



Research Article

Dynamical analysis of infectious disease models considering awareness factors and neural network numerical simulation

Minghao Song¹, Hua Liu^{1,*}, Jingyan Ma², Kai Zhang¹, Qibin Zhang³ and Yumei Wei⁴

¹ School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou 730030, China

² School of Preparatory Education, Northwest Minzu University, Lanzhou 730000, China

³ Gansu High-Tech Innovation Service Center, Lanzhou 730000, China

⁴ Experimental Teaching Department, Northwest Minzu University, Lanzhou 730000, China

* **Correspondence:** Email: jslh@xbmu.edu.cn.

Abstract: Media coverage plays a crucial role in the spread of infectious diseases. To investigate the influence of media intervention on epidemic transmission, we adopted a media-aware susceptible–infected–recovered (SIR) model based on media-intervention epidemic studies. Our primary focus was the systematic optimization of Physics-Informed Neural Network (PINN) architectures. The dynamic behavior of the epidemic system was investigated through equilibrium, stability, and bifurcation analyses. The theoretical analysis characterized the threshold dynamics of the model through the basic reproduction number R_0 , while numerical simulations further indicated that media intervention can reduce the infection peak and delay its arrival. Based on this epidemic framework, PINNs were employed for numerical simulation, and three algorithm optimization approaches were systematically investigated: (1) Replacing Euler-generated synthetic reference trajectories with RK4-generated reference trajectories to examine the influence of numerical discretization accuracy on PINN training and evaluation; (2) increasing the network depth from three to five layers to enhance the approximation of nonlinear epidemic dynamics; and (3) introducing skip-connection architectures, including Residual, Highway, and an exploratory residual-integrated Highway (ResHighway) network. Comparative results showed that the Residual network achieved the best overall accuracy and stability. Although the proposed ResHighway architecture yielded marginal accuracy gains compared with the Highway network, it underperformed the lightweight Residual network. This observation implies that overcomplicating network architectures may not bring consistent performance improvements to PINNs applied in epidemic dynamic modeling.

Keywords: media coverage; SIR model; infectious disease dynamics; PINN; residual network; highway network

1. Introduction

Infectious diseases pose significant threats to human health and social stability. Historically, large-scale pandemics such as the Black Death and smallpox caused enormous mortality [1, 2], and in recent decades outbreaks including SARS, H1N1 influenza, Ebola virus disease, and COVID-19 have further highlighted the risk of rapidly spreading pathogens in highly interconnected societies [3–6]. Despite continuous improvements in medical services and public health infrastructure, emerging infectious diseases can still spread quickly and generate substantial social and economic burdens.

Mathematical modeling remains an important tool for understanding epidemic transmission and evaluating intervention strategies. The classical susceptible–infected–recovered (SIR) framework provides a compact description of disease dynamics through key quantities such as the basic reproduction number R_0 and the stability of equilibrium points [7–9]. In particular, the threshold property of R_0 offers a principled way to distinguish disease extinction from persistence [10]. When $R_0 < 1$, the disease-free equilibrium is expected to be stable under appropriate conditions, and the epidemic tends to die out in the local stability sense.

Meanwhile, Physics-Informed Neural Network (PINN) has become a promising paradigm in scientific machine learning [11]. Since the classic PINN framework was proposed, an increasing number of studies have applied physics-driven or hybrid neural network models to the field of infectious disease modeling. For instance, Millevoi et al. proposed a split training method for PINN to address the calibration difficulty of time-varying parameters in compartmental models, and verified its advantages in accuracy and efficiency using synthetic and real epidemic datasets [12]. Similarly, Xia et al. proposed a susceptible–exposed–infected/hesitant–recovered (SEIHR) knowledge dissemination model with time-varying parameters based on PINN, further extending the application of scientific machine learning to epidemiological propagation frameworks [13]. Schiassi et al. proposed a physics-informed neural network framework for epidemiological compartmental models. Their method accurately estimated unknown model parameters from synthetic epidemic data and demonstrated the effectiveness of PINNs in inverse epidemiological problems [14]. Fujita and Akutsu proposed a physics-informed spatial identity neural network named PISID, which integrates a graph-free spatio-temporal encoding module with the SIR dynamic framework and infers epidemiological parameters via spatial embedding and neural networks. The model achieves superior prediction performance on real epidemic data and possesses good interpretability [15]. Freire et al. applied physics-informed neural networks to the classical Kermack–McKendrick epidemic model and successfully estimated epidemiological parameters from historical bubonic plague data, demonstrating the applicability of PINNs to inverse epidemic modeling [16]. In addition, researchers have further extended PINN to knowledge dissemination models with media effects, demonstrating its effectiveness in learning time-varying transmission dynamics under complex information propagation scenarios [17]. Furthermore, Mei et al. developed a direct data-driven intermittent control method for unknown dynamical systems using Lyapunov-guided attraction-region estimation and a neural feedback-loop design [18]. Nevertheless, most relevant studies rely on feedforward networks or multi layer perceptron subnetworks. The influence of typical advanced network architectures, such as Residual Network [19] and Highway Network [20], on the training characteristics of PINNs for infectious diseases has not been systematically investigated.

Studies have shown that media exposure can influence public awareness and preventive behavior during epidemics [21]. Media-awareness mechanisms have therefore been incorporated into

deterministic and stochastic epidemic models to represent behavior-induced reductions in effective contact and disease transmission [22, 23]. Aldila investigated the combined effects of media intervention and rapid testing on epidemic control, and pointed out that early high-intensity media intervention combined with moderate testing can reduce the peak number of infections and delay its arrival [24]. Cui et al. added a media influence factor to the SIR model, and proved that timely and high-intensity media reports can control the basic reproduction number below 1 to block the spread of epidemics [25]. It is clear that media effects have a great influence on epidemic transmission and should be fully considered in infectious disease modeling.

Given the research gap in the systematic comparison of different PINN network architectures and the important role of media effects in epidemic transmission, we adopt a media-aware SIR model following established modeling frameworks in the literature, where susceptible individuals are divided into aware and unaware susceptible individuals. Notably, we focus on the systematic investigation and optimization of PINN architectures for epidemic simulation. Specifically, three algorithm optimization approaches are investigated: The fourth-order Runge-Kutta (RK4) method is employed to replace the Euler method for time discretization; the network depth is increased from three to five layers to improve the approximation of nonlinear epidemic dynamics; and residual skip connections are introduced to facilitate the training of deep PINNs. Furthermore, we construct an exploratory ResHighway architecture by integrating residual connections into the Highway network for empirical comparison.

The experimental results show that the RK4 method achieves better training performance than the Euler method. Increasing the network depth further improves the model fitting capability. After introducing residual connections, the fully connected network and the Highway network exhibit improved performance. Among all investigated architectures, the Residual network achieves the best overall performance, while the proposed ResHighway architecture provides only a modest improvement over the original Highway network.

In Section 2, we present the media-aware epidemic model and analyze its mathematical properties. In Section 3, we provide theoretical validation and numerical simulations. In Section 4 we investigate the optimization of PINN architectures, and In Section 5, we conclude the paper.

2. Infectious disease model incorporating media coverage

Media coverage plays a key role in the model: By influencing the population's cognitive level of infectious diseases, it alters individuals' behavioral patterns, thereby exerting an effect on the disease transmission process. By introducing the regulatory mechanism of media coverage on public behavior, the model subdivides the susceptible population into unaware susceptible individuals (not affected by media) and aware susceptible individuals (affected by media and taking protective measures).

2.1. Model construction

In this paper, S_u represents the unaware susceptible population; S_a represents the aware susceptible population; I represents the infected population; and R represents the recovered population. This model describes the changes in these population groups through the following set of differential equations, aiming to more accurately characterize the transmission of infectious diseases under the influence of media coverage-including transmission speed, infection scale, and transmission trends of infectious diseases. Figure 1 shows the interconnections and transformation among the population groups.

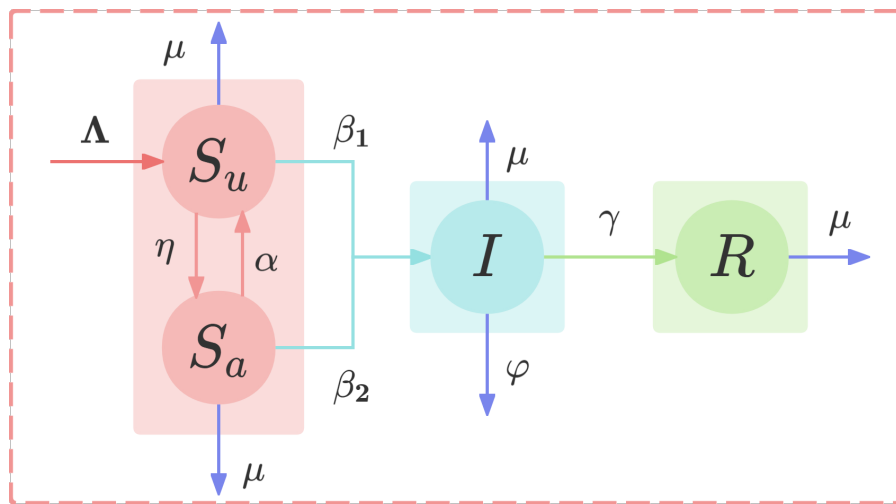


Figure 1. Compartmental diagram of the media-aware SIR epidemic model.

The model established in this study and its corresponding differential equations are as follows:

$$\begin{cases} \frac{dS_u}{dt} = \Lambda - \beta_1 S_u I - \eta S_u + \alpha S_a - \mu S_u, \\ \frac{dS_a}{dt} = \eta S_u - \alpha S_a - \beta_2 S_a I - \mu S_a, \\ \frac{dI}{dt} = \beta_1 S_u I + \beta_2 S_a I - \varphi I - \gamma I - \mu I, \\ \frac{dR}{dt} = \gamma I - \mu R. \end{cases} \quad (2.1)$$

Initial conditions are as follows: $S_u(0) \geq 0, S_a(0) \geq 0, I(0) \geq 0, R(0) \geq 0$.

In the aforementioned model, Λ is the population input rate. β_1 and β_2 are transmission rate coefficients. β_1 is the transmission rate coefficient for the unaware susceptible, and β_2 is the transmission rate coefficient for the aware susceptible. We assume $\beta_1 > \beta_2$ because the aware susceptible individuals will take protective measures to avoid infection. In practice, awareness may also influence other epidemiological processes, such as promoting earlier isolation, shortening the effective infectious period, or modifying contact patterns. Incorporating these effects would be an interesting direction for future research. η is the media conversion rate, which represents the rate at which unaware susceptible individuals convert to aware susceptible individuals. α is the awareness decay rate, which represents the rate at which aware individuals revert to the unaware susceptible individuals. φ is the disease mortality rate caused by infectious diseases. μ is the natural mortality rate. γ is the recovery rate of the disease.

2.2. Basic properties and equilibrium points of the model

After completing the model construction, conducting an in-depth analysis of the model's basic properties and equilibrium points is crucial for understanding the transmission dynamics of infectious diseases under the framework of this model. We start with the basic properties of the model and conduct a detailed analysis.

2.2.1. Basic properties

Theorem 2.1. Let $S_u(t), S_a(t), I(t), R(t)$ denote the solution of the model under the aforementioned initial conditions. If $t > 0$, we have $S_u(t) \geq 0, S_a(t) \geq 0, I(t) \geq 0, R(t) \geq 0$.

