Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 25/16

Achraf Zinihi

Identifying Memory Effects in Epidemics via a Fractional SEIRD Model and Physics-Informed Neural Networks

September 26, 2025

http://www.imacm.uni-wuppertal.de

Identifying Memory Effects in Epidemics via a Fractional SEIRD Model and Physics-Informed Neural Networks

Achraf Zinihi

University of Wuppertal, Applied and Computational Mathematics, Gaußstrasse 20, 42119 Wuppertal, Germany

Moulay Ismail University of Meknes, FST Errachidia, Department of Mathematics, AMNEA Group, Errachidia 52000, Morocco

Abstract

We develop a physics-informed neural network (PINN) framework for parameter estimation in fractional-order SEIRD epidemic models. By embedding the Caputo fractional derivative into the network residuals via the L1 discretization scheme, our method simultaneously reconstructs epidemic trajectories and infers both epidemiological parameters and the fractional memory order α. The fractional formulation extends classical integer-order models by capturing long-range memory effects in disease progression, incubation, and recovery. Our framework learns the fractional memory order α as a trainable parameter while simultaneously estimating the epidemiological rates $(\beta, \sigma, \gamma, \mu)$. A composite loss combining data misfit, physics residuals, and initial conditions, with constraints on positivity and population conservation, ensures both accuracy and biological consistency. Tests on synthetic Mpox data confirm reliable recovery of α and parameters under noise, while applications to COVID-19 show that optimal $\alpha \in (0, 1]$ captures memory effects and improves predictive performance over the classical SEIRD model. This work establishes PINNs as a robust tool for learning memory effects in epidemic dynamics, with implications for forecasting, control strategies, and the analysis of non-Markovian epidemic processes.

Keywords: Fractional epidemic models, Physics-informed neural networks, Caputo derivative, Parameter estimation, Memory effects in epidemics.
2020 Mathematics Subject Classification: 92C60, 26A33, 65L05, 68T07.

1. Introduction

The global spread of infectious diseases such as the Black Death (1347–1351 in Europe, Asia, and North Africa), smallpox (16th–20th century, eradicated the Americas in 1980), cholera (seven waves from 1817 to 1975), H1N1 influenza (2009–2010, swine flu), Ebola (notably 2014–2016 in West Africa), and COVID-19 (2019–present, caused by SARS-CoV-2) has highlighted the importance of reliable epidemic models for understanding disease dynamics and guiding public health interventions [1–7]. Traditional compartmental models, such as the classical SI, SIR, SEIR and SEIRD frameworks, have long served as fundamental tools for capturing the transmission, progression, and recovery processes in populations [8–14]. These models are widely used due to their simplicity and analytical tractability. However, they often fail to fully capture the real-world epidemic dynamics, such as the long-term memory effects and anomalous diffusion observed in disease spreading.

Email address: a.zinihi@edu.umi.ac.ma (Achraf Zinihi)

To address these limitations, researchers have increasingly turned to fractional-order epidemic models. These models incorporate memory and hereditary effects into disease dynamics by replacing classical integer-order derivatives with fractional operators [15–21]. The fractional-order parameter, usually denoted by α, characterizes the degree of memory in the system. When $\alpha = 1$, the model reduces to the standard integer-order case. When $0 < \alpha < 1$, the model reflects subdiffusive, history-dependent dynamics. Several studies have demonstrated that fractional epidemic models can provide more accurate fits to epidemiological data and reveal dynamics not captured by classical models [21–23]. Readers seeking a more in-depth discussion are referred to the following comprehensive studies [24–33]. These studies provide extensive coverage of fractional calculus applications.

Despite these advantages, applying fractional epidemic models presents several challenges. Estimating the memory order α from real epidemic data is particularly challenging. It is often nonintuitive and highly sensitive to noise or data sparsity. Furthermore, it is strongly correlated with other epidemiological rates. Furthermore, classical fitting approaches based on least-squares optimization require repeated numerical simulations of the fractional model. These simulations are computationally demanding due to the nonlocal nature of the Caputo derivative [34–37]. Conversely, purely data-driven machine learning approaches often ignore the underlying epidemiological structure, leading to biologically inconsistent outcomes, such as negative compartment sizes or loss of population conservation. These challenges motivate developing hybrid approaches that integrate the mechanistic structure of epidemic models with the flexibility of modern data-driven methods.

In this regard, *physics-informed neural networks* (PINNs) are a promising alternative. By embedding the governing equations into the neural network training process, PINNs enforce physical consistency while fitting observational data simultaneously [38–40]. This makes them particularly attractive for epidemiological models, where constraints such as positivity, conservation, and dynamical consistency are essential. While recent works have successfully applied PINNs to epidemic modeling in the integer-order setting [40–43], the integration of fractional-order operators into PINNs remains largely unexplored.

In this study, we present a PINN-based framework for estimating the memory order α and, optionally, key epidemiological parameters, such as transmission, recovery, and mortality rates, in fractional SEIRD models. Our methodology uses the L1 discretization scheme for the Caputo derivative to represent fractional dynamics in the neural network loss function. We validate our approach on synthetic datasets where the ground-truth parameters are known and on real datasets from multiple regions in Europe. In addition to demonstrating the feasibility of estimating α from data, we examine the identifiability and interpretability of fractional epidemic dynamics compared to classical integer-order models. The main contributions of this study are:

- We develop a fractional SEIRD model with Caputo derivatives that incorporates memory effects into epidemic dynamics.
- We design a PINN framework that can handle fractional operators via the L1 scheme and enforce epidemiological constraints.
- We demonstrate that the method can recover the fractional memory order α and epidemiological parameters from synthetic and real epidemic datasets.
- We perform uncertainty and identifiability analyses to highlight the role of memory effects in the spread of epidemics and their implications for forecasting.

This work lies at the intersection of fractional calculus, epidemiological modeling, and machine learning. It offers a novel perspective on data-driven parameter estimation in epidemic

systems. The rest of the paper is organized as follows. Section 2 provides the necessary preliminaries on fractional calculus and neural networks. Section 3 introduces the fractional SEIRD model and presents its well-posedness analysis. The proposed methodology, including the PINN framework, the discretization of the Caputo derivative, the parameter estimation strategy, the loss function, and the training algorithm, is detailed in Section 4. Section 5 describes the experiments on both synthetic and real epidemic datasets and discusses the obtained results. Finally, Section 6 concludes the paper and outlines directions for future work.

2. Preliminaries

This section establishes the mathematical foundation of our study. Its purpose is to enable readers, especially those unfamiliar with fractional calculus, to understand the subsequent model description. In this work, we adopt the Caputo fractional derivative of order $\alpha \in (0,1)$, which is well-suited for initial value problems. Let $T > 0$ be a finite time horizon, and let $f\colon [0,T] \to \mathbb{R}$ be a sufficiently smooth function.

First, we recall the Gamma function, which generalizes the factorial function to complex arguments and often appears in the formulation of fractional operators.

Definition 1 (See [44, p. 24])**.** *Let $\omega \in \mathbb{C}$ such that $\mathrm{Re}(\omega) > 0$. The Gamma function is given by*

$$\Gamma(\omega) := \int_0^{+\infty} e^{-\tau}\tau^{\omega-1} d\tau.$$

The Gamma function can be used to introduce the fractional integral operator, which generalizes the concept of repeated integration to include non-integer orders.

Definition 2 (See [44, p. 69])**.** *The fractional integral operator with base point 0 is written as*

$$\mathcal{I}^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} f(\tau)\, d\tau.$$

For completeness, we recall the definitions of the left- and right-sided Caputo derivatives, since they play a central role in the formulation and analysis of fractional differential equations.

Definition 3 (See [44, p. 91])**.** *The left and right fractional derivatives in the Caputo sense of f with base point 0 and T, respectively, are defined by*

$$^{\mathcal{C}}\mathcal{D}_t^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} f'(\tau)\, d\tau, \tag{1}$$

and

$$^{\mathcal{C}}_{T}\mathcal{D}_t^\alpha f(t) = \frac{-1}{\Gamma(1-\alpha)} \int_t^T (\tau-t)^{-\alpha} f'(\tau)\, d\tau.$$

Remark 1. *As α approaches 1 from below, the case reduces to the classical first-order derivative, recovering the memoryless ordinary differential equation (ODE) framework.*

We now recall a useful identity linking the Caputo derivative and the fractional integral, which will be used later in the analysis of our model.

Lemma 1 (See [44, p. 96])**.** *One has*

$$\mathcal{I}^\alpha \left({}^{\mathcal{C}}\mathcal{D}_t^\alpha f(t) \right) = f(t) - f(0).$$

3. Model Description

This section presents the mathematical framework used to describe the transmission dynamics of infectious diseases. We adopt an SEIRD compartmental model, which partitions the population into five epidemiological groups. This structure provides a flexible basis for incorporating fractional-order dynamics that capture memory effects in disease progression.

3.1. Fractional SEIRD Model

The SEIRD model is a widely adopted compartmental structure in mathematical epidemiology because it captures the fundamental stages of disease progression from exposure to removal. In this formulation, the total population is divided into five epidemiological classes: susceptible individuals $S(t)$, who are at risk of infection; exposed individuals $E(t)$, who have been infected but are not yet infectious; infectious individuals $I(t)$, who can transmit the disease to others; recovered individuals $R(t)$, who have acquired immunity; and deceased individuals $D(t)$, who have died due to the disease.

The transmission dynamics are governed by four key epidemiological parameters:

(i) the infection rate β, (ii) the incubation rate σ,

(iii) the recovery rate γ, (iv) the disease-induced mortality rate μ.

To incorporate memory effects and long-range temporal dependencies, we extend the model using a Caputo fractional derivative of order $\alpha \in (0, 1]$. This generalization allows the current rate of change in each compartment to depend on both the present state and the history of the process. Thus, it captures the nonlocal character of epidemic spread.

