LEARNING EPIDEMIOLOGICAL DYNAMICS VIA THE FINITE EXPRESSION METHOD

Jianda Du,[1] Senwei Liang,[2] & Chunmei Wang[1,*]

[1]*Department of Mathematics, University of Florida, Gainesville, Florida 32611, USA*

[2]*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

*Address all correspondence to: Chunmei Wang, Department of Mathematics, University of Florida, Gainesville, Florida 32611, USA, E-mail: chunmei.wang@ufl.edu

Modeling and forecasting the spread of infectious diseases is essential for effective public health decision-making. Traditional epidemiological models rely on expert-defined frameworks to describe complex dynamics, while neural networks (NNs), despite their predictive power, often lack interpretability due to their "black-box" nature. This paper introduces the finite expression method (FEX), a symbolic learning framework that leverages reinforcement learning (RL) to derive explicit mathematical expressions for epidemiological dynamics. Through numerical experiments on both synthetic and real-world datasets, FEX demonstrates high accuracy in modeling and predicting disease spread, while uncovering explicit relationships among epidemiological variables. These results highlight FEX as a powerful tool for infectious disease modeling, combining interpretability with strong predictive performance to support practical applications in public health.

KEY WORDS: *symbolic regression, finite expression method, reinforcement learning, epidemiological dynamics, data-driven modeling, interpretability*

1. INTRODUCTION

Partial differential equations (PDEs) and ordinary differential equations (ODEs) are used to describe physical phenomena across diverse scientific and engineering disciplines. In epidemiology, PDEs and ODEs serve as cornerstones for understanding disease dynamics, guiding interventions, and improving strategies to mitigate the impact of infectious diseases (Anderson and May, 1991; Brauer et al., 2019). Classical models such as the susceptible-infected-recovered (SIR) (Hethcote, 2000; Kermack et al., 1927), susceptible-exposed-infected-recovered (SEIR) (Anderson and May, 1991; Kermack et al., 1927), and susceptible-exposed-infected-recovered-deceased (SEIRD) (Brauer et al., 2019) frameworks rely on compartmental approaches, using differential equations to represent transitions between population states with parameters like transmission and recovery rates. These frameworks have underpinned epidemiological research for decades, offering valuable insights into disease spread and control.

However, traditional compartmental models encounter significant limitations in addressing the complexities of real-world scenarios. Incorporating factors such as time-varying transmission rates, spatial heterogeneity, or additional compartments often makes these models analytically intractable and computationally intensive (Keeling and Rohani, 2008). Moreover, their reliance on manual refinements slows response times and hampers adaptability to rapidly evolving conditions (Grassly and Fraser, 2008).

Submitted: 12/30/24; Accepted: 3/9/25; Online: 3/17/25
https://doi/10.1615/JMachLearnModelComput.2025057903
2689–3967/25/$35.00

The emergence of data-driven approaches, particularly those leveraging deep learning (Huang et al., 2024; Liang and Yang, 2022; Liu et al., 2024; Ma et al., 2019), has introduced powerful alternatives for modeling epidemiological dynamics. Neural networks (NNs)-based approaches, such as recurrent neural networks (RNNs) (Hochreiter, 1997), have shown promise in capturing intricate patterns within epidemiological data (Goodfellow, 2016; LeCun et al., 2015; Raissi et al., 2019). These methods enable rapid learning of disease dynamics, facilitating faster predictions and decision-making. However, their "black-box" nature limits interpretability, hindering their utility in understanding the mechanisms driving disease spread (Rudin, 2019). Additionally, the implicit biases in NN optimization often favor smooth functions with rapid frequency decay, restricting their ability to produce highly accurate solutions (Liang et al., 2024).

Symbolic learning has recently emerged as a promising alternative, bridging the gap between the predictive power of machine learning and the interpretability of traditional models. By discovering governing equations directly from data, symbolic learning maintains mathematical rigor while leveraging data-driven insights (Cranmer et al., 2020; Schmidt and Lipson, 2009). The finite expression method (FEX) (Liang and Yang, 2022), marks a significant advancement in symbolic learning for high-dimensional problems. FEX formulates the task of identifying mathematical expressions as a combinatorial optimization (CO) problem and leverages reinforcement learning (RL) to solve it. This approach automates the discovery of governing equations, drastically reducing development time while preserving interpretability and physical consistency. By producing parsimonious mathematical expressions, FEX becomes a powerful tool for reliable and efficient epidemiological modeling, particularly in addressing emerging public health crises.