Proof. In the first equation of the system of differential equations, all terms involving S_a are moved to the left-hand side of the equality:

$$\frac{dS_u}{dt} + (\mu + \eta + \beta_1 I)S_u = \Lambda + \alpha S_a.$$

First, multiply both sides of the equation by $\xi = e^{\int_0^t \mu + \eta + \beta_1 I dt}$ to obtain:

$$\left(\frac{dS_u}{dt} + (\mu + \eta + \beta_1 I)S_u\right)\xi = (\Lambda + \alpha S_a)\xi,$$

the left-hand side of the equation can be expressed in the form of the derivative of a product:

$$\frac{d}{dt}(S_u \xi) = (\Lambda + \alpha S_a)\xi,$$

thus, we can solve for the expression of S_u :

$$S_u = \frac{S_u(0) + \int_0^t (\Lambda + \alpha S_a) e^{\int_0^t \mu + \eta + \beta_1 I dt} dt}{e^{\int_0^t \mu + \eta + \beta_1 I dt}}.$$

Since the numerator and denominator are positive, it follows that $S_u(t) \geq 0$ holds. Similarly, it can be proved that $S_a(t) \geq 0, I(t) \geq 0, R(t) \geq 0$. $\square$

Theorem 2.2. *The model is uniformly bounded in the domain under the initial conditions, and we define the bounded domain as:*

$$\Omega = \left\{ (S_u(t), S_a(t), I(t), R(t)) \mid 0 < S_u(t) + S_a(t) + I(t) + R(t) < \frac{\Lambda}{\mu} \right\}.$$

Proof.

$$\frac{dN}{dt} = \frac{dS_u + dS_a + dI + dR}{dt} = \Lambda - \varphi I - \mu N < \Lambda - \mu N,$$

separate variables and place N on the left-hand side, then integrate both sides of the inequality

$$\int_0^t \frac{dN}{\Lambda - \mu N} < \int_0^t dt,$$

and further, we have

$$-\frac{1}{\mu} \int_0^t \frac{d(\Lambda - \mu N)}{\Lambda - \mu N} < \int_0^t dt,$$

integrating both sides simultaneously, we obtain

$$\ln \left| \frac{\Lambda - \mu N(t)}{\Lambda - \mu N(0)} \right| > -\mu t,$$

and exponentiating both sides, we obtain

$$\frac{\Lambda - \mu N(t)}{\Lambda - \mu N(0)} > e^{-\mu t},$$

and solving for N yields

$$0 < N < \frac{\Lambda}{\mu} - \frac{1}{\mu}(\Lambda - \mu N(0))e^{-\mu t},$$

when $N(0) < \frac{\Lambda}{\mu}$, and $t \rightarrow \infty$, it holds that:

$$N < \frac{\Lambda}{\mu}.$$

This proves the boundedness of the differential equation. $\square$

2.2.2. Disease-free equilibrium point and basic reproduction number

Theorem 2.3. *The disease-free equilibrium of the system in Eq (2.1) is $E_0 = (\frac{\Lambda(\alpha+\mu)}{\mu(\alpha+\eta+\mu)}, \frac{\Lambda\eta}{\mu(\alpha+\eta+\mu)}, 0, 0)$ and the basic reproduction number is*

$$R_0 = \frac{\Lambda(\beta_1(\alpha + \mu) + \beta_2\eta)}{\mu(\alpha + \eta + \mu)(\varphi + \mu + \gamma)}.$$

Proof. Set the left-hand side of Eq (2.1) to 0, which yields the following system of equations:

$$\begin{cases} \Lambda - \beta_1 S_u I - \eta S_u + \alpha S_a - \mu S_u = 0, \\ \eta S_u - \alpha S_a - \beta_2 S_a I - \mu S_a = 0, \\ \beta_1 S_u I + \beta_2 S_a I - \varphi I - \gamma I - \mu I = 0, \\ \gamma I - \mu R = 0. \end{cases}$$

The disease-free equilibrium point $E_0 = (S_u^0, S_a^0, 0, 0)$, is obtained by solving this system of equations, where:

$$S_u^0 = \frac{\Lambda(\alpha + \mu)}{\mu(\alpha + \eta + \mu)}, S_a^0 = \frac{\Lambda\eta}{\mu(\alpha + \eta + \mu)}.$$

Next, the next-generation matrix method [26] is used to calculate the basic reproduction number. Let $X = (I, R, S_u, S_a)$, then the above model can be written in the following form:

$$\frac{dX}{dt} = F(X) - V(X),$$

where

$$F(X) = \begin{pmatrix} \beta_1 S_u I + \beta_2 S_a I \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$V(X) = \begin{pmatrix} (\varphi + \gamma + \mu)I \\ \mu R - \gamma I \\ -\Lambda + \beta_1 S_u I + \eta S_u - \alpha S_a + \mu S_u \\ -\eta S_u + \alpha S_a + \beta_2 S_a I + \mu S_a \end{pmatrix},$$

we compute the partial derivatives of the two matrices at the disease-free equilibrium point, and the following equations are obtained

$$DF(X) = \begin{pmatrix} \beta_1 S_u^0 + \beta_2 S_a^0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$DV(X) = \begin{pmatrix} \varphi + \mu + \gamma & 0 & 0 & 0 \\ -\gamma & \mu & 0 & 0 \\ \beta_1 S_u^0 & 0 & \mu + \eta & -\alpha \\ \beta_2 S_a^0 & 0 & -\eta + \mu & \alpha \end{pmatrix},$$

and the product of DF and DV^{-1} is

$$DF \cdot DV^{-1} = \begin{pmatrix} \frac{\beta_1 S_u^0 + \beta_2 S_a^0}{\varphi + \mu + \gamma} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where the basic reproduction number of the model is determined by the spectral radius, as the next-generation matrix is one-dimensional, its spectral radius is equivalent to the matrix value itself, i.e.,

$$\rho(DF \cdot DV^{-1}) = \frac{\beta_1 S_u^0 + \beta_2 S_a^0}{\varphi + \mu + \gamma}.$$

Finally, the basic reproduction number is obtained as:

$$R_0 = \frac{\Lambda(\beta_1(\alpha + \mu) + \beta_2\eta)}{\mu(\alpha + \eta + \mu)(\varphi + \mu + \gamma)},$$

thus, we obtain the expression for the basic reproduction number. $\square$

It can be seen from the expression of the basic reproduction number that among these parameters, the natural mortality rate μ , disease mortality rate φ , and population input rate Λ are hardly changed by media coverage. However, media coverage can enable the aware susceptible population to adopt better self-protection measures, thereby reducing the transmission rate coefficient β_2 of this group and achieving the goal of lowering the basic reproduction number.

2.2.3. Endemic equilibrium point

Theorem 2.4. When $R_0 > 1$, the system in Eq (2.1) has a unique endemic equilibrium.

Proof. Let the endemic equilibrium of the system be $E^* = (S_u^*, S_a^*, I^*, R^*)$. The calculation of the endemic equilibrium point is carried out as follows:

$$\begin{cases} \Lambda - \beta_1 S_u^* I^* - \eta S_u^* + \alpha S_a^* - \mu S_u^* = 0, \\ \eta S_u^* - \alpha S_a^* - \beta_2 S_a^* I^* - \mu S_a^* = 0, \\ \beta_1 S_u^* I^* + \beta_2 S_a^* I^* - \varphi I^* - \gamma I^* - \mu I^* = 0, \\ \gamma I^* - \mu R^* = 0. \end{cases} \quad (2.2)$$

From the third equation, we cancel I^* (since $I^* \neq 0$ at the endemic equilibrium point) from both sides, yielding:

$$\beta_1 S_u^* + \beta_2 S_a^* = \varphi + \gamma + \mu,$$

combining the first and second equations of Eq (2.2) we obtain the system:

$$\eta S_u^* = (\alpha + \mu + \beta_2 I^*) S_a^*, \quad (2.3)$$

$$\Lambda + \alpha S_a^* = \beta_1 S_u^* I^* + \mu S_u^* + \eta S_u^*, \quad (2.4)$$

and further, we have

$$S_a^* = \frac{\eta(\varphi + \mu + \gamma)}{\beta_1(\alpha + \mu + \beta_2 I^*) + \eta\beta_2}, \quad (2.5)$$

and

$$S_u^* = \frac{\varphi + \mu + \gamma}{\beta_1 + \frac{\beta_2 \eta}{\alpha + \mu + \beta_2 I^*}}, \quad (2.6)$$

at the endemic equilibrium point, using $N^* = \frac{\Lambda - \varphi I^*}{\mu}$, and the equation for the recovered population, this transforms into:

$$R^* = \frac{\gamma I^*}{\mu}, \quad (2.7)$$

and combining Eqs (2.5)–(2.7), we obtain:

$$AI^{*2} + BI^* + C = 0,$$

where

$$A = (1 + \frac{\gamma + \varphi}{\mu})\beta_2\beta_1 > 0,$$

$$B = (1 + \frac{\gamma + \varphi}{\mu})[\beta_1(\alpha + \mu) + \beta_2\eta] + \beta_2(\varphi + \mu + \gamma) - \frac{\Lambda\beta_1\beta_2}{\mu},$$

$$C = (1 - R_0)(\varphi + \mu + \gamma)(\alpha + \mu + \eta),$$

considering the case of $R_0 > 1$, we have

$$C = (1 - R_0)(\varphi + \mu + \gamma)(\alpha + \mu + \eta) < 0.$$

From the discriminant of the quadratic equation

$$\Delta = B^2 - 4AC > 0,$$

therefore, the equation has two roots; assume the two roots of the equation are I_1^* and I_2^* :

$$I_1^* I_2^* = \frac{C}{A} < 0.$$

It can be concluded that the equation has two roots with different signs (one positive and one negative). The negative root does not conform to the practical context (as population size cannot be negative) and is discarded. Thus, this quadratic equation in one variable has only one positive solution under $R_0 > 1$, which indicates that the endemic equilibrium point exists and is unique. $\square$

2.3. Stability of equilibrium points

The stability analysis of equilibrium points can reveal whether infectious diseases tend to be stably controlled or continue to spread and diffuse under different conditions. Below, we conduct an analysis of the stability of the disease-free equilibrium point and the endemic equilibrium point, respectively.

2.3.1. Stability analysis of the disease-free equilibrium point

Theorem 2.5. *When $R_0 < 1$, the disease-free equilibrium point is locally asymptotically stable.*

Proof. At the disease-free equilibrium point $E_0 = (S_u^0, S_a^0, 0, 0)$, the Jacobian matrix of the system after linearization is:

$$J(E_0) = \begin{pmatrix} -(\mu + \eta) & \alpha & -\beta_1 S_u^0 & 0 \\ \eta & -(\alpha + \mu) & -\beta_2 S_a^0 & 0 \\ 0 & 0 & \beta_1 S_u^0 + \beta_2 S_a^0 - (\varphi + \mu + \gamma) & 0 \\ 0 & 0 & \gamma & -\mu \end{pmatrix}.$$

After calculation, the characteristic equation of the Jacobian matrix is as follows:

$$(\lambda - (\varphi + \mu + \gamma)(R_0 - 1))(\lambda + \mu)(\lambda^2 + (2\mu + \eta + \alpha)\lambda + \mu^2 + \alpha\mu + \eta\mu) = 0.$$

We now focus on the equation:

$$\lambda^2 + (2\mu + \eta + \alpha)\lambda + \mu^2 + \alpha\mu + \eta\mu = (\lambda + \mu)(\lambda + \mu + \eta + \alpha) = 0,$$

therefore, the eigenvalues are as follows:

$$\lambda_1 = (\varphi + \mu + \gamma)(R_0 - 1), \lambda_{2,3} = -\mu, \lambda_4 = -(\mu + \eta + \alpha).$$

Hence, when $R_0 < 1$, all eigenvalues have negative real parts, and the disease-free equilibrium point E_0 is locally asymptotically stable at this time. $\square$

2.3.2. Stability analysis of the endemic equilibrium point

Theorem 2.6. *When $R_0 > 1$, the endemic equilibrium point is locally asymptotically stable.*

Proof. We consider the Jacobian matrix of the system at the endemic equilibrium point, and use the condition: $\beta_1 S_u^* + \beta_2 S_a^* - (\varphi + \mu + \gamma) = 0$

$$J(E^*) = \begin{pmatrix} -(\beta_1 I^* + \mu + \eta) & \alpha & -\beta_1 S_u^* & 0 \\ \eta & -(\alpha + \mu + \beta_2 I^*) & -\beta_2 S_a^* & 0 \\ \beta_1 I^* & \beta_2 I^* & 0 & 0 \\ 0 & 0 & \gamma & -\mu \end{pmatrix},$$

the characteristic equation corresponding to this matrix is:

$$(\lambda + \mu)(\lambda^3 + h_1\lambda^2 + h_2\lambda + h_3) = 0,$$

where:

$$\begin{aligned} h_1 &= \beta_1 I^* + \beta_2 I^* + 2\mu + \eta + \alpha, \\ h_2 &= \beta_1 \beta_2 I^{*2} + (\alpha + \mu) \beta_1 I^* + (\eta + \mu) \beta_2 I^* + \mu(\mu + \eta + \alpha) + \beta_1^2 S_u^* I^* + \beta_2^2 S_a^* I^*, \\ h_3 &= \beta_1^2 S_u^* I^* (\beta_2 I^* + \mu + \alpha) + \beta_2^2 S_a^* I^* (\beta_1 I^* + \mu + \eta) + \alpha \beta_1 \beta_2 S_a^* I^* + \eta \beta_1 \beta_2 S_u^* I^*. \end{aligned}$$