Mathematically, the resulting fractional-order SEIRD system can be written as

$$\begin{cases} {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} S(t) = -\beta \dfrac{S(t)I(t)}{N_L(t)}, \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} E(t) = \beta \dfrac{S(t)I(t)}{N_L(t)} - \sigma E(t), \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} I(t) = \sigma E(t) - (\gamma + \mu) I(t), \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} R(t) = \gamma I(t), \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} D(t) = \mu I(t), \end{cases} \tag{2}$$

supplied with the initial conditions

$$S(0) = S_0 \geq 0, \quad E(0) = E_0 \geq 0, \quad I(0) = I_0 \geq 0, \quad R(0) = R_0 \geq 0, \quad D(0) = D_0 \geq 0. \tag{3}$$

Here, $N_L(t) = S(t) + E(t) + I(t) + R(t)$ denotes the total living population, and $N = N_L(t) + D(t) = S(t) + E(t) + I(t) + R(t) + D(t)$ represents the total population. In the absence of natural births and deaths, N remains constant over time, as shown in Section 3.2.

This fractional SEIRD formulation (2) generalizes the classical integer-order model: when $\alpha = 1$ (i.e. $\alpha \to 1^-$), the system reduces to the standard SEIRD model; when $\alpha < 1$, the system captures memory-driven dynamics, reflecting the influence of past states on present transmission. This property is particularly relevant in real epidemics, where delays in infection, incubation, and recovery processes often exhibit non-exponential waiting time distributions that cannot be adequately captured by ODEs.

Remark 2. *In this formulation (2), we do not include vital dynamics such as natural births and deaths. This choice is justified by the fact that epidemic outbreaks typically evolve on a*

much shorter timescale than demographic changes. By excluding vital dynamics, we ensure that the total population remains constant, allowing us to isolate the effects of disease-induced transitions and the fractional memory order α, which is the main focus of this study.

Before we proceed to the methodological aspects of model learning with PINNs, we must first establish the mathematical soundness of the proposed system. The following subsection examines the well-posedness of the fractional SEIRD model (2)–(3), focusing on the existence, uniqueness, and positivity of solutions under biologically meaningful conditions.

3.2. Well-posedness Analysis

This section establishes that the model (2)–(3)is mathematically and biologically well-posed. Consider each equation in (2) on the boundary of the non-negative orthant

$$ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} S(t)\big|_{S=0} = 0,\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} E(t)\big|_{E=0} = \beta \frac{S(t)I(t)}{N_L(t)} \geq 0,\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} I(t)\big|_{I=0} = \sigma E(t) \geq 0, $$

$$ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} R(t)\big|_{R=0} = \gamma I(t) \geq 0,\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} D(t)\big|_{D=0} = \mu I(t) \geq 0. $$

Thus, on each coordinate hyperplane, the corresponding derivative points inward to $\mathbb{R}^5_+$. By the generalized mean value theorem for Caputo derivatives [45, p. 288], the solution remains non-negative for all $t \geq 0$.

Now, summing the first four equations of (2) yields

$$ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} N_L(t) = -\mu I(t) \leq 0. $$

Hence, N_L is non-increasing and bounded above by its initial value, i.e. $N_L(0) = N_{L_0}$. Moreover,

$$ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} N(t) = 0, $$

which means $N(t) \equiv N(0) = N_0$. Therefore, every trajectory of (2) remains in the compact positively invariant set

$$ \Omega = \Big\{ (S, E, I, R, D) \in \mathbb{R}^5_+ \,:\, S + E + I + R + D = N_0 \Big\}. $$

Let us denote $X(t) = \big(S(t), E(t), I(t), R(t), D(t)\big)^T$. The system (2) can be written in a compact form as

$$ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} X(t) = F(X(t)), \quad \text{where} \quad X(0) = X_0 \in \mathbb{R}^5_+, $$

where $F\colon \mathbb{R}^5_+ \to \mathbb{R}^5$, the right-hand side of (2), is continuous and locally Lipschitz, because of the boundedness of X. By the general theory of fractional differential equations (see e.g. [19]), the system admits a unique solution defined on $[0, \infty)$. The proposition below summarizes this section.

Proposition 1. *For any given positive initial conditions and positive parameters β, σ, γ, μ, the fractional system (2)–(3) has a unique global solution $X(t)$ on $[0, \infty)$. Furthermore, this solution remains positive and bounded for all $t \geq 0$.*

3.3. Normalization and Scaling

To facilitate the analysis and design of the PINN framework, we introduce dimensionless variables that represent proportions of the total population

$$s(t) = \frac{S(t)}{N},\ e(t) = \frac{E(t)}{N},\ i(t) = \frac{I(t)}{N},\ r(t) = \frac{R(t)}{N},\ \text{ and }\ d(t) = \frac{D(t)}{N},$$

where $N_L(t) = N(1 - d(t))$. Dividing the fractional SEIRD model (2) by N, and using the linearity of the Caputo operator ${}^{\mathcal{C}}\mathcal{D}_t^{\alpha}$, we obtain the normalized system

$$\begin{cases} {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} s(t) = -\beta \dfrac{s(t)\, i(t)}{1 - d(t)}, \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} e(t) = \beta \dfrac{s(t)\, i(t)}{1 - d(t)} - \sigma e(t), \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} i(t) = \sigma e(t) - (\gamma + \mu) i(t), \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} r(t) = \gamma i(t), \\ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} d(t) = \mu i(t), \end{cases} \tag{4}$$

supplied with initial conditions

$$s(0) = s_0, \quad e(0) = e_0, \quad i(0) = i_0, \quad r(0) = r_0, \quad \text{and} \quad d(0) = d_0. \tag{5}$$

Since ${}^{\mathcal{C}}\mathcal{D}_t^{\alpha}(s + e + i + r + d) = 0$, showing that the simplex

$$\Delta = \big\{(s, e, i, r, d) \in \mathbb{R}_+^5 \ :\ s + e + i + r + d = 1 \text{ and } d < 1\big\},$$

is forward invariant. Thus, all state variables remain nonnegative and bounded by 1 whenever the initial conditions belong to Δ.

This normalized formulation (4)–(5) offers several advantages for the subsequent PINN methodology: (i) the state variables are naturally bounded in $[0, 1]$, which improves numerical conditioning and facilitates the use of positivity or conservation constraints within the neural network; (ii) working with proportions yields more easily interpretable and comparable parameters across different datasets; (iii) the explicit denominator $1 - d(t)$ preserves the impact of mortality on transmission while keeping the equations scale-free. All methodological developments in the subsequent subsections are based on this normalized system (4)–(5).

4. Methodology

This section describes the proposed methodology for learning epidemic trajectories and estimating the parameters of the fractional SEIRD model. Our approach leverages PINNs, which integrate observational data with the governing fractional differential equations to ensure consistency with epidemic dynamics. Using system (4)–(5), the PINN simultaneously approximates the compartmental trajectories and recovers the parameters, including the fractional order α, while enforcing positivity and conservation of the total population.

4.1. PINN Framework

The key idea of PINNs is to approximate the solution to a system of differential equations using a neural network whose training is guided by both observational data and the governing equations. In our setting, the network learns the normalized epidemic trajectories

$$\big(s(t), e(t), i(t), r(t), d(t)\big) \in \Delta,$$

along with the unknown parameters β, σ, γ, μ, and, most importantly, the fractional order α.

4.1.1. Network Architecture

We employ a fully connected feedforward neural network (multilayer perceptron) that takes time $t \in [0, T]$ as input and outputs approximations

$$\mathcal{NN}^{\theta,\Theta} \colon t \mapsto \big(\hat{s}(t;\theta), \hat{e}(t;\theta), \hat{i}(t;\theta), \hat{r}(t;\theta), \hat{d}(t;\theta)\big),$$

where θ and Θ denote the trainable weights and biases of the network, respectively.

To maintain epidemiological validity, we enforce nonnegativity and approximate population conservation through the output layer. This can be achieved either by using *softplus*, an activation function defined as $softplus(x) = \ln(1 + e^x)$. This function smoothly maps real numbers to positive values, ensuring nonnegativity without sharp cutoffs, or activations to guarantee positivity. Alternatively, we can apply a *softmax*, a normalization function that transforms a vector into positive components that sum to one: $softmax(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$. It is commonly used to represent probabilities or normalized fractions. The softmax projection ensures

$$\hat{s}(t) + \hat{e}(t) + \hat{i}(t) + \hat{r}(t) + \hat{d}(t) \approx 1,$$

so that the total population is conserved. This constraint can be imposed explicitly via a simplex-based architecture or softly enforced through a penalty term in the loss function (see Sections 4.4 and 4.5).

4.1.2. Physics-Informed Residuals

Rather than training solely on observed data, the PINN incorporates the fractional SEIRD equations (4) into the loss function by ensuring that the neural outputs satisfy the dynamics in the sense of the Caputo derivative. Specifically, the fractional derivatives ${}^{\mathcal{C}}\mathcal{D}_t^{\alpha}\hat{s}(t)$, ${}^{\mathcal{C}}\mathcal{D}_t^{\alpha}\hat{e}(t)$, etc., are approximated numerically (see Section 4.2) and substituted into the right-hand side of the system. The resulting residuals vanish for the true solution; therefore, minimizing them enforces consistency between the neural predictions and the governing epidemic model.

4.1.3. Joint Learning of Trajectories and Parameters

The network weights θ and the epidemiological parameters $\Theta = \{\beta, \sigma, \gamma, \mu, \alpha\}$ are updated simultaneously. This hybrid approach allows the model to reconstruct unobserved states and estimate parameters from limited data. The embedded physics serve as a regularizer, mitigating overfitting and enhancing interpretability.