This paper investigates the application of the FEX for learning epidemiological models from both synthetic and real-world data. We demonstrate FEX's potential to address critical limitations of traditional and NN-based approaches. Specifically, our contributions include the following:

- *Advantages over NN-based methods*. We compare the FEX with NN-based approaches, including RNNs, across three classical epidemiological models—SIR, SEIR, and SEIRD. FEX not only achieves competitive predictive performance but also derives explicit governing equations, offering superior interpretability.
- *Advantages over traditional modeling approaches*. Using real-world COVID-19 data, we compare FEX with the fractional-order SEIQRDP model (Bahloul et al., 2020). FEX demonstrates its versatility in handling complex epidemiological dynamics, providing actionable insights and enabling rapid model development.

The remainder of this paper is organized as follows. Section 2 introduces the foundational concepts and procedural framework of the FEX, supplemented by a detailed flowchart. Section 3 describes the loss function and optimization strategy used by FEX to derive governing equations for synthetic and real-world epidemiological data. Section 4 presents experimental results, benchmarking FEX against both NN-based and traditional methods. Finally, Section 5 discusses the broader implications of our findings, acknowledges current limitations, and outlines directions for future research on extending FEX to other dynamical systems.

2. THE FINITE EXPRESSION METHOD

The FEX (Jiang et al., 2023; Liang and Yang, 2022; Song et al., 2024) provides a versatile framework for identifying governing equations of dynamical systems. It explores a function

space composed of finite mathematical expressions constructed from a predefined set of operators. These expressions are represented as binary trees, denoted by $\mathcal{T}$ (Fig. 1), where each node corresponds to an operator, forming an operator sequence $\mathbf{e}$. Each operator is associated with trainable scaling and bias parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, collectively denoted as $\boldsymbol{\theta}$. This setup enables the representation of a FEX as $f(\boldsymbol{x}; \mathcal{T}, \mathbf{e}, \boldsymbol{\theta})$. The FEX framework seeks to identify governing equations by minimizing a functional $\mathcal{L}$ (e.g., derived from ODEs or PDEs). Formally, this optimization problem is expressed as

$$\min\{\mathcal{L}(f(\cdot; \mathcal{T}, \boldsymbol{e}, \boldsymbol{\theta}))|\boldsymbol{e}, \boldsymbol{\theta}\}.$$

To solve this CO problem, FEX employs a search loop powered by RL, as illustrated in Fig. 2(a). The search process comprises four key components:

Score computation (rewards in RL). The suitability of each operator sequence $\mathbf{e}$ is assessed via a score, $S(\boldsymbol{e})$, which is defined as

$$S(\boldsymbol{e}) := \big(1 + L(\boldsymbol{e})\big)^{-1},$$

where $L(\boldsymbol{e}) := \min\{\mathcal{L}(f(\cdot; \mathcal{T}, \boldsymbol{e}, \boldsymbol{\theta}))|\boldsymbol{\theta}\}$. A smaller $L(\boldsymbol{e})$ indicates better approximation of the governing dynamics, corresponding to a higher score.

To address the computational challenges of minimizing $\mathcal{L}$ globally, FEX employs a two-stage optimization strategy. First, a first-order algorithm runs for T_1 iterations to provide an initial parameter estimate. Subsequently, a second-order algorithm, such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Fletcher, 2000), refines these parameters over T_2 iterations. If $\boldsymbol{\theta}_0^{\boldsymbol{e}}$ represents the initial parameters and $\boldsymbol{\theta}_{T_1+T_2}^{\boldsymbol{e}}$ the refined ones, the score is approximated as

$$S(\boldsymbol{e}) \approx \big(1 + \mathcal{L}(f(\cdot; \mathcal{T}, \boldsymbol{e}, \boldsymbol{\theta}_{T_1+T_2}^{\boldsymbol{e}}))\big)^{-1}.$$

Operator sequence generation (actions in RL). The controller, denoted by $\boldsymbol{\chi}_\Phi$, generates operator sequences $\boldsymbol{e}$ [Fig. 2(b)]. The parameters Φ of the controller are updated iteratively to favor high-scoring sequences. Operator sequences are constructed by sampling from probability mass functions $\boldsymbol{p}_\Phi^1, \boldsymbol{p}_\Phi^2, \ldots, \boldsymbol{p}_\Phi^s$, which define the distributions of node values in $\mathcal{T}$. A sequence $\boldsymbol{e} = (e_1, e_2, \ldots, e_s)$ is generated by sampling each e_j from $\boldsymbol{p}_\Phi^j$,