Next, we mainly focus on $\lambda^3 + h_1 \lambda^2 + h_2 \lambda + h_3$ whether the roots of this characteristic equation have negative real parts. According to the Routh–Hurwitz criterion [27, 28], we consider the following determinants Δ_1, Δ_2

$$\Delta_1 = h_1 > 0, \Delta_2 = \begin{vmatrix} h_1 & 1 \\ h_3 & h_2 \end{vmatrix},$$

if $\Delta_1 > 0$ and $\Delta_2 > 0$, it can be concluded from the Routh–Hurwitz criterion that all characteristic roots have negative real parts. Next, we focus on verifying whether $\Delta_2 > 0$

$$\Delta_2 = \begin{vmatrix} h_1 & 1 \\ h_3 & h_2 \end{vmatrix} = h_1 h_2 - h_3.$$

Simplify the expression step by step:

$$\begin{aligned} \Delta_2 &= h_1 h_2 - h_3, \\ &= \beta_1^2 \beta_2 I^{*3} + (\mu + \alpha) \beta_1^2 I^{*2} + (\mu + \eta) \beta_1 \beta_2 I^{*2} + \mu(\mu + \eta + \alpha) \beta_1 I^* \\ &\quad + \beta_1^3 S_u^* I^{*2} + \beta_1 \beta_2^2 I^{*3} + (\mu + \alpha) \beta_1 \beta_2 I^{*2} + (\mu + \eta) \beta_2^2 I^{*2} \\ &\quad + \mu(\mu + \eta + \alpha) \beta_2 I^* + \beta_2^3 S_a^* I^{*2} + (2\mu + \eta + \alpha) \beta_1 \beta_2 I^{*2} \\ &\quad + (2\mu + \eta + \alpha)(\alpha + \mu) \beta_1 I^* + (2\mu + \eta + \alpha)(\eta + \mu) \beta_2 I^* \\ &\quad + \mu(2\mu + \eta + \alpha)(\eta + \mu + \alpha) + \mu \beta_1^2 S_u^* I^* + \mu \beta_2^2 S_a^* I^* \\ &\quad + (\beta_1 - \beta_2)(\eta \beta_1 S_u^* - \alpha \beta_2 S_a^*) I^*. \end{aligned} \tag{2.8}$$

Most terms in the above expression are positive; we need to analyze the sign of the last terms:

$$(\beta_1 - \beta_2)(\eta \beta_1 S_u^* - \alpha \beta_2 S_a^*) I^*.$$

By simplifying the second equation of the system in Eq (2.3) at the endemic equilibrium, we obtain $\eta S_u^* = (\alpha + \mu + \beta_2 I^*) S_a^*$. Given $\mu + \beta_2 I^* > 0$, it directly follows that $\eta S_u^* > \alpha S_a^*$. Under the standing assumption $\beta_1 > \beta_2$, we first multiply both sides of $\eta S_u^* > \alpha S_a^*$ by β_1 to derive $\eta \beta_1 S_u^* > \beta_1 \alpha S_a^*$. Combined with $\beta_1 > \beta_2$, we further have $\beta_1 \alpha S_a^* > \alpha \beta_2 S_a^*$. The two inequalities together yield $\eta \beta_1 S_u^* > \beta_1 \alpha S_a^* > \alpha \beta_2 S_a^*$.

Therefore, we obtain $\Delta_2 > 0$, which satisfies the Routh–Hurwitz stability criterion. Thus, the equilibrium E^* is locally asymptotically stable. $\square$

2.4. Bifurcation analysis

Theorem 2.7. When $R_0 = 1$, the system in Eq (2.1) exhibits forward bifurcation.

Proof. To analyze the bifurcation behavior of the model, we focus on its dynamical characteristics at $R_0 = 1$. We choose the transmission rate coefficient β_2 of the aware susceptible individuals as the bifurcation parameter. The critical value determined by $R_0 = 1$ is calculated as follows:

$$\beta_2^* = \frac{\mu(\alpha + \eta + \mu)(\varphi + \mu + \gamma) - \Lambda\beta_1(\alpha + \mu)}{\Lambda\eta}.$$

Now, the method in [29, 30] is used to study the model's dynamical behavior at $R_0 = 1$. Let:

$$S_u(t) = x_1, S_a(t) = x_2, I(t) = x_3, R(t) = x_4,$$

and

$$\frac{dS_u}{dt} = f_1, \frac{dS_a}{dt} = f_2, \frac{dI}{dt} = f_3, \frac{dR}{dt} = f_4.$$

In Theorem 2.5, we obtain the Jacobian matrix of the disease-free equilibrium point, Based on this matrix, we calculate the right eigenvector and left eigenvector of the model at (E_0, β_2^*) :

$$J(E_0)\omega = 0, J^T(E_0)v = 0,$$

and we can obtain the right eigenvector $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)^T$ and left eigenvector $v = (v_1, v_2, v_3, v_4)$, where

$$\begin{aligned}\omega_1 &= -\frac{\Lambda\eta[\alpha\eta\beta_2 + (\alpha\mu + \mu^2 + \alpha\eta + \mu\eta)\beta_1(\alpha + \mu)]}{\mu^2(\alpha + \eta + \mu)^2} < 0, \\ \omega_2 &= -\frac{\Lambda\eta(\mu + \eta)(\beta_2 + \beta_1(\alpha + \mu))}{\mu^2(\alpha + \eta + \mu)^2} < 0, \\ \omega_3 &= 1, \omega_4 = \frac{\gamma}{\mu}, v_1 = v_2 = v_4 = 0, v_3 = 1.\end{aligned}$$

When calculating the right eigenvector, we may set $\omega_3 = 1$, thereby obtaining a relatively simplified expression. When calculating the left eigenvector v_3 , we determine the value of the corresponding component ω_4 of the left eigenvector at this time according to the normalization condition:

$$v^T\omega = 1,$$

and we can have $v_3 = 1$. Next, calculate the partial derivatives of f_k ($k = 1, 2, 3, 4$) with respect to variables S_u, S_a, I, R , and the bifurcation parameter β_2 at the specific condition (E_0, β_2^*) , for brevity, only non-zero second-order partial derivatives are listed, and all vanishing terms are omitted

$$\begin{aligned}\frac{\partial^2 f_1}{\partial S_u \partial I} &= -\beta_1, \frac{\partial^2 f_2}{\partial S_a \partial I} = -\beta_2, \frac{\partial^2 f_3}{\partial S_u \partial I} = \beta_1, \\ \frac{\partial^2 f_3}{\partial S_a \partial I} &= \beta_2, \frac{\partial^2 f_2}{\partial I \partial \beta_2} = -S_a^0, \frac{\partial^2 f_3}{\partial I \partial \beta_2} = S_a^0.\end{aligned}$$

Now, we discuss the model's dynamical behavior. The following formula is obtained from [29, 30]:

$$a = \sum_{k,i,j=1}^4 v_k \omega_i \omega_j \frac{\partial^2 f_k}{\partial x_i \partial x_j}, b = \sum_{k,i=1}^4 v_k \omega_i \frac{\partial^2 f_k}{\partial x_i \partial \beta_2},$$

thus, we have:

$$a = \nu_3(2\omega_1\omega_3\beta_1 + 2\omega_2\omega_3\beta_2^*) = 2(\omega_1\beta_1 + \omega_2\beta_2^*) < 0,$$

$$b = \nu_3\omega_3S_a^0 = S_a^0 > 0.$$

Since the specific condition $a < 0, b > 0$, holds at this time, the model exhibits a forward bifurcation. $\square$

3. Numerical simulations and result analysis

To verify the theoretical conclusions of the Section 2 infectious disease model, in this section we use Python and PINN for numerical simulations, focusing on the model's dynamic behavior under different conditions. Core work covers two aspects: (1) Verifying the model's basic properties via PINN to ensure theoretical accuracy; (2) simulating media coverage's impact on infectious disease spread and quantifying its role in lowering infection peaks and delaying outbreaks.

3.1. Verification of model properties

To verify the basic properties of the constructed infectious disease model, we build a standard 3-Layer fully connected neural network with the structure "Affine transformation $\rightarrow$ ReLU activation $\rightarrow$ Affine transformation $\rightarrow$ ReLU activation $\rightarrow$ Affine transformation" for fitting the system's dynamic process [11]. The input layer takes time as the sole feature to capture temporal dynamics of disease transmission; the network has two hidden layers (128 neurons each) using ReLU activation, which introduces nonlinearity (adapting to complex transmission dynamics) and is valued for simplicity and training stability in deep learning [31]. The output layer corresponds to the four population groups in the model: unaware susceptible, aware susceptible, infected, and recovered.

For the iterative computation in network training, we solve the temporal evolution of the ODE system via both the explicit forward Euler method and the fourth-order Runge–Kutta (RK4) method as numerical discretization schemes, and generate synthetic time-series data for all compartments (S_u, S_a, I, R). These datasets are derived entirely from the theoretical model equations rather than real-world epidemic observations, so the experiments in this section serve as self-consistency tests to verify the PINN's ability to reproduce the model's intrinsic dynamical behaviors, rather than fully data-driven simulations with empirical data. The numerical solutions generated by the corresponding solver are used throughout this section as synthetic reference solutions for model training and accuracy evaluation [32]. In this study, RK4-generated trajectories are used as high-accuracy synthetic reference solutions for controlled methodological benchmarking. Therefore, the comparisons between PINN predictions and numerical reference trajectories should be interpreted as evaluations of approximation accuracy under a synthetic dynamical system, rather than as validation against real-world epidemiological observations. It should be clarified that, throughout this section, the terms "true solution", "true trajectory", and "true value" refer to the synthetic numerical reference solutions generated by the corresponding classical numerical solver, namely the Euler method or the RK4 method, rather than to real-world epidemiological observations. Therefore, the comparison between PINN predictions and the "true" trajectories should be interpreted as a controlled numerical benchmark for evaluating approximation accuracy under the prescribed model equations.

In the training phase, to ensure efficiency and stability of convergence, the AdamW optimizer is selected for minimizing the network's loss function; AdamW decouples weight decay from the gradient

update and improves regularization compared to standard Adam [33]. Moreover, a cosine annealing learning rate scheduling strategy (SGDR) is adopted to dynamically adjust the learning rate, which alleviates the risk of being trapped in local minima and improves the accuracy of fitting the model's dynamical characteristics [34]. In the PINN proposed in this paper, the design of the loss function integrates data fitting and physical law constraints [11]. The formula for calculating the loss function is as follows:

$$L = \delta L_{data} + \beta L_{physics},$$

among them, L_{data} denotes the data-fitting loss, which is defined as the mean squared error (MSE) [35] between the corresponding numerical reference solutions and the PINN predictions. Specifically, in the Euler-based and RK4-based experiments, the reference trajectories are generated by the Euler method and the RK4 method, respectively, so that each PINN model is evaluated under its corresponding numerical discretization setting. These numerically generated trajectories are used as high-accuracy synthetic reference solutions for methodological benchmarking under controlled dynamical conditions, rather than as epidemiological ground truth obtained from real outbreak observations. $L_{physics}$ is computed from the residuals of the governing differential equations, encouraging the predicted solutions to remain consistent with the epidemic dynamics. The corresponding formulation is given by

$$L_{data} = \frac{1}{N} \sum_{i=1}^N |y_{pred}(x_i) - y_{true}(x_i)|^2,$$

and

$$L_{physics} = \frac{1}{N} \sum_{i=1}^N \frac{1}{4} \sum_{j=1}^4 \left| \frac{dy_j}{dx_i} - f_j(x_i) \right|,$$

where f_j denotes the differential equation for the j -th population. δ and β are hyperparameters, i.e., predefined parameters not learned during training—that serve as weight ratios to balance the two error components (data-fitting loss and physics-informed residual loss). Through systematic ablation experiments, we determine the optimal weight combination as $\delta = 0.6$ and $\beta = 0.4$ for this study.

The neural network approximates the temporal trajectories of the four compartments. In addition to the data-fitting loss against synthetic reference solutions, the PINN minimizes physics residuals derived from the governing equations, thereby encouraging the predicted trajectories to remain consistent with the epidemic dynamics.

Next, based on the theory in Section 2, we present the phase diagrams of the disease-free equilibrium point and the endemic equilibrium point, as well as the forward bifurcation diagram.