4.2. Caputo Derivative Discretization

For a sufficiently smooth function ψ, we recall the Caputo derivative ${}^{\mathcal{C}}\mathcal{D}_t^{\alpha}\psi$ defined in (1). Its nonlocal nature introduces long-range memory effects, which are central to our analysis. To efficiently approximate the derivative at discrete time points $t_n = n\Delta t$, we adopt the classical L1 scheme (see, e.g., [46, p. 34-40])

$$ {}^{\mathcal{C}}\mathcal{D}_t^{\alpha}\psi(t_n) \;\approx\; \frac{1}{\Delta t^{\alpha}} \sum_{k=0}^{n-1} c_k^{(n)} \big(\psi(t_{n-k}) - \psi(t_{n-k-1})\big), \tag{6}$$

with weights

$$c_k^{(n)} = \frac{1}{\Gamma(2-\alpha)} \big[(k+1)^{1-\alpha} - k^{1-\alpha}\big], \quad \text{with} \quad k = 0, 1, \ldots, n-1.$$

This discretization plays a crucial role in the PINN framework developed in this paper because it allows fractional-order dynamics to be directly embedded into the loss function. The L1

scheme (6) is differentiable with respect to both $\psi(t)$ and α, enabling gradients to propagate through the discretization during training. In practice, the L1 operator is implemented as a differentiable layer within the PINN to ensure efficient backpropagation.

The neural network outputs $\hat{s}(t)$, $\hat{e}(t)$, $\hat{i}(t)$, $\hat{r}(t)$, and $\hat{d}(t)$ are evaluated at collocation points $\{t_j\}$, and their Caputo derivatives are computed using the L1 scheme (6). These approximations are substituted into the residuals of the fractional SEIRD system to ensure that the network predictions satisfy the underlying dynamics.

4.3. Parameter Estimation Strategy

The PINN is designed not only to approximate epidemic trajectories but also to identify the epidemiological parameters

$$\Theta = \{\beta, \sigma, \gamma, \mu, \alpha\}.$$

All parameters are treated as trainable variables and updated jointly with the network weights. To ensure admissibility and stability, the following design choices are imposed:

- **Positivity:** All rates are parameterized via *softplus* transformations, e.g. $\beta = softplus(\mathscr{C}_\beta)$, ensuring $\beta > 0$, where $\mathscr{C}_\beta \in \mathbb{R}$ and $\beta \leq \beta_{max}$.
- **Fractional Order:** The memory parameter α is restricted to $(\alpha_{\min}, 1]$ using a sigmoid mapping

  $$\alpha = \alpha_{\min} + (1 - \alpha_{\min})\,\sigma(z_\alpha),$$

  where $z_\alpha \in \mathbb{R}$ and $\sigma(\cdot)$ denotes the sigmoid function.
- **Staged Optimization:** Training proceeds in two phases:
  - (i) an initialization phase where α is fixed at 1 (classical SEIRD) and only the epidemiological parameters are trained;
  - (ii) a joint phase where α is released and estimated simultaneously with the other parameters.

Initial guesses for Θ are obtained from classical SEIRD fits or epidemiological literature, and parameter bounds are imposed to exclude unrealistic values.

4.4. Loss Function

In the PINN framework, the loss function plays a central role in guiding the neural network to approximate both the epidemic trajectories and the underlying dynamics of the fractional SEIRD system (4)–(5). Unlike purely data-driven models, which only minimize prediction error, here the loss function in the PINN framework must balance multiple objectives: fitting available observations, enforcing the governing fractional differential equations, satisfying initial conditions, and maintaining epidemiological consistency, such as positivity and population conservation. To achieve this balance, we construct a composite loss function that integrates data misfit with physics-based residuals and additional structural constraints. This design ensures that the trained network accurately reflects the observed data while remaining faithful to the model's mathematical and biological principles.

The PINN is trained by minimizing the composite loss functional

$$\mathcal{L}(\theta, \Theta) = \lambda_{\text{data}}\mathcal{L}_{\text{data}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}} + \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}. \tag{7}$$

Each component is defined as follows:

Data Misfit: Penalizes discrepancies between predicted and observed compartments $\mathcal{O} = \{s, e, i, r, d\}$,

$$\mathcal{L}_{\text{data}} = \frac{1}{N_{\text{data}}} \sum_{j=1}^{N_{\text{data}}} \sum_{\mathrm{x} \in \mathcal{O}} \left| \hat{\mathrm{x}}(t_j) - \mathrm{x}_j^{\text{obs}} \right|^2.$$

Physics Residual: Enforces consistency with the fractional SEIRD dynamics at collocation points

$$\mathcal{L}_{\text{phys}} = \frac{1}{N_{\text{coll}}} \sum_{j=1}^{N_{\text{coll}}} \sum_{\mathrm{x} \in \mathcal{O}} \left| \mathcal{R}_{\mathrm{x}}(t_j) \right|^2,$$

where $\mathcal{R}_{\mathrm{x}}(t) = {}^{\mathcal{C}}\mathcal{D}_t^{\alpha} \hat{\mathrm{x}}(t) - F_{\mathrm{x}}(\hat{\mathrm{x}}, \beta, \sigma, \gamma, \mu)$ is the residual of the governing equation (4), and ${}^{\mathcal{C}}\mathcal{D}_t^{\alpha}$ is discretized using the L1 scheme (6).

Initial Conditions: Ensures agreement with prescribed initial values

$$\mathcal{L}_{\text{IC}} = \sum_{\mathrm{x} \in \mathcal{O}} \left| \hat{\mathrm{x}}(0) - \mathrm{x}_0 \right|^2.$$

Conservation: Penalizes violations of the population balance $\hat{s} + \hat{e} + \hat{i} + \hat{r} + \hat{d} = 1$ at collocation points

$$\mathcal{L}_{\text{cons}} = \frac{1}{N_{\text{coll}}} \sum_{j=1}^{N_{\text{coll}}} \left(\hat{s}(t_j) + \hat{e}(t_j) + \hat{i}(t_j) + \hat{r}(t_j) + \hat{d}(t_j) - 1 \right)^2.$$

Regularization: ℓ^2 penalties on the network weights (and optionally the parameters θ) to avoid overfitting

$$\mathcal{L}_{\text{reg}} = \lambda_{\Theta} \|\Theta\|_2^2 + \lambda_{\theta} \|\theta\|_2^2.$$

The weights λ_* balance the contributions of each term and are tuned empirically to achieve stability between data fidelity and physics enforcement.

Figure 1 summarizes the complete architecture of the proposed PINN. The neural network receives time t as input and outputs approximations of the compartmental trajectories $(\hat{s}, \hat{e}, \hat{i}, \hat{r}, \hat{d})$. These outputs are processed through the Caputo derivative operator ${}^{\mathcal{C}}\mathcal{D}_t^{\alpha}$, discretized using the L1 scheme (6), and compared against the fractional SEIRD dynamics (4) to form physics residuals. Along with the data misfit, physics and initial condition enforcement, conservation penalties, and regularization, these residuals contribute to the composite loss (7), which guides the training.

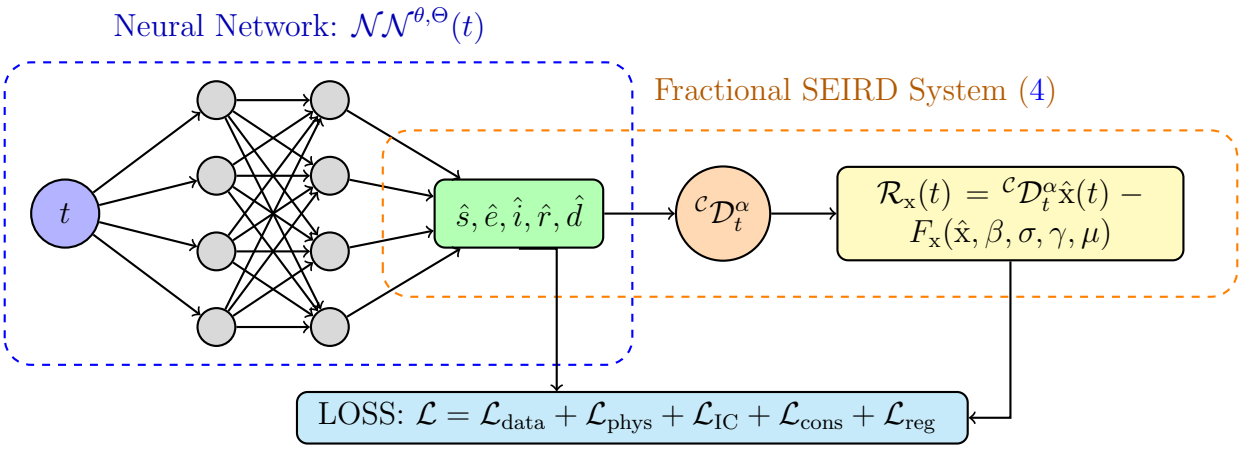


Figure 1: Schematic diagram of the proposed PINN for the fractional SEIRD model (4)–(5).

4.5. Training Algorithm

The training procedure integrates the loss function defined in Section 4.4 with staged optimization strategies that are designed to improve the stability and identifiability of the fractional order α. The workflow is summarized below.

Algorithm 1 Fractional-SEIRD PINN Training Algorithm

1: *Initialization:* Neural network weights Θ are initialized using Xavier initialization [47]. Epidemiological parameters $\theta = (\beta, \sigma, \gamma, \mu)$ are drawn from admissible ranges informed by prior studies, while the fractional order α is initialized close to 1 to reflect the classical Markovian case (see Remark 3)

2: *Staged Optimization:* Training proceeds in two phases:

- i : *Pretraining*: Fix $\alpha = 1$ and minimize a reduced loss $\mathcal{L}_{\text{data}} + \mathcal{L}_{\text{IC}}$ to fit observations and initial conditions;
- ii : *Joint training*: Release α and optimize the full composite loss $\mathcal{L}$ in (7), updating both network weights and epidemiological parameters.