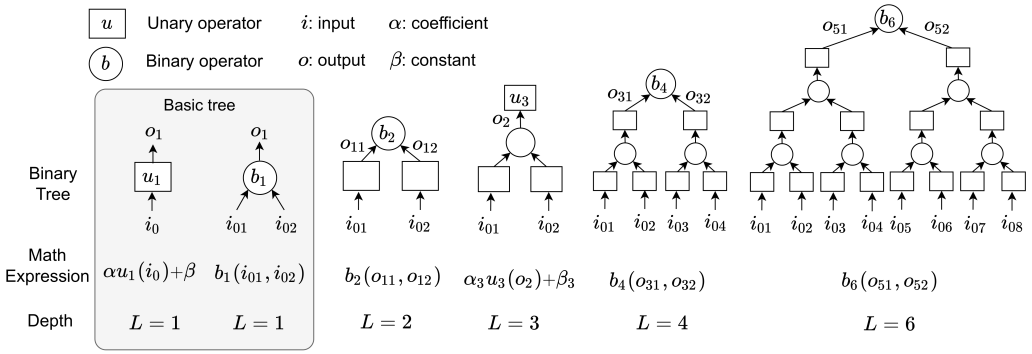


FIG. 1: The computational structure is represented using binary trees, where each node is assigned either a binary or unary operator. Expressions are recursively constructed, starting from depth-1 trees. Binary operators include $\mathbb{B} := \{+, -, \times, \div, \ldots\}$, and unary operators include $\mathbb{U} := \big\{\sin, \exp, \log, \mathrm{Id}, (\cdot)^2, \int \cdot \mathrm{d}x_i, \partial \cdot / \partial x_i, \ldots\big\}$.

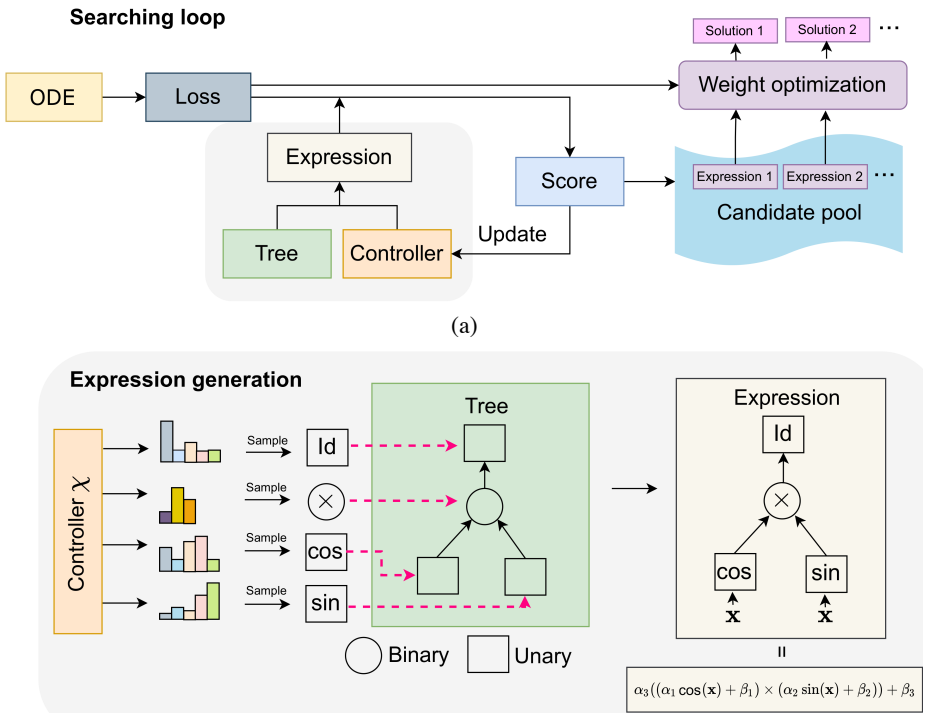


FIG. 2: Flowchart of the FEX. The process consists of an iterative search loop (a), weight optimization (b), and expression generation to identify solutions for the target ODEs or PDEs. Key components include the expression tree, controller, and candidate pool, which collaboratively refine expressions through sampling, scoring, and optimization mechanisms.

To encourage exploration, an ϵ-greedy strategy (Sutton and Barto, 2018) is employed: with probability ϵ, e_i is sampled uniformly from the operator set, while with probability $1-\epsilon$, it is sampled from $\boldsymbol{p}_\Phi^i$. Larger values of ϵ promote broader exploration of the search space.

Controller update (policy optimization in RL). The controller's parameters Φ are updated to increase the likelihood of generating high-performing sequences. Using a policy gradient approach (Petersen et al., 2019; Ruder, 2016), the objective is to maximize,

$$\mathcal{J}(\Phi) = \mathbb{E}_{\boldsymbol{e}\sim\boldsymbol{\chi}_\Phi}\{S(\boldsymbol{e})|S(\boldsymbol{e}) \geq S_{\nu,\Phi}\},$$

where $S_{\nu,\Phi}$ is the $(1-\nu)\times 100\%$-quantile of the score distribution produced by $\boldsymbol{\chi}_\Phi$ in a batch. This prioritizes top-performing sequences. Gradient ascent is used to update

$$\Phi \leftarrow \Phi + \eta\nabla_\Phi \mathcal{J}(\Phi).$$

Candidate optimization (policy deployment). A candidate pool $\mathbb{P}$, with a fixed capacity K, stores the highest-scoring operator sequences discovered during the search. Additional optimization is performed for each sequence $\boldsymbol{e} \in \mathbb{P}$, using a first-order algorithm over T_3 iterations with a smaller learning rate. This reduces the risk of missing promising solutions due to local minima encountered in the initial search.