First, we verify the theoretical conclusions obtained in Section 2. As shown in Figure 2, this phase diagram uses the unaware susceptible population S_u , aware susceptible population S_a , and infected population I as the coordinate axes, and shows the phase space trajectories of the system when the basic reproduction number $R_0 < 1$ (where $R_0 = 0.57$).

It can be observed that under different initial conditions, all trajectories converge to the disease-free equilibrium point, and the number of infected people eventually decreases to 0. This not only confirms the theory that the disease will die out when the basic reproduction number $R_0 < 1$, but also verifies that the proof of the disease-free equilibrium point in the previous section is correct.

As shown in Figure 3, this phase diagram presents the phase space trajectories of the system when $R_0 > 1$ (with $R_0 = 28.53$). All trajectories converge to the endemic equilibrium. By comparing the two phase diagrams and analyzing the system behavior around the threshold of the basic reproduction number, we depict the “extinction” and “prevalence” phenomena in infectious disease dynamics. This comparison provides an intuitive dynamic framework for understanding the model. Next, we demonstrate the forward bifurcation whose existence has been proven previously.

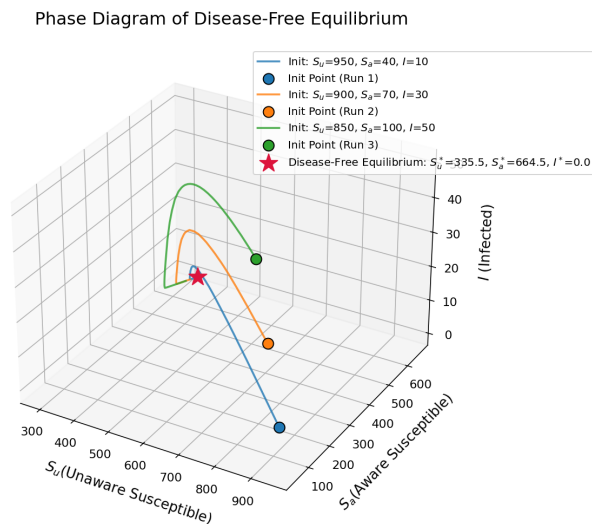


Figure 2. Phase diagram of the disease-free equilibrium.

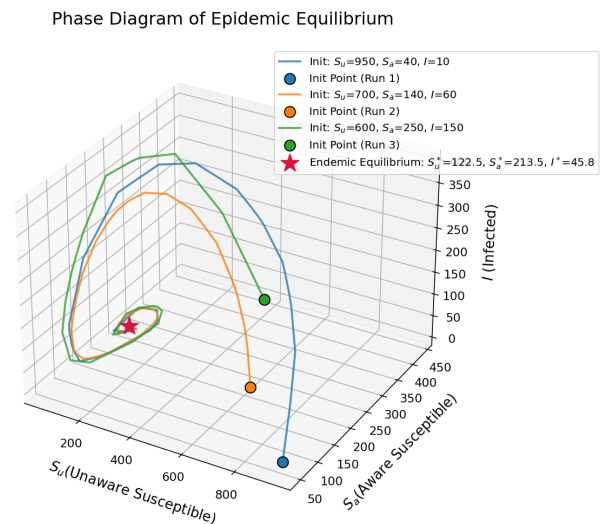


Figure 3. Phase diagram of the endemic equilibrium.

As shown in Figure 4, this diagram presents the forward bifurcation of the model established in this study, with the bifurcation parameter being β_2 . The blue line represents the stable disease-free equilibrium point, the red line represents the unstable disease-free equilibrium point, and the green line represents the stable endemic equilibrium point. It can be observed that the stable disease-free equilibrium point becomes unstable when the basic reproduction number $R_0 = 1$ and a bifurcation of the stable endemic equilibrium point occurs, confirming the conclusions derived earlier in this study.

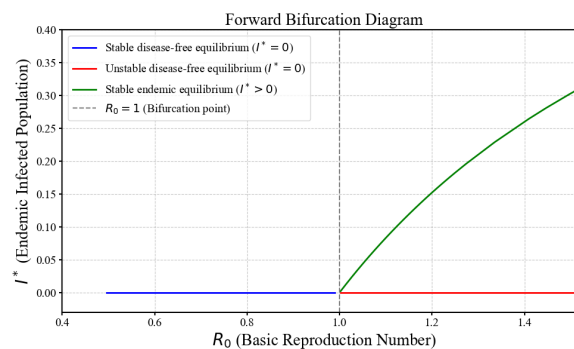


Figure 4. Forward bifurcation diagram.

3.2. Simulation of the media effect

We derive the ranges of some parameters based on references: In [24], focusing on the media effect, the intensity of the media coverage effect was analyzed using a set of real infectious disease data. Through the interpolation method, the media conversion rate η was 0.398 per day, and the awareness decay rate α was 0.1 per day. According to [36], the recovery rate γ was 0.11 per day. Based on the setting in the article by Feng et al. [37], the average human lifespan is approximately 70 years. Converted to the parameter value required in this study (i.e., the average daily natural mortality rate μ), this equals $1/(70 \times 365)$, which is approximately 3.9×10^{-5} per day. Finally, using the relevant data in [38], the disease mortality rate φ is set to approximately 0.015 per day.

Subsequently, using the population steady-state formula $N = \Lambda/\mu$, we calculated the population input rate as 0.39 per day. Since this study is not calibrated to a specific infectious disease, and the transmission rate coefficients of different diseases vary significantly, for the convenience of calculation and based on the assumed relationship $\beta_1 > \beta_2$, we make assumptions about the transmission rate coefficients of the two susceptible individuals, as follows: $\beta_1 = 8 \times 10^{-5}$, $\beta_2 = 2 \times 10^{-5}$.

The parameter values used in this study are listed in Table 1.

Table 1. Parameter values.

Symbol	Parameter	Values	Source
Λ	Population input rate	0.39	Calculated
β_1	Unaware transmission rate coefficient	8×10^{-5}	Assumed
β_2	Aware transmission rate coefficient	2×10^{-5}	Assumed
η	Media conversion rate	0.398	[24]
γ	Recovery rate	0.11	[36]
μ	Natural mortality rate	3.9×10^{-5}	[37]
φ	Disease mortality rate	0.015	[38]
α	Awareness decay rate	0.1	[24]

It should be emphasized that the parameter values adopted in this study are representative demonstrative settings selected to illustrate the model's dynamical properties and to evaluate different PINN architectures under a controlled computational environment. Since we focus on dynamical analysis and network architecture comparison, fixed parameter values are used to keep the experimental conditions identical across all PINN structures. This setting enables a fair comparison of numerical discretization schemes, network depth, and skip-connection architectures. The theoretical findings and algorithmic conclusions provide methodological insight for compartmental epidemic systems with similar awareness-based stratification, and systematic parameter calibration and empirical validation with real-world disease datasets will be pursued in future work.

Next, we will target several scenarios covered in Section 2, mainly using the aforementioned PINN to conduct parameter simulations, and present numerical results of relevant conclusions to more intuitively verify the results we obtained earlier.

As shown in Figure 5, for the scenario without media coverage, we set the media-related parameters η , α to 0. This means merging the susceptible populations—specifically, the aware susceptible population no longer exists. In the no-media scenario (red line), three key issues emerge: The peak proportion of

infected individuals is significantly higher, the rising-phase slope is steeper, and the peak arrives earlier. These stem from limited public access to epidemic information-leading to weak protective awareness and no proactive preventive actions, which leave disease transmission unchecked. By contrast, the media scenario (blue line) shows marked improvements: a much lower infection peak, a gentler rising trend, and a delayed peak. This is because media enables the public to get timely epidemic updates and take proactive protection, curbing rapid disease spread.

In summary, media coverage inhibits transmission by boosting public protective awareness and guiding behavior changes. It lowers the infection peak, slows growth, delays the peak, and accelerates the post-peak decline-ultimately reducing the overall impact of infectious diseases on the population.

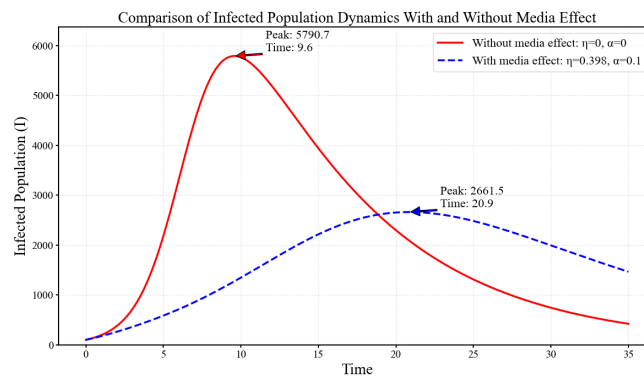


Figure 5. Comparison of infected population dynamics with and without media effects.

Figure 6 shows the change of the basic reproduction number R_0 with the media conversion rate η . As the media conversion rate η increases, the basic reproduction number R_0 first decreases and then tends to stabilize. In the theory of infectious disease transmission, R_0 represents the average number of people an infected individual can transmit the disease to in a fully susceptible population. The threshold line of $R_0 = 1$ in the figure enables an intuitive judgment of the infectious disease transmission trend: When η is small, $R_0 > 1$, meaning there is a risk of infectious disease transmission; as η increases, R_0 gradually decreases; when $R_0 < 1$, the transmission trend of the infectious disease is inhibited.

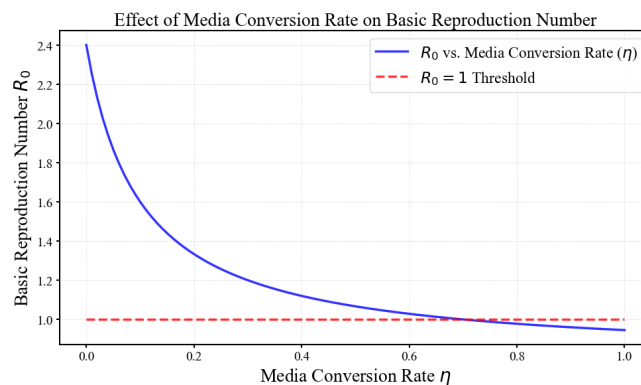


Figure 6. Effect of media conversion rate on basic reproduction number.

This indicates that increasing the media conversion rate can effectively control the spread of infectious

diseases, providing a theoretical basis for infectious disease prevention and control.

As can be seen from Figure 7, R_0 and β_2 exhibit a positive correlation. Here, β_2 represents the transmission rate coefficient of aware susceptible individuals: The larger the value of β_2 , the more likely this group is to be infected after coming into contact with infected individuals. This further increases the possibility of disease transmission, consequently raising R_0 and leading to a higher risk of infectious disease spread. Conversely, reducing β_2 can effectively decrease R_0 , thereby achieving the effect of inhibiting the spread of infectious diseases.

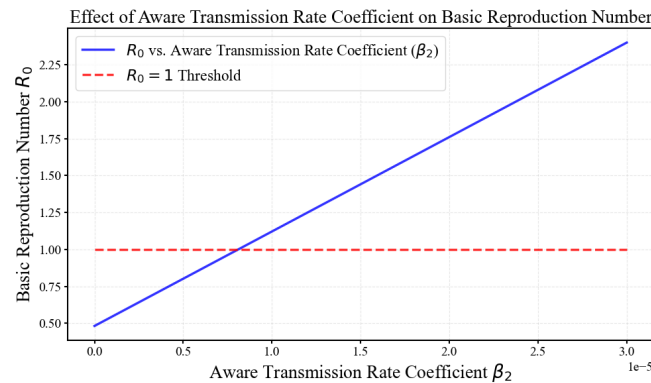


Figure 7. Effect of aware transmission rate coefficient on basic reproduction number.

4. Optimization of the neural network

In this section, we focus on optimizing the neural network simulation presented earlier, as well as providing reflections and comparisons. Specifically, we introduce the RK4 method, increase the number of network layers, and adopt Residual Network and Highway Network. Additionally, we conduct statistical analysis on the final results to validate the optimization effects.

4.1. Algorithm updates and network deepening

We replace the first-order Euler method in residual calculation with the RK4 method [32], which has a global truncation error of $O(h^4)$ (vs. $O(h)$ for Euler), to investigate its impact on model performance [39]. All experimental parameters are kept identical for a fair comparison.