3: *Optimization Scheme:* We employ a hybrid optimizer [38, 48, 49]: Adam with a decaying learning rate is used for initial exploration, followed by L-BFGS for fine-tuning. This hybrid strategy has been shown to be effective in PINNs.

4: *Constraints and Monitoring:* Positivity and conservation are enforced through the loss penalties described in Section 4.4. Training is monitored via the total loss and its components. Early stopping is applied when improvements fall below the tolerance threshold.

Remark 3. *When $\alpha = 1$, the fractional derivative reduces to the classical first-order derivative, resulting in the standard Markovian SEIRD model with memoryless dynamics. Therefore, initializing α near 1 therefore, lets the model begin training close to the classical case and adapt toward sub-exponential, memory-dependent dynamics if the data supports it.*

Algorithm 1 simultaneously reconstructs latent epidemic trajectories and reliably estimates the fractional order α, thereby quantifying memory effects while maintaining epidemiological interpretability.

5. Results and Discussion

To evaluate the proposed fractional SEIRD model (4)–(5) with PINNs, we conduct experiments on both synthetic and real epidemic datasets, designed to address the following questions:

- Can the framework accurately recover the fractional memory order α and epidemiological parameters from noisy data?
- How does the inclusion of memory effects improve predictive performance compared to classical integer-order SEIRD models?
- What is the role of different architectural and methodological choices (e.g., fixed vs trainable α, conservation constraints) in the model's performance?
- How reliable are the estimated parameters and predictions, given uncertainty and data limitations?

To this end, we first validate the approach using synthetic epidemic trajectories that mimic the dynamics of the Mpox epidemic, for which the ground-truth parameters $(\beta, \sigma, \gamma, \mu, \alpha)$ are

known. We simulate the dynamics of the fractional SEIRD model (4) through the L1 discretization of the Caputo derivative (6) for various values of $\alpha \in (0, 1]$, with Gaussian noise added to replicate observation errors. Next, we use the procedure in Section 4.5 to train the PINN to infer α and the epidemiological parameters from partial and noisy observations. This confirms the identifiability and stability of the proposed learning strategy before turning to real-world data. We then apply the method to COVID-19 datasets, focusing on confirmed, recovered, and deceased case counts from several European regions. We assess performance through trajectory fitting, estimation of the optimal fractional order α, and comparison with the predictive accuracy of the fractional SEIRD model.

Section 5.1 presents the results for the synthetic setting, based on numerical solutions obtained using the L1 discretization scheme (6). Section 5.2 details the evaluation on real datasets, emphasizing predictive accuracy, parameter estimation, and epidemiological interpretation.

5.1. Synthetic Data Experiments

The parameters were selected within biologically plausible ranges reported in the literature for Mpox, as shown in Table 1.

Table 1: Epidemiological parameters for the SEIRD model applied to Mpox (clade Ib, 2024–2025).

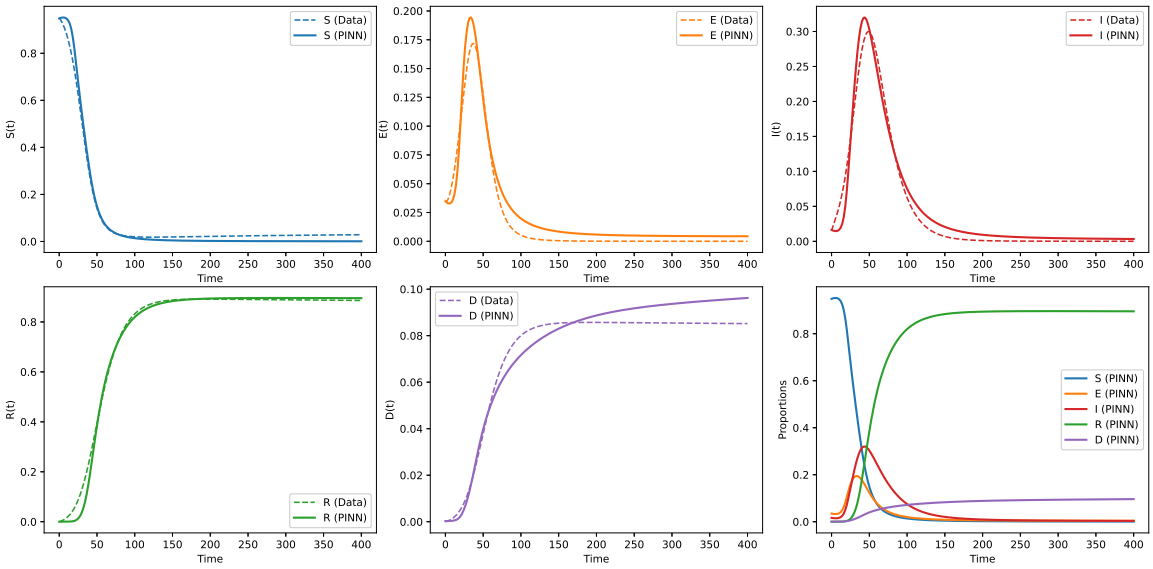
Epidemic	β	σ	γ	μ	**References**
Mpox	0.1 – 0.3	0.077 – 0.2	0.036 – 0.071	0.001 – 0.03	[50, 51]

Specifically, the fixed values were set to

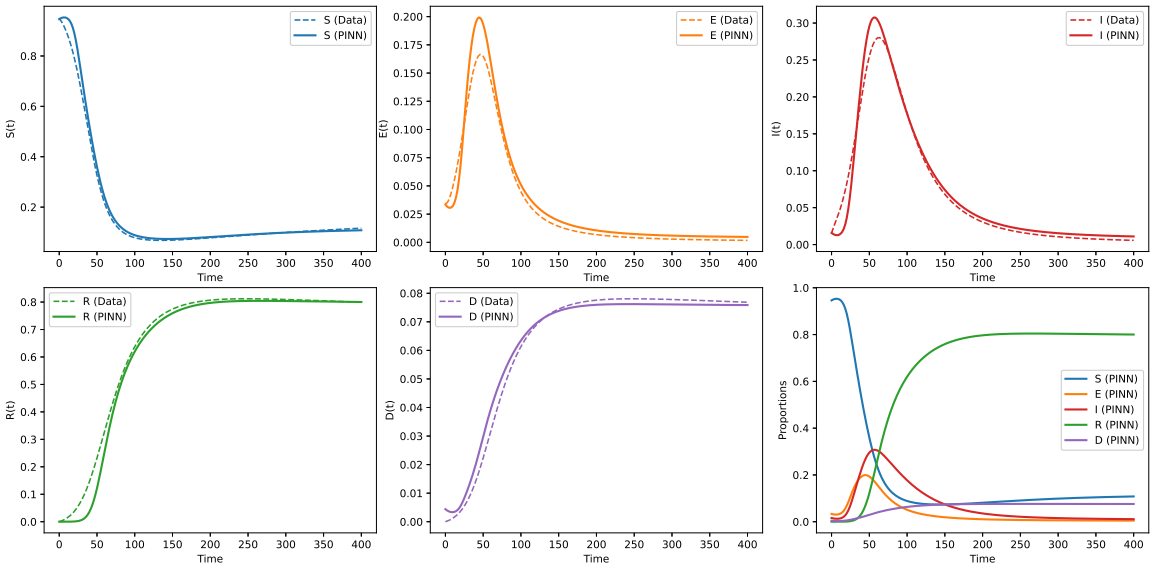
$$\beta = 0.25, \quad \sigma = 0.13, \quad \gamma = 0.052, \quad \mu = 0.005,$$

while the memory order varied as $\alpha \in \{1, 0.95, 0.9\}$. The case $\alpha = 1$ corresponds to the classical integer-order SEIRD model, while $\alpha < 1$ incorporates fractional memory effects. The values $\alpha = 0.95$ and $\alpha = 0.9$ were chosen to introduce moderate deviations from the classical case. These values remain close enough to $\alpha = 1$ to ensure biological plausibility and numerical stability, while still reflecting realistic long-memory effects, such as incubation variability, reporting delays, or behavioral adaptation.

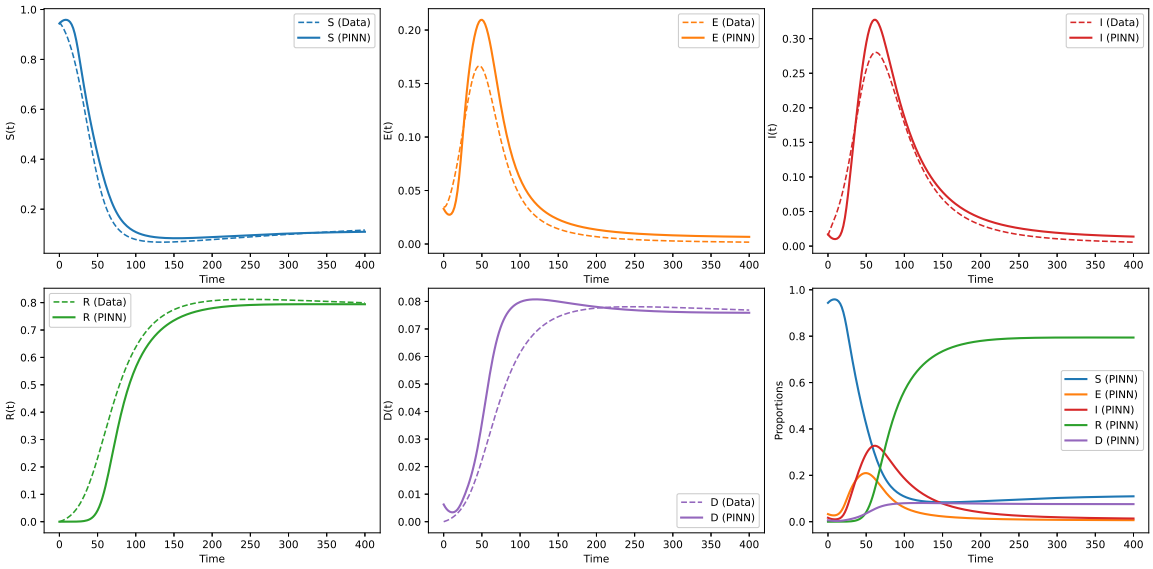
Figures 2 and 3 provide a visual comparison of the reconstructed epidemic dynamics and the corresponding parameter estimates under different fixed values of α. Figure 2 shows that the predicted trajectories for all compartments (S, E, I, R, D) closely align with the ground-truth synthetic data, demonstrating the PINN's ability to capture both the transient and steady-state behaviors of the system. As α decreases, the infectious peak is slightly delayed and flattened, highlighting the moderating role of memory effects in epidemic progression. Figure 3 complements this analysis by illustrating the estimated epidemiological parameters across the three scenarios. Table 2 summarizes these estimated values.



