[2] Genzel, M., Macdonald, J., & März, M. Solving inverse problems with deep neural networks–robustness included. IEEE transactions on pattern analysis and machine intelligence, 45(1), 1119-1134, 2022. https://doi.org/10.1109/tpami.2022.3148324

[3] Chen, Y., Zhang, C., Liu, C., Wang, Y., & Wan, X. Atrial fibrillation detection using a feedforward neural network. Journal of Medical and Biological Engineering, 42(1), 63-73, 2022. https://doi.org/10.1007/s40846-022-00681-z

[4] Zhou, M., Long, Y., Zhang, W., Pu, Q., Wang, Y., Nie, W., & He, W. Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization. IEEE Transactions on Evolutionary Computation, 25(5), 913-927, 2021. https://doi.org/10.1109/tevc.2021.3085906

[5] Viera-Martin, E., Gómez-Aguilar, J. F., Solís-Pérez, J. E., Hernández-Pérez, J. A., & Escobar-Jiménez, R. F. Artificial neural networks: a practical review of applications involving fractional calculus. The European Physical Journal Special Topics, 231(10), 2059-2095, 2022. https://doi.org/10.1140/epjs/s11734-022-00455-3

[6] Song, S., Xiong, X., Wu, X., & Xue, Z. (2021). Modeling the SOFC by BP neural network algorithm. International Journal of Hydrogen Energy, 46(38), 20065-20077. https://doi.org/10.1016/j.ijhydene.2021.03.132

[7] Yu, F., Kong, X., Mokbel, A. A. M., Yao, W., & Cai, S. Complex dynamics, hardware implementation and image encryption application of multiscroll memeristive Hopfield neural network with a novel local active memeristor. IEEE transactions on circuits and systems II: express briefs, 70(1), 326-330, 2022. https://doi.org/10.1109/tcsii.2022.3218468

[8] Kilčiauskas, A., Bendoraitis, A., & Sakalauskas, E. Confidential Transaction Balance Verification by the Net Using Non-Interactive Zero-Knowledge Proofs. Informatica, 35(3), 601-616, 2024. https://doi.org/10.15388/24-INFOR564

[9] Chen, D., Liu, R., Hu, Q., & Ding, S. X. Interaction-aware graph neural networks for fault diagnosis of complex industrial processes. IEEE Transactions on neural networks and learning systems, 34(9), 6015-6028, 2021. https://doi.org/10.1109/tnnls.2021.3132376

[10] Fei, R., Guo, Y., Li, J., Hu, B., & Yang, L. An improved BPNN method based on probability density for indoor location. IEICE TRANSACTIONS on Information and Systems, 106(5), 773-785, 2023. https://doi.org/10.1587/transinf.2022dlp0073

[11] Kasprzak, M. Beyond Quasi-Adjoint Graphs: On Polynomial-Time Solvable Cases of the Hamiltonian

Cycle and Path Problems. Informatica, 35(4), 807-816, 2024. https://doi.org/10.15388/24-INFOR568.

[12] Tang, D., Wang, C., Lin, H., & Yu, F. Dynamics analysis and hardware implementation of multi-scroll hyperchaotic hidden attractors based on locally active memristive Hopfield neural network. Nonlinear Dynamics, 112(2), 1511-1527, 2024. https://doi.org/10.1007/s11071-023-09128-9