As shown in Figure 8, the two methods show nearly identical loss curves in early training (0–1500 epochs). After 1500 epochs, RK4 achieves slightly faster residual convergence. Both models produce nearly identical predicted $I(t)$ curves, with no perceptible difference in prediction accuracy.

In conclusion, RK4 provides measurable but very limited improvement for this low-dimensional epidemic model. The first-order Euler method delivers sufficient numerical accuracy for the PINN residual calculation here.

Generally speaking, deeper neural networks have stronger representational capacity [40, 41]. To enhance the network's representational capacity, we expand the depth of the fully connected network (FCN) from 3-Layers to 5-Layers, aiming to better fit the complex nonlinear relationships in the system. However, it is important to note that increasing depth does not always yield better results: A deeper network may lead to training difficulties (e.g., gradient vanishing/exploding) and

overfitting risks [42].

As shown in Figure 9, the 5-Layer network outperforms the 3-Layer network in the initial training stage and the final convergence stage, exhibiting stronger nonlinear fitting capability. It demonstrates better performance throughout the training process. Moreover, we can intuitively observe that the 5-Layer neural network shows a significant improvement over the 3-Layer network. It learns faster and achieves lower loss.

On this basis, we further test fully connected structures with more layers. However, experimental results show that such deeper networks not only fail to further reduce the loss during training but also exhibit certain instability; their final performance is similar to that of the 5-Layer network or slightly decreases. Analysis suggests that as network depth increases, the model may encounter issues such as gradient vanishing or optimization difficulties, preventing continuous improvement in performance [42]. Thus, in subsequent simulations of this paper, the 5-Layer network structure is adopted as the main model architecture and will be used for comparison with other network architectures.

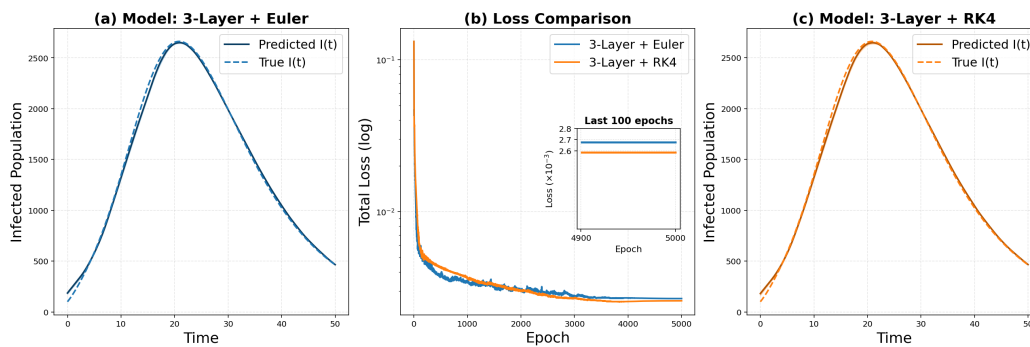


Figure 8. Training loss and prediction accuracy: RK4 vs. Euler discretization scheme. (a) Predicted $I(t)$ vs. true $I(t)$: 3-Layer + Euler; (b) Loss curve comparison: 3-Layer + Euler vs. 3-Layer + RK4; (c) Predicted $I(t)$ vs. true $I(t)$: 3-Layer + RK4.

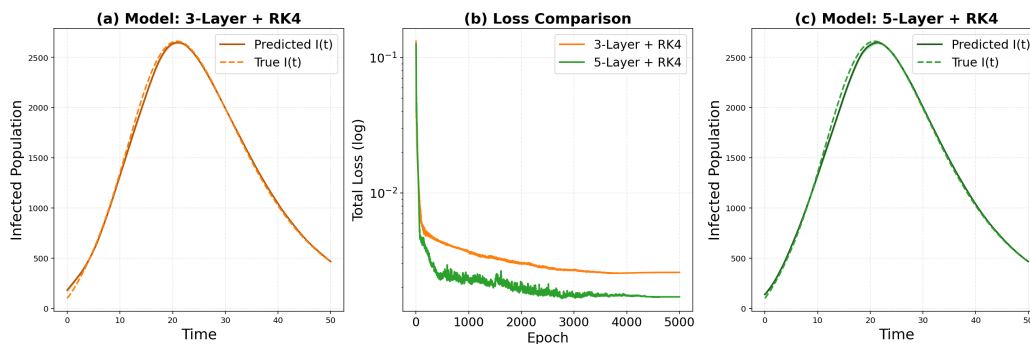


Figure 9. Impact of network depth on training performance: 3-Layer vs. 5-Layer Fully Connected Network. (a) Predicted $I(t)$ vs. true $I(t)$: 3-Layer + RK4; (b) loss curve comparison: 3-Layer + RK4 vs. 5-Layer + RK4; (c) predicted $I(t)$ vs. true $I(t)$: 5-Layer + RK4.

4.2. Residual network

To further enhance the representational capability and training stability of PINN in solving differential equations, we introduce two structural optimization schemes based on fully connected networks: Residual Network and Highway Network. Both alleviate the gradient vanishing problem in training deep network through a skip connection mechanism, thereby improving the model's fitting accuracy and training efficiency [19].

First, we explain the residual network involved in the network: The core idea of residual network is that instead of directly learning the mapping relationship between input and output, the neural network learns the “residual”-i.e., the slight change in output relative to input. Each residual block includes a linear transformation, a Layer Normalization, and an activation function (ReLU), with a skip connection introduced. The output of the module is expressed as: $y = x + f(x)$, where x denotes the input value, $f(x)$ represents the residual term after several nonlinear transformations. x is added to the output $f(x) = out$ after a series of transformations to obtain the overall output y . In fact, what this network learns is the residual between the target function and the input $f(x) = y - x$. During backpropagation, the equation for calculating the gradient of the loss function becomes:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial x} = \frac{\partial L}{\partial y} \left(1 + \frac{\partial f}{\partial x}\right),$$

even if the increment is small each time, the presence of 1 alleviates the gradient vanishing problem in neural network learning, enhances the model's representational capability, and thereby improves the model's training [19]. This core idea is implemented in the residual network, as shown in Figure 10.

As shown in Figure 11, with the vertical axis (loss value) on a logarithmic scale and the horizontal axis representing training epochs, this figure compares the loss evolution of the fully connected network and the residual network. The red curve denotes the residual network, and the green curve denotes the fully connected network. An analysis of the curve dynamics is as follows.

In the initial stage of training, the initial loss value of the fully connected network is slightly lower than that of the residual network, and it decreases faster in the 0 to approximately 800-epoch interval. Thus, it remains in a leading position around the 800 epochs mark. However, after 800 epochs, the loss reduction of the fully connected network slows down significantly, and it is even accompanied by slight fluctuations before gradually leveling off. In contrast, the loss of the residual network continues to decrease rapidly during this stage, with the magnitude and duration of the decrease being significantly better than those of the fully connected network. When training progresses to around 4000 epochs, the losses of both networks gradually enter the convergence stage, with fluctuations reducing sharply.

From the perspective of the final convergence result, the residual network can reduce the loss to a level far lower than that of the fully connected network. It can be seen that after modifying the 5-Layer fully connected network into a residual network, the model not only achieves more efficient loss reduction in the later stage of training but also reaches a lower loss value at final convergence. This improvement significantly enhances the model's fitting accuracy and makes the prediction results more accurate.

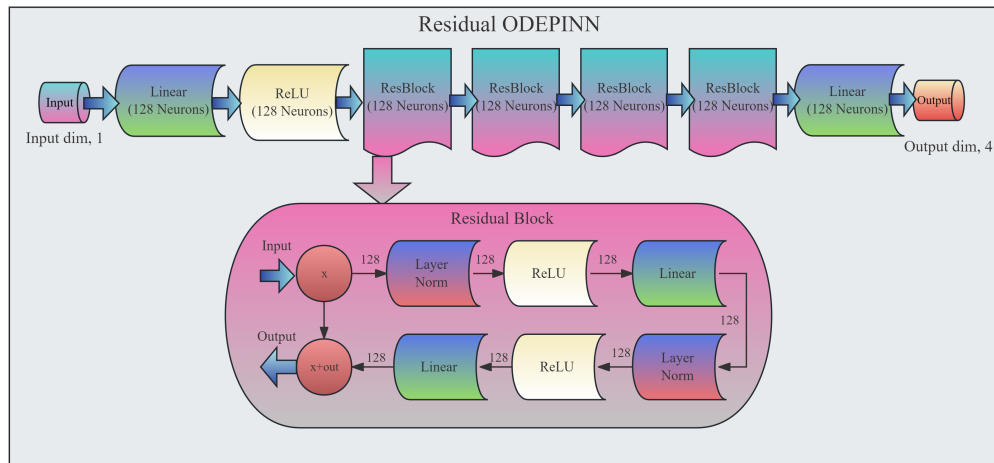


Figure 10. Residual network.

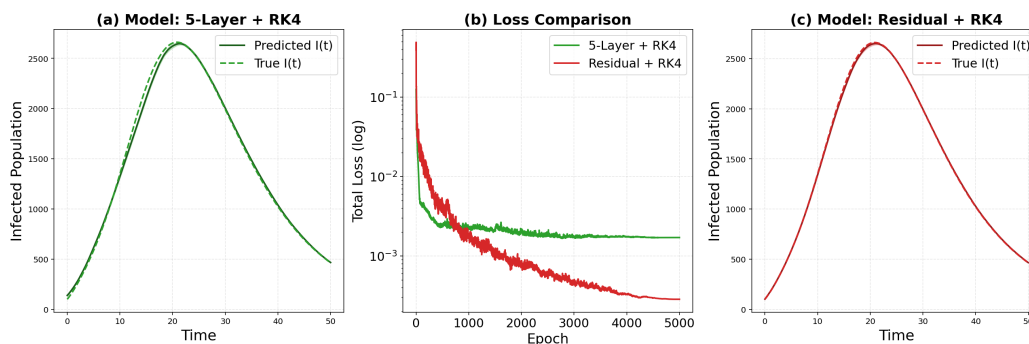


Figure 11. Training loss comparison: Residual network vs. 5-Layer Fully Connected Network. (a) Predicted $I(t)$ vs. true $I(t)$: 5-Layer + RK4; (b) loss curve comparison: 5-Layer + RK4 vs. Residual + RK4; (c) predicted $I(t)$ vs. true $I(t)$: Residual + RK4.

4.3. Highway network

In the aforementioned comparison involving Residual Network, fixed skip connections have demonstrated performance advantages over Fully Connected Network. Next, we conduct simulations using the Highway Network.

The Highway Network is a type of neural network proposed by Srivastava et al. in 2015 [20], whose core design objective is to address the vanishing gradient problem in the training of deep neural network. By introducing gating mechanisms and skip connections, it enables flexible information transfer between network layers, overcoming the limitation of “layer-by-layer dependency” in information transfer of traditional Fully Connected Network.

The core unit of a standard Highway Network is the Highway block, which mostly consists of a main transformation branch (H) and a gating branch (T). The main transformation branch is responsible for performing nonlinear transformations on input features and extracting deep-seated features; while the

gating branch controls the weight of information in the main branch through learnable gating coefficients and simultaneously adjusts the proportion of direct transfer of input information. In accordance with the network formulas designed in [20], the application in this study is as follows:

$$y = (1 - T(x))x + T(x)H(x).$$

The Highway network used in this study is illustrated in Figure 12, which depicts its internal working mechanism. Within the Highway block, input features x simultaneously enter the main transformation branch and the gating branch. The output of the main branch and the original input are ultimately fused through a gating-weighted method to achieve information transfer. The overall network consists of one input mapping layer, four Highway blocks, and one output mapping layer, which will be used for training in the following sections.

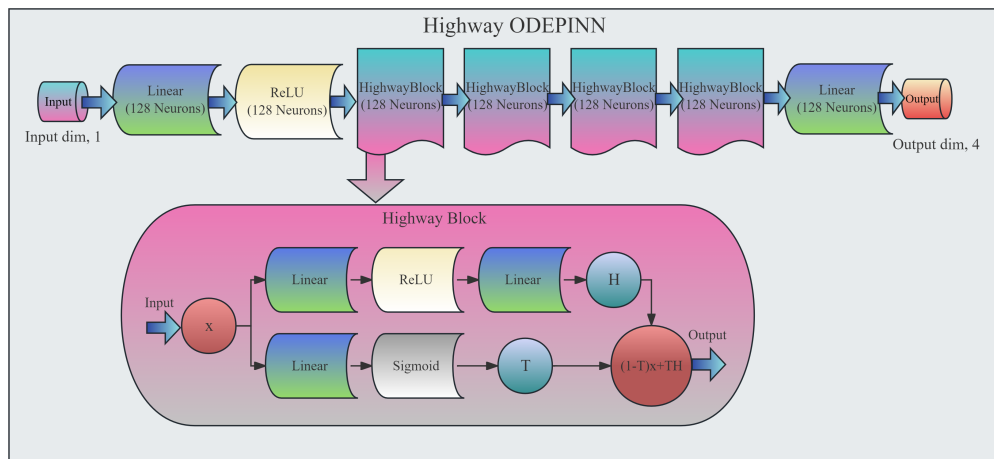


Figure 12. Highway network.