(a) $\alpha = 1$

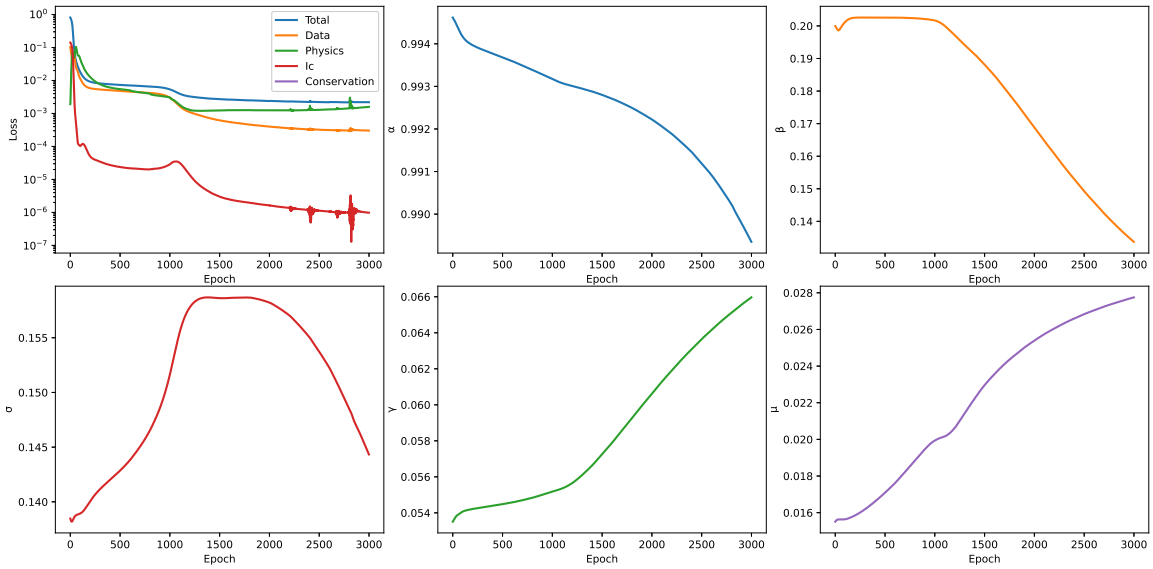


(b) $\alpha = 0.95$

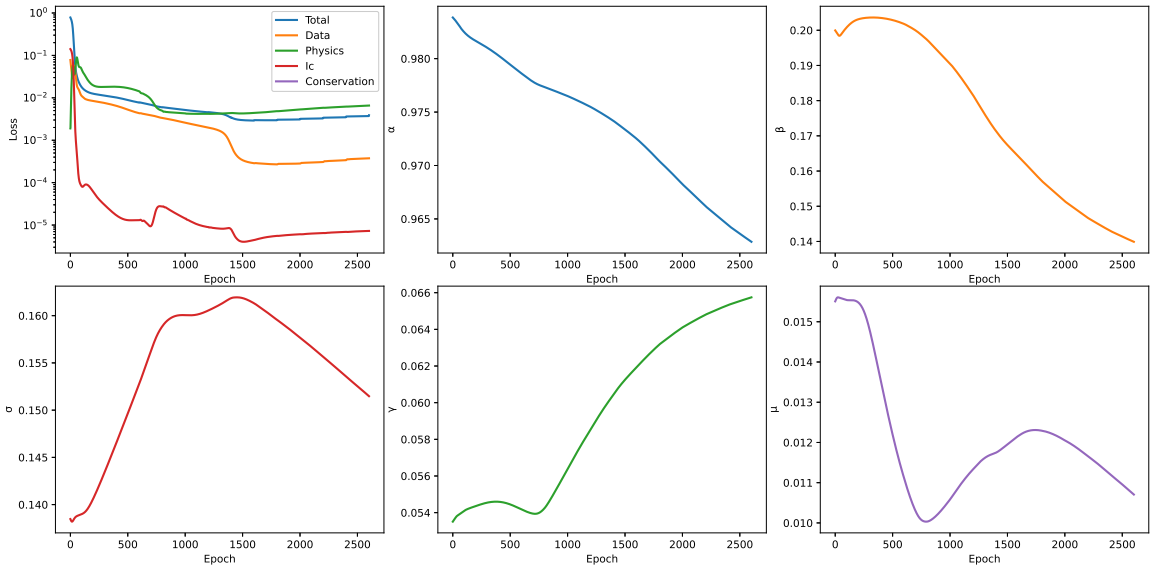


(c) $\alpha = 0.9$

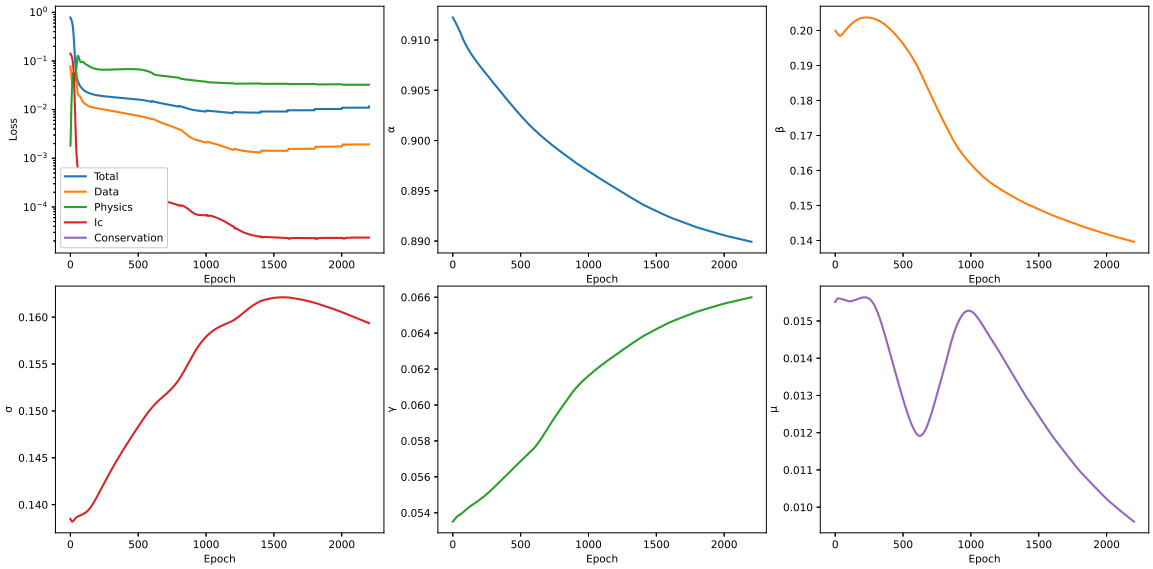
Figure 2: PINN predictions of SEIRD dynamics under different fixed values of α.



(a) $\alpha = 1$



(b) $\alpha = 0.95$



(c) $\alpha = 0.9$

Figure 3: Estimated epidemiological parameters under different fixed α.

The recovered parameters remain within the biologically relevant ranges listed in Table 1, demonstrating good consistency with the ground-truth settings. Notably, as α decreases, σ increases slightly, while μ decreases. This trend aligns with the idea that fractional dynamics introduce memory effects that slow down epidemic progression while effectively reducing its severity. Meanwhile, β and γ remain relatively stable across all cases.

Table 2: Estimated epidemiological parameters for Mpox with α fixed at 1, 0.95, and 0.9.