To intuitively compare the performance differences between the Highway Network and the Fully Connected Network during the training process, we plot a comparison chart of their loss function value changes (as shown in Figure 13).

Figure 13 compares the training performance of the 5-Layer fully connected network (FCN) and the Highway network (both with RK4). The green curve represents the 5-Layer FCN, and the purple curve represents the Highway network. The Highway network shows overall better performance than the FCN. It converges faster in early training. The Highway network maintains a steady loss decline and reaches a much lower final stable loss with smaller fluctuations, showing stronger convergence and stability. In summary, the Highway network alleviates training instability via its gating mechanism [20], and outperforms the 5-Layer FCN in convergence speed, residual error, and stability for epidemic dynamics modeling.

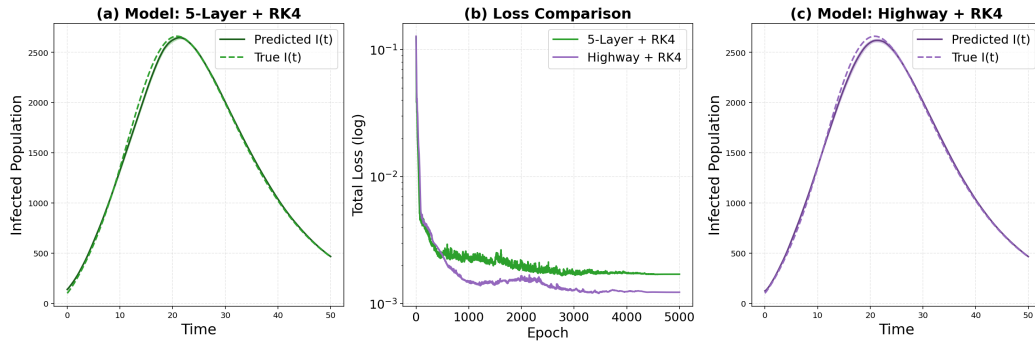


Figure 13. Training loss comparison: Highway network vs. 5-Layer Fully Connected Network. (a) Predicted $I(t)$ vs. true $I(t)$: 5-Layer + RK4; (b) loss curve comparison: 5-Layer + RK4 vs. Highway + RK4; (c) predicted $I(t)$ vs. true $I(t)$: Highway + RK4.

Building on the superior performance of the Highway network, we further integrate residual connections to enhance training stability and gradient propagation. Combining the adaptive gating of Highway network with the direct identity mapping of residual connections, we propose a residual-integrated Highway network. As shown in Figure 14, this adopts a three-parallel branch architecture. Input features are fed simultaneously into the main transformation branch (H), gating branch (T) (following the classic Highway design [20]), and an additional independent residual branch (Res) to stabilize feature learning and provide output corrections [43].

It should be noted that the residual structure proposed in this study is similar to that in [43], as both integrate residual ideas into the Highway network. However, there is a key difference: The researchers in [44] utilize the formula $y = \text{Highway}(x) + x$, treating the entire Highway network as a nonlinear transformation and adding skip connections externally. In contrast, we introduce an additional residual branch $R(x)$, inside the Highway block, integrating the residual mechanism with the gating mechanism within the same module, thereby further enhancing network stability.

On this basis, we construct an exploratory residual-integrated Highway variant by introducing an additional independent residual branch $R(x)$ inside the Highway block.

$$y = (1 - T(x))x + T(x)H(x) + R(x) = x + R(x) + T(x)(H(x) - x).$$

Here $H(x)$ represents the output similar to a feedforward neural network, which is the main prediction result; while $R(x)$ is residual, and $T(x)$ is the Sigmoid-gated coefficient (range [0,1]) controlling layer output change intensity, when $T(x) = 0$ only the residual and input remain, which is consistent with the results in the previous section; when $T(x) = 1$, the formula becomes $y = H(x) + R(x)$, which is different from the residual network in the previous section that learns $y - x$, instead, it learns $y - H(x)$, i.e., the difference between the prediction of the main path and the true value. At this time, $R(x)$ acts as an auxiliary correction. Subsequently, by adjusting the value of the gating coefficient $T(x)$ within [0, 1], we can control how much old information to retain and how many new changes to introduce in the ResHighway network. Based on the content of the three pathways mentioned above, Figure 14 shows the ResHighway network used in this study.

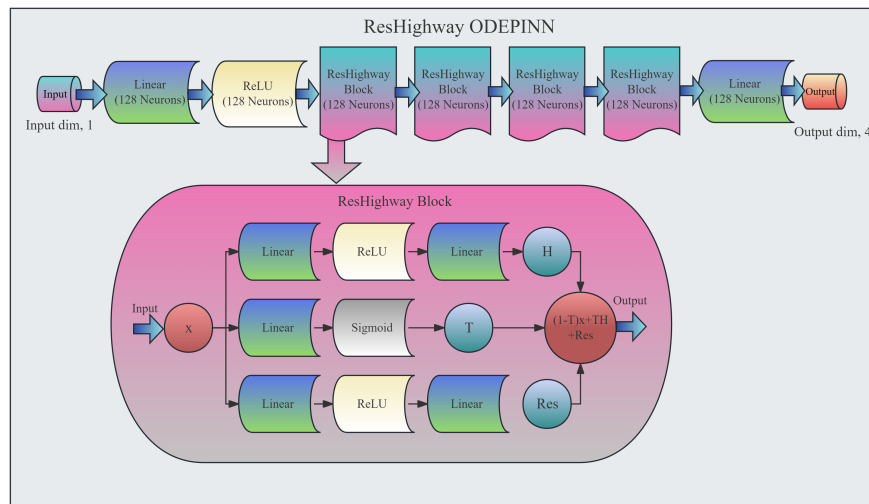


Figure 14. ResHighway network.

Figure 15 compares the training loss between the 5-Layer fully connected network and the proposed ResHighway network. In the early training stage, both networks show rapid loss reduction, but the ResHighway network converges faster, with the advantage of residual integration already evident. At convergence, the ResHighway network achieves a significantly lower final loss, showing improved performance compared with the 5-Layer fully connected baseline.

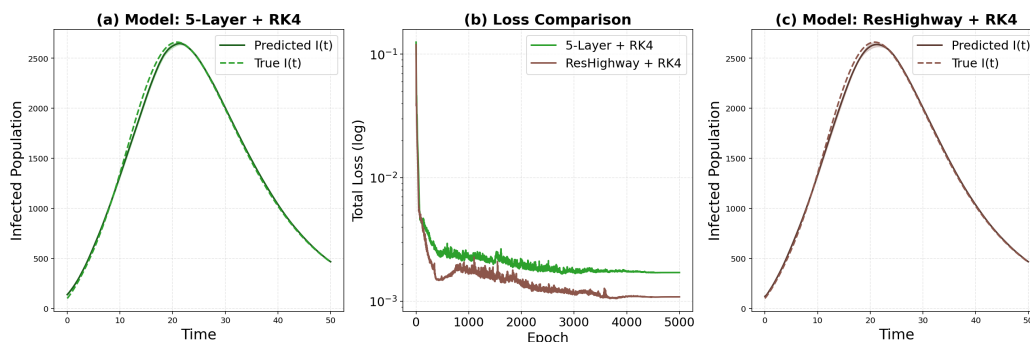


Figure 15. Training loss comparison: ResHighway Network vs. 5-Layer Fully Connected Network. (a) Predicted $I(t)$ vs. true $I(t)$: 5-Layer + RK4; (b) loss curve comparison: 5-Layer + RK4 vs. ResHighway + RK4; (c) predicted $I(t)$ vs. true $I(t)$: ResHighway + RK4.

Figure 16 compares the training performance of the standard Highway network and our proposed ResHighway network. The ResHighway network converges faster throughout training and achieves a lower final stable loss, demonstrating better convergence and training stability. This result verifies that introducing the independent residual branch effectively enhances the network's performance, indicating that the additional residual branch can improve the standard Highway network in this controlled benchmark, although it still does not outperform the simpler Residual network in the overall comparison.

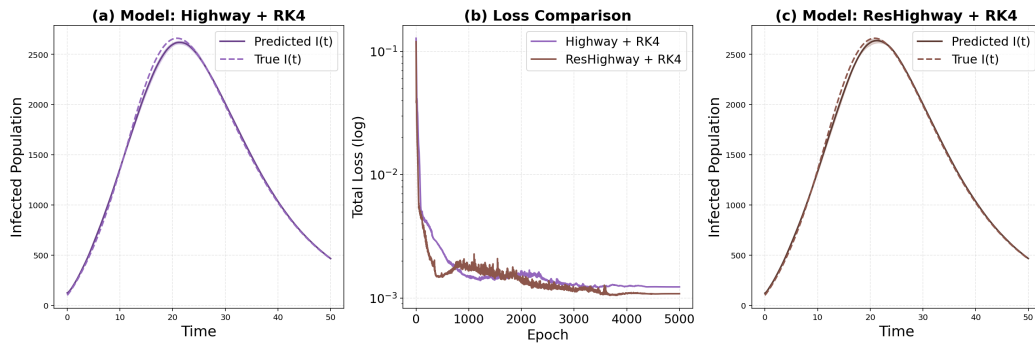


Figure 16. Training loss comparison: ResHighway Network vs. Standard Highway Network. (a) Predicted $I(t)$ vs. true $I(t)$: Highway + RK4; (b) loss curve comparison: Highway + RK4 vs. ResHighway + RK4; (c) predicted $I(t)$ vs. true $I(t)$: ResHighway + RK4.

4.4. Network comparison and analysis

In this section, we focus on comparing and discussing the training performance, convergence characteristics, and final fitting accuracy of the six PINN architectures designed in this study (3-Layer + Euler, 3-Layer + RK4, 5-Layer + RK4, Residual + RK4, Highway + RK4, ResHighway + RK4) for infectious disease dynamics tasks.

The comparison is based on the average loss curves and statistical results obtained from ten independent training runs for each model. The ten runs differ only in the random seed used for network initialization and the stochastic components of training (seeds 42–51); the architecture, optimizer, learning-rate schedule, training data, and number of epochs are identical across runs. Therefore, the reported variability and the corresponding confidence intervals mainly reflect each architecture's sensitivity to random initialization and its training stability across independent runs, rather than resampling of a single trained model. For the RMSE and physics-residual metrics reported in Figures 19 and 20 and Table 2, we present the sample mean over the ten runs together with its 95% confidence interval, expressed as $\bar{x} \pm \text{CI}$, where the half-width CI is derived from the t distribution:

$$\text{CI} = t_{0.975,9} \frac{s}{\sqrt{n}},$$

where s is the sample standard deviation, $n = 10$, the degrees of freedom are $n - 1 = 9$, and $t_{0.975,9} = 2.262$. Key indicators are presented with intuitive visualizations [45].

The six curves in Figure 17 show the average total loss over ten independent training runs, illustrating the convergence behavior of different PINN architectures. As shown in the figure, the Residual Network achieves the lowest final loss and maintains the most stable convergence throughout the training process, demonstrating its superior optimization capability. The ResHighway Network exhibits rapid loss reduction during the early training stage and achieves better performance than the original Highway Network and the 5-Layer fully connected network. However, its final accuracy remains inferior to that of the Residual Network, indicating that the additional gating mechanism does not provide further improvement for the present task. The 5-Layer fully connected network converges to a lower loss than the 3-Layer models, confirming the benefit of increased network depth. A comparison between the two 3-Layer networks shows that the RK4 solver consistently achieves slightly lower loss than the Euler

solver, suggesting that a more accurate numerical discretization provides a modest improvement in PINN training.

As shown in Figure 18, the six subplots correspond to the evaluated PINN architectures: 3-Layer + Euler, 3-Layer + RK4, 5-Layer + RK4, Residual + RK4, Highway + RK4, and ResHighway + RK4. Each subplot compares the predicted and reference trajectories of the four epidemic compartments (S_u , S_a , I , and R), where the reference solution is obtained using the RK4 method. The average R^2 across all compartments is reported in each subplot as a quantitative measure of prediction accuracy. Consistent with the loss curves, the Residual Network provides the closest agreement with the reference solution, while the ResHighway Network offers moderate improvement over the Highway architecture but does not outperform the Residual Network.

Among these models, 3-Layer + Euler and 3-Layer + RK4 have relatively low average R^2 and poor fitting performance, and their overall performance is similar. The correlation coefficient of the 5-Layer + RK4 model is greatly improved, which achieves a notable promotion in fitting capability. The Highway + RK4 model obtains a higher correlation coefficient and better fitting effect than the 5-Layer structure. The ResHighway + RK4 model outperforms Highway + RK4 in terms of correlation coefficient and fitting stability. In particular, the Residual + RK4 model owns the highest correlation coefficient and presents the best fitting result, where the predicted curves are almost completely consistent with the real curves. In summary, the comprehensive performance ranking is Residual + RK4 > ResHighway + RK4 > Highway + RK4 > 5-Layer + RK4 > 3-Layer + RK4 $\approx$ 3-Layer + Euler. This indicates that adopting fourth-order RK4 method instead of Euler method and introducing residual structure can effectively increase the correlation coefficient, and significantly improve the fitting accuracy and stability of the proposed model.

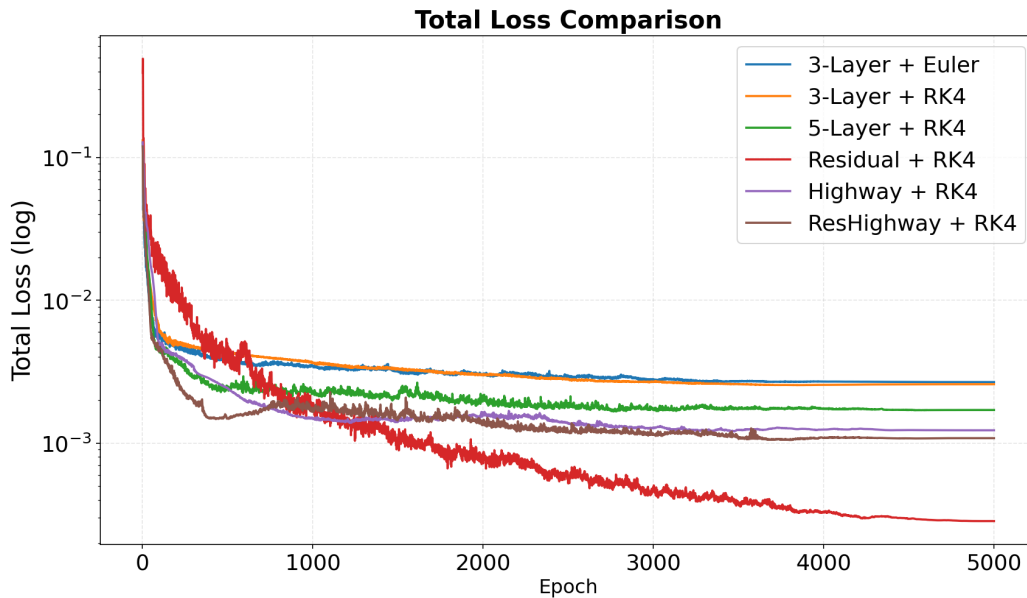


Figure 17. Averaged total loss curves of all six PINN architectures over 5000 training epochs.

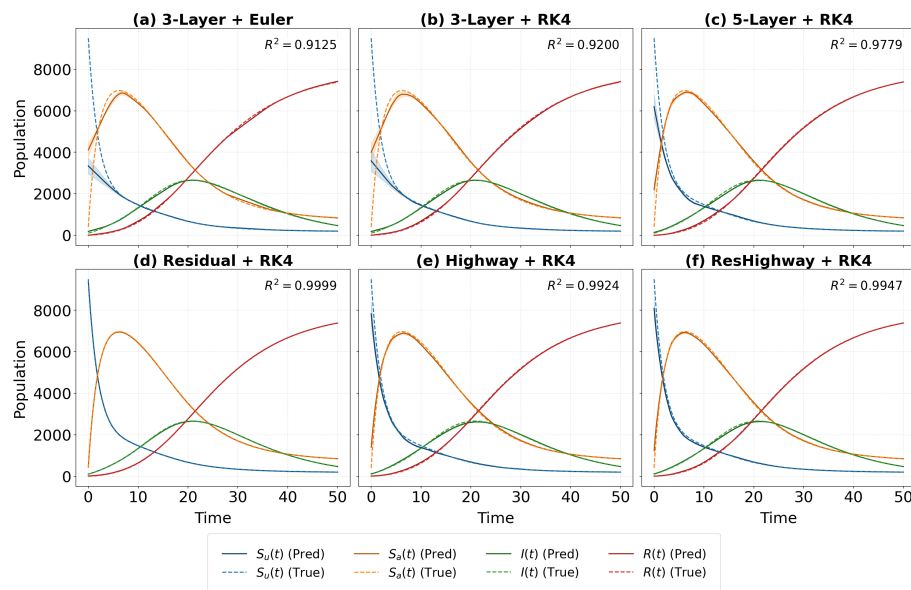


Figure 18. True vs. predicted epidemic compartment trajectories for all six PINN architectures. (a) 3-Layer + Euler ($R^2 = 0.9125$); (b) 3-Layer + RK4 ($R^2 = 0.9200$); (c) 5-Layer + RK4 ($R^2 = 0.9779$); (d) Residual + RK4 ($R^2 = 0.9999$); (e) Highway + RK4 ($R^2 = 0.9924$); (f) ResHighway + RK4 ($R^2 = 0.9947$).

Figure 19 shows the mean RMSE for infected population predictions with 95% CI. Among all models, Residual + RK4 achieves the lowest RMSE by a substantial margin, followed by ResHighway + RK4. The other models perform worse in comparison. Figure 20 presents the mean physics residual loss with 95% CIs, which shows a consistent trend: Residual + RK4 has the lowest physics residual, with ResHighway + RK4 ranking second. It should be noted that these two metrics are independent evaluations of the models' final prediction performance and are not directly equivalent to the loss terms in the training objective function.

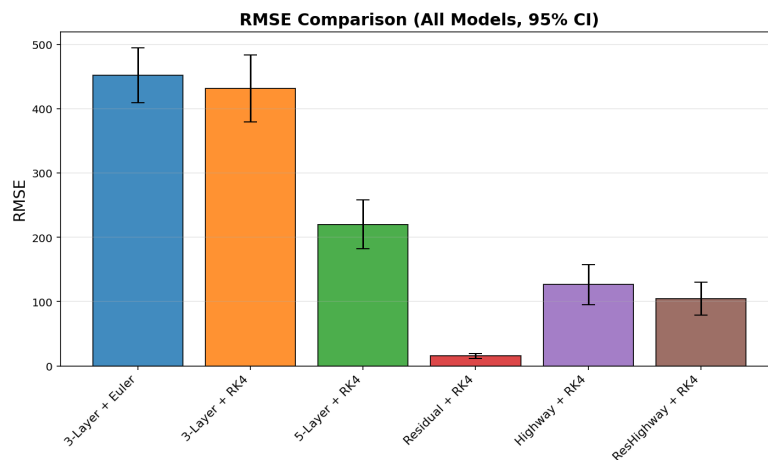


Figure 19. Mean RMSE of infected population predictions with 95% CI.

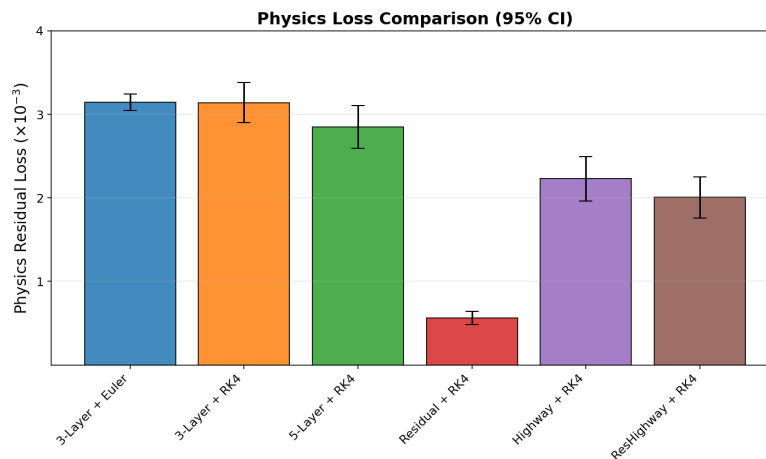


Figure 20. Mean physics residual loss with 95% CI across all models.

Table 2 summarizes the core performance indicators of the six PINN architectures, corresponding to the aforementioned figures.

Table 2. Core performance indicators.

Network Name	RMSE	CI(RMSE)	Physics Residual	CI(PR)
3-Layer + Euler	452.67	[410.05, 495.29]	3.14	[3.05, 3.24]
3-Layer + RK4	431.59	[379.38, 483.81]	3.14	[2.90, 3.38]
5-Layer + RK4	220.08	[182.12, 258.05]	2.85	[2.59, 3.11]
Residual + RK4	15.61	[11.96, 19.26]	0.56	[0.48, 0.64]
Highway + RK4	126.59	[95.21, 157.97]	2.23	[1.96, 2.50]
ResHighway + RK4	105.08	[79.47, 130.68]	2.00	[1.76, 2.25]

Note: Physics residual values are amplified by 1000 times; CI = 95% confidence interval.

Experimental results show that the original Highway network outperforms the plain 5-Layer fully connected network, but remains inferior to the Residual Network and the proposed ResHighway network. By embedding an independent residual branch into the Highway block, the ResHighway architecture achieves moderate improvements in training stability and prediction accuracy over the original Highway network. However, it does not outperform the simpler Residual Network in the present task. This result indicates that increasing architectural complexity does not necessarily improve PINN performance. Instead, the ResHighway architecture should be regarded as an exploratory design that provides a useful negative result for PINN architecture selection, while further confirming the effectiveness of residual connections in improving PINN training for differential equation solving.

The major conclusions are summarized as follows:

- (1) The RK4 method provides a modest improvement over the Euler method owing to its higher numerical accuracy.
- (2) Appropriately increasing network depth is conducive to improving the overall training performance of the model.
- (3) Among all investigated architectures, the Residual Network achieves the best overall performance

in terms of convergence, prediction accuracy, and stability, indicating that residual connections are particularly effective for improving PINN training in the epidemic modeling tasks considered in this work.

5. Conclusions

In this study, we explore the influence of media coverage on infectious disease transmission. Following the core modeling logic of published media intervention epidemic research, we construct a media-aware compartmental model extended from the classic SIR structure. During model construction and theoretical analysis, the susceptible population is stratified into unaware and aware subgroups. We formulate the evolutionary dynamics of all compartments via differential equations and systematically investigate how media intervention influences epidemic transmission characteristics. The basic mathematical properties of the model are rigorously established to support the subsequent theoretical analysis. By deriving the equilibrium points and the basic reproduction number, we characterize the relationship between epidemic evolution and equilibrium stability.

In numerical simulations, PINNs are adopted for model solving, and the numerical results agree well with the theoretical analysis. The results show that media publicity promotes the transition from unaware to aware susceptible individuals, thereby slowing epidemic growth, reducing the infection peak, and delaying its arrival. These findings further demonstrate the effectiveness of media intervention in mitigating epidemic transmission.

To improve the prediction accuracy and training stability of PINNs for epidemic dynamics, in Section 4 we investigate three algorithm optimization approaches, including numerical integration schemes, network depth, and network architectures. The RK4 method replaces the Euler scheme to improve the accuracy of physics residuals, while deeper networks enhance the approximation of nonlinear epidemic dynamics. Residual connections are further introduced, and an exploratory ResHighway architecture is constructed by combining residual learning with Highway networks for empirical comparisons.

Comparative experiments show that the optimized PINN models, especially the Residual network combined with RK4, achieve the best overall prediction accuracy and training stability. Although the proposed ResHighway architecture slightly improves upon the original Highway network, it does not outperform the simpler Residual network. This result suggests that increasing architectural complexity does not necessarily lead to better PINN performance, and highlights the effectiveness of residual connections for epidemic dynamics modeling.