Parameter	α	β	σ	γ	μ
$\alpha = 1$	0.989349	0.133711	0.144335	0.065966	0.027749
$\alpha = 0.95$	0.962859	0.139901	0.151477	0.065746	0.010705
$\alpha = 0.9$	0.914439	0.136780	0.155774	0.066356	0.008520

5.2. Application to Real-World COVID-19 Data

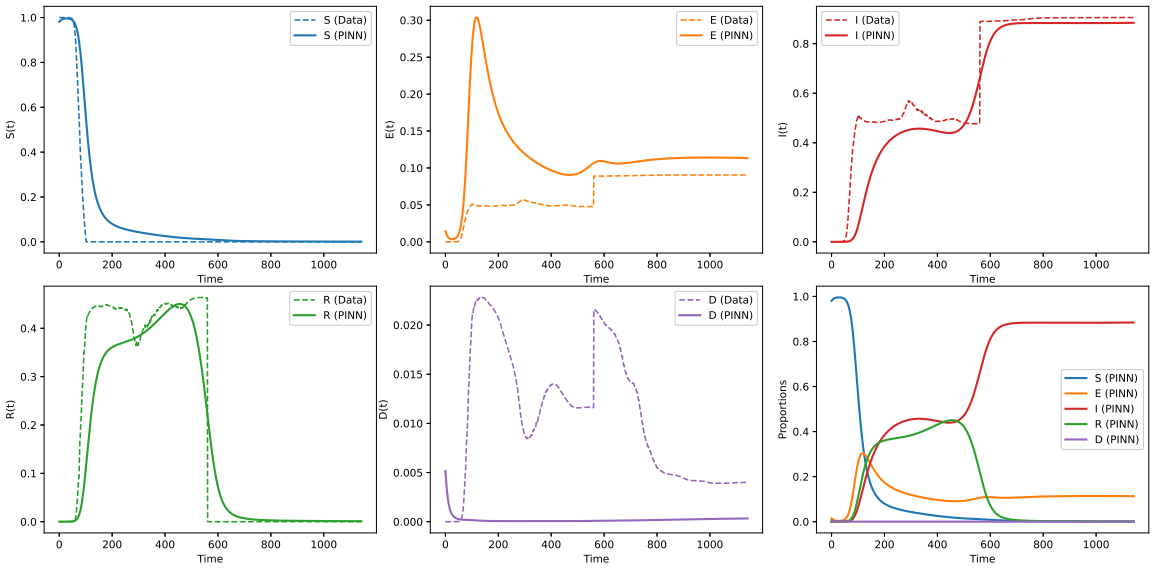
To further assess the applicability of the proposed framework further, we examined real-world epidemic data from Germany and Sweden during the COVID-19 outbreaks. We selected these two countries due to the availability of their daily case and mortality reports and their contrasting epidemic trajectories within Europe. The corresponding epidemiological parameters were constrained within biologically plausible ranges, as shown in Table 3.

Table 3: Epidemiological parameters for the SEIRD model applied to COVID-19 in Europe (2024).

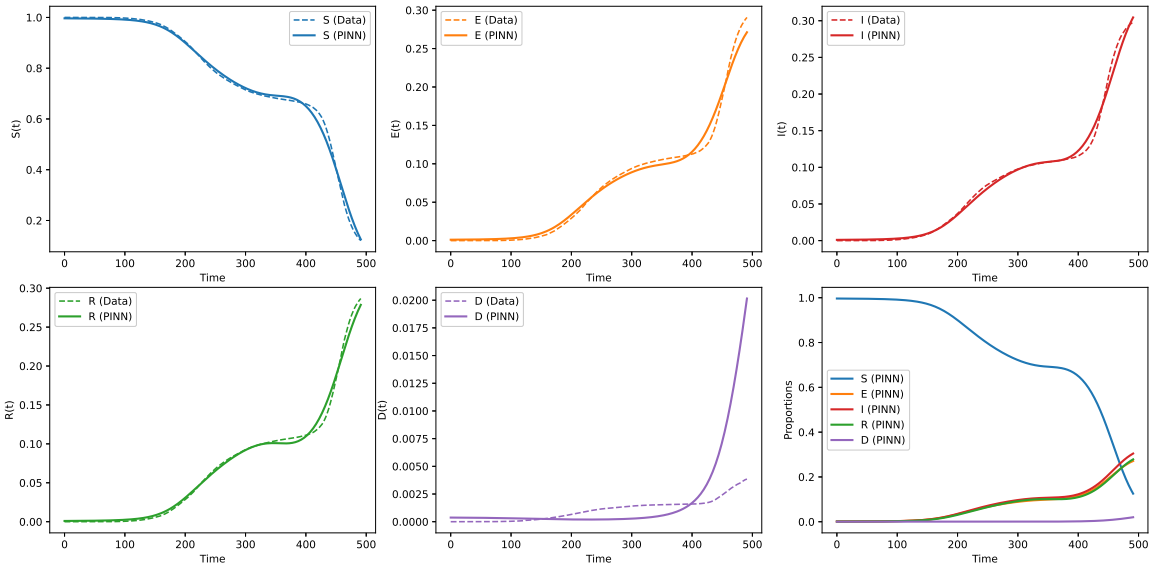
Epidemic	β	σ	γ	μ	**References**
COVID-19	0.2 – 0.4	0.1 – 0.3	0.05 – 0.1	0.001 – 0.01	[52, 53]

The daily number of confirmed cases and deaths were extracted for each country from publicly available databases [54–57]. Cumulative counts were transformed into compartmental trajectories consistent with the SEIRD structure. Specifically, the infected class I was reconstructed from active case estimates, the deceased class D from reported daily deaths, and the removed class R by aggregating recoveries and deaths, when recovery data were available. Since the exposed population E was not directly observed, it was treated as a latent state to be inferred by the model.

Figures 4 and 5 present the application of the proposed fractional-SEIRD PINN framework to real-world COVID-19 data from Germany and Sweden. Figure 4 compares the reconstructed epidemic dynamics with the reported data for both countries. The trajectories of the susceptible, exposed, infected, recovered, or deceased populations show close agreement with the observations. This indicates that the PINN can capture the underlying transmission dynamics despite data noise and reporting variability. Notably, the German outbreak exhibits a sharper infection peak and a faster decline, while the Swedish epidemic has a more prolonged infectious curve. These differences are consistent with the different public health interventions implemented in the two countries.



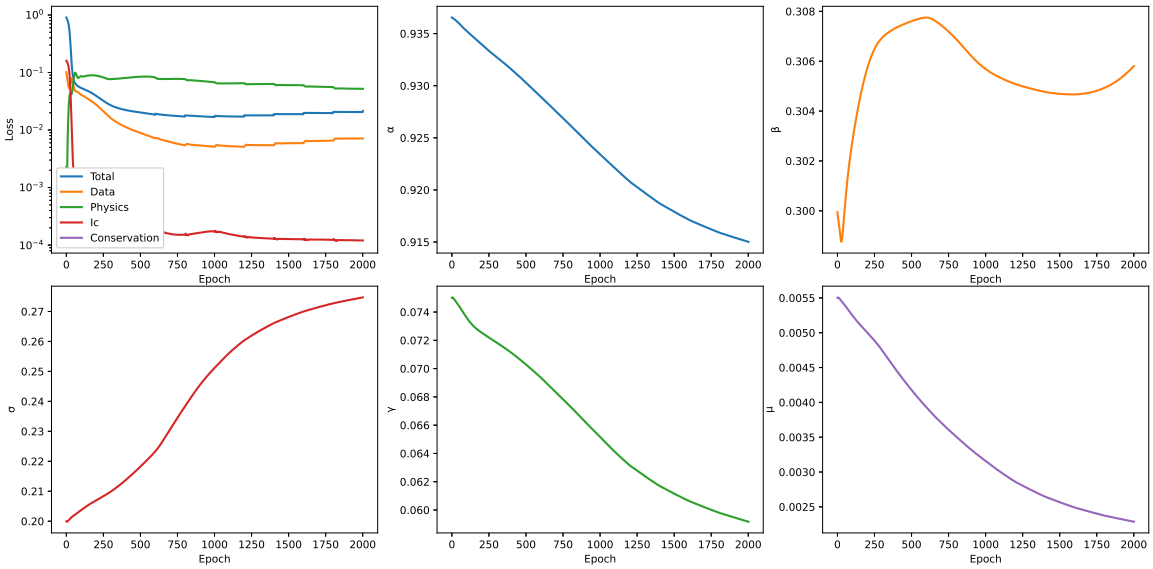
(a) Germany data



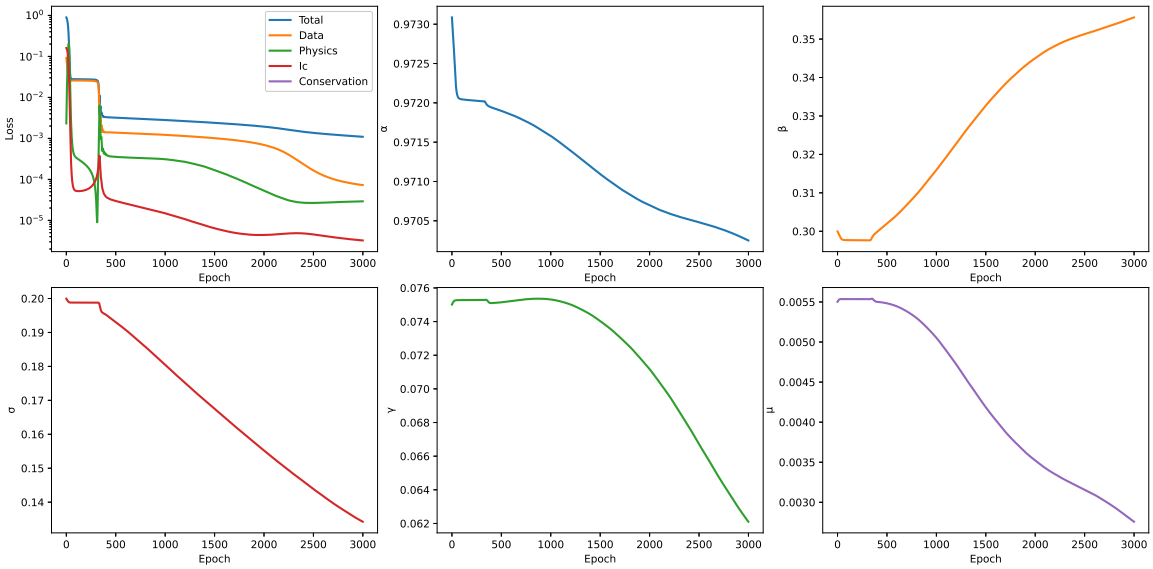
(b) Sweden data

Figure 4: PINN predictions of SEIRD dynamics for real data.

Figure 5 illustrates the corresponding parameter estimation outcomes. The recovered values remain within the admissible ranges provided in Table 3, confirming both the validity of the training process and the biological plausibility of the estimates. The fractional order α was consistently identified as less than one for both Germany and Sweden, reflecting the presence of long-memory effects in the epidemic data. Additionally, subtle differences emerge across the two case studies: the Swedish data suggest a slightly lower α, accompanied by higher incubation and recovery rates, whereas the German data yield higher transmission and mortality rates. These contrasts demonstrate how fractional dynamics can reveal country-specific heterogeneity in epidemic progression that standard integer-order models may miss.



(a) Germany data



(b) Sweden data

Figure 5: Estimated epidemiological parameters for real data.

Table 4 summarizes the estimated parameters for Germany and Sweden. All recovered values fall within the biologically plausible ranges reported in Table 3, confirming the reliability of the framework when applied to real data.

Table 4: Estimated epidemiological parameters for COVID-19.

Parameter	α	β	σ	γ	μ
Germany	0.915016	0.305806	0.274704	0.059172	0.002286
Sweden	0.970250	0.355629	0.134239	0.062103	0.002757

Overall, the results confirm that the proposed PINN framework generalizes well to real-world COVID-19 data, providing both accurate trajectory reconstruction and interpretable parameter estimates. The identification of $\alpha < 1$ reinforces the epidemiological relevance of fractional-order models in accounting for memory effects such as delayed reporting, behavioral adaptation, and variability in disease progression.

5.3. Limitations of the PINN Training

Although the proposed fractional-SEIRD PINN framework yields promising results on synthetic and real datasets, it has several limitations.

First, the framework relies on the availability and quality of epidemiological data. For real-world COVID-19 datasets, reporting delays, under-reporting, and missing recovery information can introduce uncertainties that affect the accuracy of reconstructed trajectories.

Estimating the susceptible and exposed compartments is inherently challenging because they are usually unobserved and must be inferred indirectly. While the PINN provides a principled way to address this issue, further refinements could enhance robustness.

These limitations do not undermine the validity of the present results. Rather, they highlight areas for methodological improvement.

6. Conclusion

In this work, we developed a PINN framework for the fractional SEIRD epidemic model. Our focus was on estimating the fractional memory order α alongside classical epidemiological parameters. Leveraging the Caputo fractional derivative and the L1 discretization scheme, our method incorporates memory effects into epidemic dynamics, and captures long-range dependencies that are absent in classical integer-order formulations. Validation on Mpox synthetic datasets demonstrated that the proposed framework can accurately recover α and epidemiological parameters, even in the presence of observational noise. Applications to real datasets further confirmed that $\alpha < 1$, supporting the hypothesis that epidemic spreading processes exhibit non-Markovian behavior. Let us note that a fractional derivative with order $\alpha < 1$ implies that system dynamics depend on the entire history of the process, not just the present state. This nonlocality captures the heavy tails of the incubation and infectious periods, reflecting empirical evidence that epidemic processes exhibit memory effects that are inconsistent with purely Markovian (exponential waiting time) assumptions. A comparative analysis against the classical SEIRD model revealed clear improvements in predictive accuracy and interpretability when memory was incorporated. Overall, our findings suggest that the fractional SEIRD model coupled with PINNs is a powerful tool for epidemic analysis, offering improved predictions and deeper insights into the underlying dynamics of infectious diseases.

Figure 6 illustrates the overall workflow of the proposed approach. Both synthetic and real datasets serve as inputs to the fractional SEIRD model, which uses a PINN architecture to maintain dataset fidelity and ensure physical consistency. The model outputs compartmental trajectories, from which key quantities such as the fractional order α and epidemiological parameters are inferred. The schematic also illustrates the importance of conservation constraints and uncertainty analysis inensuring, that the learned parameters are interpretable and reliable. The diagram emphasizes the integration of data-driven learning and fractional epidemic modeling into a unified framework.

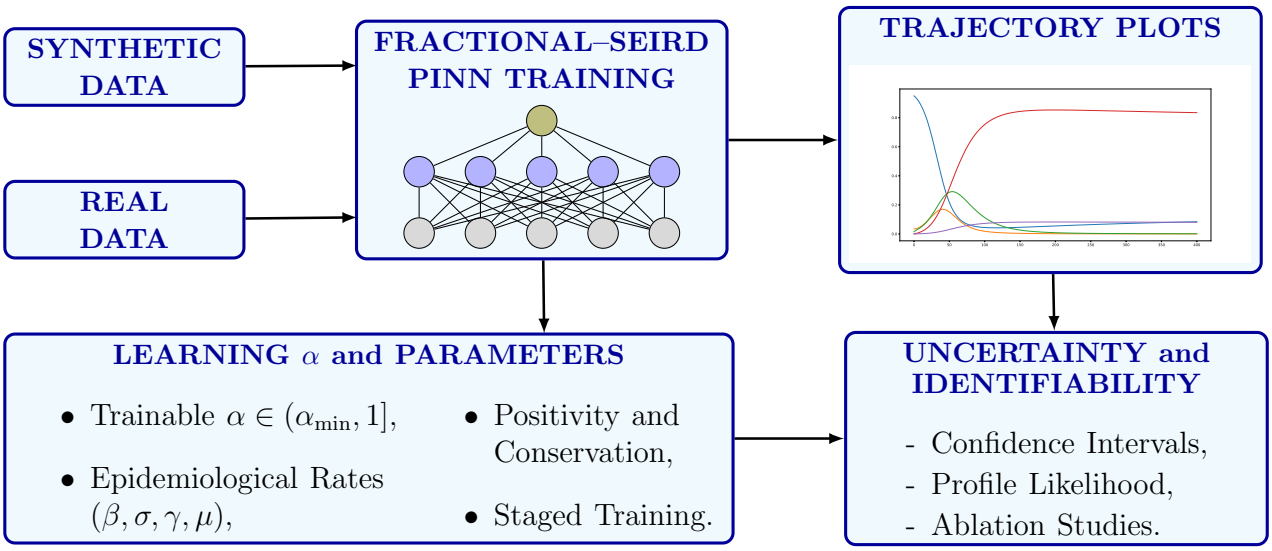


Figure 6: Workflow of PINNs in the Fractional SEIRD Model.

This study reveals several promising directions. One natural extension is introducing a time-varying fractional order $\alpha(t)$, which would enable the model to capture evolving memory effects during different epidemic phases, such as periods of intervention or changes in population behavior. Another important avenue is integrating multi-region and network-based epidemic models. Spatial coupling and mobility data could be incorporated to infer α across interconnected regions, improving large-scale epidemic forecasting. Developing spatiotemporal extensions of the fractional SEIRD model within the PINN framework would be particularly appealing in this context because it would enable the simultaneous representation of memory effects, temporal evolution, and spatial heterogeneity in epidemic dynamics.

In terms of uncertainty quantification, shifting from bootstrap-based methods to Bayesian formulations, such as Bayesian neural operators or variational inference, could yield more reliable posterior distributions for α and epidemiological parameters. Improving scalability and efficiency is essential on the computational side, especially by designing fast algorithms for fractional derivative evaluation, such as reduced-memory kernels or GPU-accelerated schemes, to make the framework practical for real-time epidemic monitoring. Finally, applying the approach to other infectious diseases, such as seasonal influenza, dengue, and vector-borne epidemics, in addition to Mpox and COVID-19, would demonstrate its generality and highlight the role of long-memory effects in different epidemiological contexts.

In summary, this work establishes a foundation for integrating fractional dynamics with PINNs in epidemic modeling. By demonstrating the identifiability and usefulness of the fractional memory order α, it paves the way for richer, spatiotemporal, memory-aware epidemic models that can improve scientific understanding and public health decision-making.

Acknowledgments

The author would like to express sincere thanks to *Dr. Rachid Benoudi*, *Prof. Dr. Matthias Ehrhardt*, and *Prof. Dr. Moulay Rchid Sidi Ammi* for their valuable feedback and guidance. Furthermore, the author is grateful for the hospitality of the Applied and Computational Mathematics group at the University of Wuppertal, where most of this study's work was carried out.

Declarations

Data Availability

This study uses both synthetic and real epidemic datasets. Real datasets were obtained from the following publicly available source: [54–57].

Conflict of Interest

The author declares that no conflict of interest exists that could have influenced the study presented in this paper.

References

[1] Encyclopaedia Britannica. Black death; https://www.britannica.com/event/Black-Death, Accessed on: August 18, 2025.

[2] Centers for Disease Control and Prevention. History of smallpox; https://www.cdc.gov/smallpox/about/history.html, Accessed on: August 18, 2025.

[3] D. Hu, B. Liu, L. Feng, P. Ding, X. Guo, M. Wang, B. Cao, P. R. Reeves, and L. Wang. Origins of the current seventh cholera pandemic. *Proceedings of the National Academy of Sciences*, 113(48), 2016.

[4] S. M. Zimmer and D. S. Burke. Historical perspective - Emergence of influenza A (H1N1) viruses. *New England Journal of Medicine*, 361(3):279–285, 2009.

[5] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. A Koopman operator framework for nonlinear epidemic dynamics: Application to an SIRSD model, 2025. arXiv preprint 2508.16766.

[6] World Health Organization. Ebola outbreak 2014-2016 - West Africa; https://www.who.int/emergencies/situations/ebola-outbreak-2014-2016-West-Africa, Accessed on: August 19, 2025.

[7] Centers for Disease Control and Prevention. Coronavirus disease 2019 (COVID-19); https://www.cdc.gov/covid/index.html, Accessed on: August 19, 2025.

[8] R. Balderrama, M. I. Prieto, C. Sánchez de la Vega, and F. Vázquez. Optimal control for an SIR model with limited hospitalised patients. *Mathematical Biosciences*, 378:109317, 2024.