It should be emphasized that the parameter values adopted in this study are typical demonstrative settings for illustrating the model's dynamical properties and evaluating the PINN architectures, and are not calibrated to epidemiological data of any specific real-world disease. As this work is positioned as a methodological study focusing on dynamical analysis and network architecture comparison, we use fixed representative parameters to maintain a fully controlled experimental setting and ensure the fairness of performance comparisons across different PINN structures. Moreover, the RK4-generated solutions used in the numerical experiments should be regarded as high-accuracy reference solutions for a synthetic benchmark, rather than as epidemiological ground truth obtained from real outbreak data. Therefore, our results mainly demonstrate the approximation and training performance of different PINN architectures under controlled dynamical conditions, rather than their real-world

forecasting capability. In practical epidemiological applications, calibrating core parameters such as transmission coefficients (β) from real surveillance data is a standard and indispensable step for obtaining disease-specific quantitative estimates. Consequently, applying the proposed PINN framework to real epidemic datasets, including parameter inference and forecasting from noisy observations, remains an important direction for future research.

In summary, this study confirms the role of media intervention in epidemic control through theoretical analysis and numerical simulation, and presents a systematically optimized PINN framework for epidemic dynamics. Building upon existing media-aware epidemic models, the main contribution of this work lies in the systematic investigation of RK4 numerical integration and different network architectures for PINN-based epidemic modeling. Among the investigated architectures, the Residual network achieves the best overall performance. The exploratory ResHighway design provides a useful empirical finding that a more complex architecture does not necessarily yield better PINN performance than a simpler residual structure.

Use of AI tools declaration

The authors declare that no artificial intelligence (AI) tools were used in the preparation of this manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. 31920240068), the 2025 Education and Teaching Reform Research Project of Gansu Province (Grant No. GJJGB041), the Key Demonstration Teaching Team of Northwest Minzu University (Grant No. 2025SFTD-03), and the Innovation Team of Ecosystem Restoration Modeling Theory and Application of Northwest Minzu University (Grant No. 10018632).

Author contributions

Minghao Song: Conceptualization, Methodology, Software, Formal analysis, Validation, Investigation, Visualization, Writing—original draft, Writing—review & editing. Hua Liu: Conceptualization, Methodology, Supervision, Project administration, Funding acquisition, Writing—review & editing. Jingyan Ma: Validation, Writing—review & editing. Kai Zhang: Methodology, Formal analysis, Validation, Writing—review & editing. Qibin Zhang: Validation, Resources, Writing—review & editing. Yumei Wei: Validation, Supervision, Writing—review & editing.

References

1. F. M. Snowden, *Epidemics and Society: From the Black Death to the Present*, Yale University Press, New Haven, 2019. <https://doi.org/10.12987/9780300249149>
2. G. P. Vigezzi, R. Vecchio, C. Barbati, G. Bonazza, P. Mazzarello, A. Odone, Historical analysis of the first smallpox vaccination campaigns in early 19-century northern Italy: Organisation and communication insights for contemporary epidemics' prevention and control, *Vaccine*, **49** (2025), 126764. <https://doi.org/10.1016/j.vaccine.2025.126764>
3. Y. Li, S. Hou, Y. Zhang, J. Liu, H. Fan, C. Cao, Effect of travel restrictions of Wuhan City against COVID-19: A modified SEIR model analysis, *Disaster Med. Public Health Prep.*, **16** (2022), 1431–1437. <https://doi.org/10.1017/dmp.2021.5>
4. World Health Organization, *WHO Guidelines for the Global Surveillance of Severe Acute Respiratory Syndrome (SARS): Updated Recommendations, October 2004*, Technical Report WHO/CDS/CSR/ARO/2004.1, World Health Organization, Geneva, 2004. Available from: <https://www.who.int/publications/i/item/WHO-CDS-CSR-ARO-2004.1>.
5. J. H. Jones, M. Salathé, Early assessment of anxiety and behavioral response to novel swine-origin influenza A(H1N1), *PLoS One*, **4** (2009), e8032. <https://doi.org/10.1371/journal.pone.0008032>
6. WHO Ebola Response Team, Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections, *N. Engl. J. Med.*, **371** (2014), 1481–1495. <https://doi.org/10.1056/nejmoa1411100>
7. W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. A*, **115** (1927), 700–721. <https://doi.org/10.1098/rspa.1927.0118>
8. R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, 1991. <https://doi.org/10.1093/oso/9780198545996.001.0001>
9. M. J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, Princeton, NJ, 2008. <https://doi.org/10.1515/9781400841035>
10. O. Diekmann, J. A. P. Heesterbeek, J. A. J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, *J. Math. Biol.*, **28** (1990), 365–382. <https://doi.org/10.1007/bf00178324>
11. M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.*, **378** (2019), 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
12. C. Millevoi, D. Pasetto, M. Ferronato, A physics-informed neural network approach for compartmental epidemiological models, *PLoS Comput. Biol.*, **20** (2024), e1012387. <https://doi.org/10.1371/journal.pcbi.1012387>
13. D. Xia, B. Diao, Q. Min, J. Mei, A SEIHR knowledge dissemination model with time-varying parameters based on physics-informed neural networks, *Neurocomputing*, **650** (2025), 130812. <https://doi.org/10.1016/j.neucom.2025.130812>

14. E. Schiassi, M. De Florio, A. D'Ambrosio, D. Mortari, R. Furfaro, Physics-informed neural networks and functional interpolation for data-driven parameters discovery of epidemiological compartmental models, *Mathematics*, **9** (2021), 2069. <https://doi.org/10.3390/math9172069>
15. S. Fujita, T. Akutsu, Enhancing epidemic forecasting with a physics-informed spatial identity neural network, *PLoS One*, **20** (2025), e0331611. <https://doi.org/10.1371/journal.pone.0331611>
16. I. V. Freire, M. M. Diniz, M. dos S. Cecconello, R. C. Bassanezi, Application of physics-informed neural networks (PINN) in the Kermack–McKendrick approach to epidemiological models, in *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, **12** (2026), 1–7. <https://doi.org/10.5540/03.2026.012.01.0255>
17. D. Xia, M. He, J. Mei, Q. Min, Physics-informed neural networks prediction approach for knowledge dissemination model considering knowledge media, *Knowl. Based Syst.*, **324** (2025), 113514. <https://doi.org/10.1016/j.knosys.2025.113514>
18. J. Mei, R. Ye, Z. Xiang, J. Hu, W. Wang, A direct data-driven intermittent control and learning via Lyapunov-guided attraction region estimation with neural feedback loop design, *IEEE Trans. Autom. Sci. Eng.*, **23** (2026), 6504–6519. <https://doi.org/10.1109/tase.2026.3669581>
19. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, 770–778. <https://doi.org/10.1109/cvpr.2016.90>
20. R. K. Srivastava, K. Greff, J. Schmidhuber, Highway networks, preprint, arXiv:1505.00387. <https://doi.org/10.48550/arXiv.1505.00387>
21. L. Liu, J. Xie, K. Li, S. Ji, Exploring how media influence preventive behavior and excessive preventive intention during the COVID-19 pandemic in China, *Int. J. Environ. Res. Public Health*, **17** (2020), 7990. <https://doi.org/10.3390/ijerph17217990>
22. S. R. Gani, S. V. Halawar, Optimal control analysis of deterministic and stochastic epidemic model with media awareness programs, *Int. J. Optim. Control Theor. Appl.*, **9** (2019), 24–35. <https://doi.org/10.11121/ijocta.01.2019.00423>
23. S. Funk, E. Gilad, V. A. A. Jansen, Endemic disease, awareness, and local behavioural response, *J. Theor. Biol.*, **264** (2010), 501–509. <https://doi.org/10.1016/j.jtbi.2010.02.032>
24. D. Aldila, Analyzing the impact of the media campaign and rapid testing for COVID-19 as an optimal control problem in East Java, Indonesia, *Chaos, Solitons Fractals*, **141** (2020), 110364. <https://doi.org/10.1016/j.chaos.2020.110364>
25. J. Cui, Y. Sun, H. Zhu, The impact of media on the control of infectious diseases, *J. Dyn. Differ. Equ.*, **20** (2008), 31–53. <https://doi.org/10.1007/s10884-007-9075-0>
26. P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Math. Biosci.*, **180** (2002), 29–48. [https://doi.org/10.1016/s0025-5564\(02\)00108-6](https://doi.org/10.1016/s0025-5564(02)00108-6)

27. E. J. Routh, *A Treatise on the Stability of a Given State of Motion: Particularly Steady Motion*, Macmillan and Company, London, 1877.
28. A. Hurwitz, Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt, *Math. Ann.*, **46** (1895), 273–284. <https://doi.org/10.1007/bf01446812>
29. K. Zhang, Y. Ji, Q. Pan, Y. Wei, Y. Ye, H. Liu, Sensitivity analysis and optimal treatment control for a mathematical model of human papillomavirus infection, *AIMS Math.*, **5** (2020), 2646–2670. <https://doi.org/10.3934/math.2020172>
30. C. Castillo-Chavez, B. Song, Dynamical models of tuberculosis and their applications, *Math. Biosci. Eng.*, **1** (2004), 361–404. <https://doi.org/10.3934/mbe.2004.1.361>
31. V. Nair, G. E. Hinton, Rectified linear units improve restricted Boltzmann machines, in *Proceedings of the 27th International Conference on Machine Learning*, PMLR, **9** (2010), 807–814.
32. J. C. Butcher, *Numerical Methods for Ordinary Differential Equations*, 2nd ed., John Wiley & Sons, Chichester, 2008. <https://doi.org/10.1002/9780470753767>
33. I. Loshchilov, F. Hutter, Decoupled weight decay regularization, preprint, arXiv:1711.05101. <https://doi.org/10.48550/arXiv.1711.05101>
34. I. Loshchilov, F. Hutter, SGDR: Stochastic gradient descent with warm restarts, preprint, arXiv:1608.03983. <https://doi.org/10.48550/arXiv.1608.03983>
35. Z. Zhang, A gentle introduction to artificial neural networks, *Ann. Transl. Med.*, **4** (2016), 370. <https://doi.org/10.21037/atm.2016.06.20>
36. B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Tang, Y. Xiao, et al., Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions, *J. Clin. Med.*, **9** (2020), 462. <https://doi.org/10.3390/jcm9020462>
37. Z. Feng, J. X. Velasco-Hernández, Competitive exclusion in a vector-host model for the dengue fever, *J. Math. Biol.*, **35** (1997), 523–544. <https://doi.org/10.1007/s002850050064>
38. P. Riyapan, S. E. Shuaib, A. Intarasit, A mathematical model of COVID-19 pandemic: A case study of Bangkok, Thailand, *Comput. Math. Methods Med.*, **2021** (2021), 6664483. <https://doi.org/10.1155/2021/6664483>
39. G. Akrivis, C. G. Makridakis, C. Smaragdakis, Runge–Kutta physics informed neural networks: Formulation and analysis, *Numer. Math.*, **157** (2025), 1975–2016. <https://doi.org/10.1007/s00211-025-01501-7>
40. M. Telgarsky, Benefits of depth in neural networks, in *Proceedings of the 29th Annual Conference on Learning Theory*, PMLR, **49** (2016), 1517–1539. Available from: <https://proceedings.mlr.press/v49/telgarsky16.html>.
41. G. F. Montúfar, R. Pascanu, K. Cho, Y. Bengio, On the number of linear regions of deep neural networks, preprint, arXiv:1402.1869. <https://doi.org/10.48550/arXiv.1402.1869>

42. Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, *IEEE Trans. Neural Networks*, **5** (1994), 157–166. <https://doi.org/10.1109/72.279181>
43. G. Zhang, H. Yang, G. Pan, Y. Duan, F. Zhu, Y. Chen, Constrained self-adaptive physics-informed neural networks with ResNet block-enhanced network architecture, *Mathematics*, **11** (2023), 1109. <https://doi.org/10.3390/math11051109>
44. T. Zia, S. Razzaq, Residual recurrent highway networks for learning deep sequence prediction models, *J. Grid Comput.*, **18** (2020), 169–176. <https://doi.org/10.1007/s10723-018-9444-4>
45. G. Casella, R. L. Berger, *Statistical Inference*, 2nd ed., Duxbury Press, Pacific Grove, CA, 2002.



AIMS Press

©2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)