[9] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. Spatiotemporal SEIQR epidemic mdeling with optimal control for vaccination, treatment, and social measures, 2025. arXiv preprint 2507.09328.

[10] Q. Sun, T. Miyoshi, and S. Richard. Analysis of COVID-19 in Japan with extended SEIR model and ensemble Kalman filter. *Journal of Computational and Applied Mathematics*, 419:114772, 2023.

[11] S. Ghersheen, V. Kozlov, V. G. Tkachev, and U. Wennergren. Dynamical behaviour of SIR model with coinfection: The case of finite carrying capacity. *Mathematical Methods in the Applied Sciences*, 42(17):5805–5826, 2019.

[12] X. Duan, S. Yuan, Z. Qiu, and J. Ma. Global stability of an SVEIR epidemic model with ages of vaccination and latency. *Computers & Mathematics with Applications*, 68(3):288–308, 2014.

[13] A. Zinihi, M. Ehrhardt, and M. R. Sidi Ammi. Actuarial analysis of an infectious disease insurance based on an SEIARD epidemiological model, 2025. arXiv preprint 2508.06580.

[14] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[15] S. Adak, S. Barman, S. Jana, S. Majee, and T. K. Kar. Modelling and analysis of a fractional-order epidemic model incorporating genetic algorithm-based optimization. *Journal of Applied Mathematics and Computing*, 71(1):901–925, 2024.

[16] I. K. Adu, F. A. Wireko, S. A. Adarkwa, and G. O. Agyekum. Mathematical analysis of Ebola considering transmission at treatment centres and survivor relapse using fractal-fractional Caputo derivatives in Uganda. *Mathematical Modelling and Numerical Simulation with Applications*, 4(3):296–334, 2024.

[17] A. Zinihi, M. R. Sidi Ammi, and D. F. M. Torres. Fractional differential equations of a reaction-diffusion SIR model involving the Caputo-fractional time-derivative and a nonlinear diffusion operator. *Evolution Equations and Control Theory*, 14(5):944–967, 2025.

[18] L. C. Barros, M. M. Lopes, F. S. Pedro, E. Esmi, J. P. C. Santos, and D. E. Sánchez. The memory effect on fractional calculus: An application in the spread of COVID-19. *Computational and Applied Mathematics*, 40(3), 2021.

[19] K. Diethelm. *The analysis of fractional differential equations: An application-oriented exposition using differential operators of Caputo type*. Springer Berlin Heidelberg, 2010.

[20] S. Barman, S. Jana, S. Majee, S. Patra, and T. K. Kar. Impact of vaccination and media on a Caputo derivative-based fractional-order epidemic model with PRCC analysis. *International Journal of Dynamics and Control*, 13(9), 2025.

[21] A. Zinihi, M. R. Sidi Ammi, and M. Ehrhardt. Mathematical modeling and Hyers-Ulam stability for a nonlinear epidemiological model with Φ_p operator and Mittag-Leffler kernel. *Advances in Applied Mathematics and Mechanics*, 17(5):1481–1508, 2025.

[22] M. Almuzini, F. A. Abdullah, M. O. Adewole, S. Momani, and S. A. Khuri. Generalized mathematical model of infectious disease during medicinal intervention by employing fractional differential equations. *Mathematical Methods in the Applied Sciences*, 48(12):12153–12173, 2025.

[23] J. P. S. Maurício de Carvalho and B. Moreira-Pinto. A fractional-order model for COVID-19 dynamics with reinfection and the importance of quarantine. *Chaos, Solitons & Fractals*, 151:111275, 2021.

[24] L. Li and J. G. Liu. A generalized definition of Caputo derivatives and its application to fractional ODEs. *SIAM Journal on Mathematical Analysis*, 50(3):2867–2900, 2018.

[25] Z. Feng and Y. Su. Ground state solutions of fractional equations with Coulomb potential and critical exponent. *Studies in Applied Mathematics*, 153(2), 2024.

[26] A. Maione, D. Mugnai, and E. Vecchi. Variational methods for nonpositive mixed local-nonlocal operators. *Fractional Calculus and Applied Analysis*, 26(3):943–961, 2023.

[27] Z. W. Fang, H. W. Sun, and H. Wang. A fast method for variable-order Caputo fractional derivative with applications to time-fractional diffusion equations. *Computers & Mathematics with Applications*, 80(5):1443–1458, 2020.

[28] B. K. Lenka. New stability theory for non-autonomous fractional time-varying order derivative systems and its applications. *Chaos, Solitons & Fractals*, 200:116935, 2025.

[29] M. Verma, A. Priyadarshi, and S. Verma. Analytical and dimensional properties of fractal interpolation functions on the Sierpiński gasket. *Fractional Calculus and Applied Analysis*, 26(3):1294–1325, 2023.

[30] M. Caputo and C. Cametti. Fractional derivatives in the diffusion process in heterogeneous systems: The case of transdermal patches. *Mathematical Biosciences*, 291:38–45, 2017.

[31] R. N. Pandey and H. M. Srivastava. Fractional calculus and its applications involving certain classes of functional relations. *Studies in Applied Mathematics*, 89(2):153–165, 1993.

[32] A. D. U. Antony and A. J. Gnanaprakasam. Development of a unified framework for fractal-fractional derivatives and its application to the modified Gompertz growth equation using the Sawi transform. *Journal of Computational and Applied Mathematics*, 475:117028, 2026.

[33] M. Khalid and S. Alha. New generalized integral transform on Hilfer-Prabhakar fractional derivatives and its applications. *International Journal of Dynamics and Control*, 12(1):24–31, 2023.

[34] A. Zinihi, M. R. Sidi Ammi, and M. Ehrhardt. Optimal control of a diffusive epidemiological model involving the Caputo-Fabrizio fractional time-derivative. *Partial Differential Equations in Applied Mathematics*, 14:101188, 2025.

[35] K. Diethelm. A fractional calculus based model for the simulation of an outbreak of Dengue fever. *Nonlinear Dynamics*, 71(4):613–619, 2012.

[36] P. A. Naik, K. M. Owolabi, J. Zu, and M. U. D. Naik. Modeling the transmission dynamics of COVID-19 pandemic in Caputo type fractional derivative. *Journal of Multiscale Modelling*, 12(03), 2021.

[37] P. Kumar and V. Suat Erturk. A case study of COVID-19 epidemic in India via new generalised Caputo type fractional derivatives. *Mathematical Methods in the Applied Sciences*, 46(7):7930–7943, 2021.

[38] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[39] F. Heldmann, S. Berkhahn, M. Ehrhardt, and K. Klamroth. PINN training using biobjective optimization: The trade-off between data loss and residual loss. *Journal of Computational Physics*, 488:112211, 2023.

[40] S. Berkhahn and M. Ehrhardt. A physics-informed neural network to model COVID-19 infection and hospitalization scenarios. *Advances in Continuous and Discrete Models*, 2022(1):61, 2022.

[41] S. Han, L. Stelz, H. Stoecker, L. Wang, and K. Zhou. Approaching epidemiological dynamics of COVID-19 with physics-informed neural networks. *Journal of the Franklin Institute*, 361(6):106671, 2024.

[42] C. Millevoi, D. Pasetto, and M. Ferronato. A physics-informed neural network approach for compartmental epidemiological models. *PLOS Computational Biology*, 20(9):e1012387, 2024.

[43] Y. Hwang, H. D. Kwon, and J. Lee. Optimal control problem of various epidemic models with uncertainty based on deep reinforcement learning. *Numerical Methods for Partial Differential Equations*, 38(6):2142–2162, 2022.

[44] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo. *Theory and applications of fractional differential equations*, volume 204. Elsevier, 2006.

[45] Z. M. Odibat and N. T. Shawagfeh. Generalized Taylor's formula. *Applied Mathematics and Computation*, 186(1):286–293, 2007.

[46] G. Gao, Z. Sun, and H. Zhang. A new fractional numerical differentiation formula to approximate the Caputo fractional derivative and its applications. *Journal of Computational Physics*, 259:33–50, 2014.

[47] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[48] L. N. Smith. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, page 464–472. IEEE, 2017.

[49] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989.

[50] World Health Organization. Mpox: Multi-country external situation report no.44; https://www.who.int/docs/default-source/coronaviruse/situation-reports/20241223_mpox-external-sitrep_-44.pdf, Accessed on: September 15, 2025.

[51] The Lancet Global Health. Modelling vaccination approaches for mpox containment; https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(24)00384-X/fulltext, Accessed on: September 15, 2025.

[52] UK Health Security Agency. Epidemiology of COVID-19 in England: January 2020 to December 2024; https://www.gov.uk/government/publications/epidemiology-of-covid-19-in-england/epidemiology-of-covid-19-in-england-january-2020-to-december-2024, Accessed on: September 20, 2025.

[53] World Health Organization. COVID-19 epidemiological update, 24 December 2024; https://www.who.int/publications/m/item/covid-19-epidemiological-update---24-december-2024, Accessed on: September 20, 2025.

[54] Our World in Data. Estimated cumulative excess deaths during COVID-19, Jan 1, 2020 to Aug 24, 2025; https://ourworldindata.org/explorers/covid?country=USA~JPN~DEU~SWE&Metric=Excess+mortality+%28estimates%29&Interval=Cumulative&Relative+to+population=false, Accessed on: September 23, 2025.

[55] J. P. Gehrcke. COVID-19 case numbers for Germany; https://github.com/jgehrcke/covid-19-germany-gae?utm_source, Accessed on: September 23, 2025.

[56] Our World in Data. Sweden: Coronavirus pandemic country profile; https://ourworldindata.org/coronavirus/country/sweden?utm_source, Accessed on: September 24, 2025.

[57] Worldometers.info. Sweden COVID-Coronavirus statistics; https://www.worldometers.info/coronavirus/country/sweden/, Accessed on: September 24, 2025.