3. LEARNING EPIDEMIOLOGICAL DYNAMICS VIA FEX

3.1 Epidemiological Models

The dynamics of an epidemiological system are governed by

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}),$$

where $\boldsymbol{x} \in \mathbb{R}^d$ denotes the state variables (e.g., susceptible, infected, and recovered populations), and $f(\boldsymbol{x})$ encapsulates the system's temporal evolution.

To evaluate the effectiveness of the FEX, we apply it to three commonly used epidemiological models:

1. *SIR model*. The SIR model (Hethcote, 2000; Kermack et al., 1927) divides the total population N into three compartments: susceptible (S), infectious (I), and recovered (R). This model is suitable for diseases with long-term immunity postrecovery, such as measles or chickenpox. The governing equations are

$$\frac{dS}{dt} = \mu(N - S) - \frac{\beta SI}{N}, \qquad \frac{dI}{dt} = \frac{\beta SI}{N} - (\mu + \gamma)I, \qquad \frac{dR}{dt} = \gamma I - \mu R, \quad (3.1)$$

   where β, γ, and μ are the transmission, recovery, and natural death rates, respectively.

2. *SEIR model*. The SEIR model (Anderson and May, 1991; Kermack et al., 1927; Su et al., 2021) extends the SIR framework by introducing an exposed (E) compartment to account for a latent, noninfectious period. Additionally, vaccination effects are incorporated through a νS term. The dynamics are

$$\begin{aligned} \frac{dS}{dt} &= \mu(N - S) - \frac{\beta SI}{N} - \nu S, \qquad \frac{dE}{dt} = \frac{\beta SI}{N} - (\mu + \sigma)E \\ \frac{dI}{dt} &= \sigma E - (\mu + \gamma)I, \qquad \frac{dR}{dt} = \gamma I - \mu R + \nu S, \end{aligned} \quad (3.2)$$

   where σ is the rate of progression from exposed to infectious, and ν is the rate of immunity acquisition.

3. *SEIRD model*. The SEIRD model (Brauer et al., 2019) further extends the SEIR by including a deceased (D) compartment to account for disease-induced mortality. The equations are

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N}, \qquad \frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E, \qquad \frac{dI}{dt} = \sigma E - (\gamma + \delta)I \\ \frac{dR}{dt} &= \gamma I, \qquad \frac{dD}{dt} = \delta I, \end{aligned} \quad (3.3)$$

   where δ is the disease-induced death rate. This model is crucial for analyzing diseases with significant mortality.

In all models, the variables are normalized such that their sum equals 1 ($N = 1$).

3.2 Learning Epidemiological Dynamics via FEX

Given historical data $\{\boldsymbol{x}^{s\Delta}\}_{s=0}^{M}$, where s is the step index, M is the total number of steps, and Δ is the time step size, our goal is to construct a surrogate model $\phi(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}^d$ that approximates the true dynamics f. This surrogate model $\phi(\boldsymbol{x})$ aims to replicate the system's behavior and predict future states. Let $\hat{\boldsymbol{x}}^{(s+1)\Delta} = \text{Integrator}(\phi, \boldsymbol{x}^{s\Delta}, \Delta)$ denote the predicted state, where the integrator advances the system by one time step Δ, starting from $\boldsymbol{x}^{s\Delta}$. Accurate replication requires $\hat{\boldsymbol{x}}^{(s+1)\Delta} \approx \boldsymbol{x}^{(s+1)\Delta}$, for $s = 0, \cdots, M-1$.

Training. The surrogate model is trained by minimizing the discrepancy between observed data and model predictions. The objective function is

$$\min_{\phi} \mathcal{L}(\phi) := \frac{1}{M} \sum_{s=0}^{M-1} \left\| \boldsymbol{x}^{(s+1)\Delta} - \hat{\boldsymbol{x}}^{(s+1)\Delta} \right\|_2^2,$$

where ϕ belongs to a specified function space. For NNs, $\phi_{\text{NN}}(\boldsymbol{x}; \theta)$ is parameterized by θ, and the objective becomes

$$\min_{\theta} \mathcal{L}(\phi_{\text{NN}}(\boldsymbol{x}; \theta)). \tag{3.4}$$

This optimization can be performed using methods such as Adam or BFGS.

Forecasting. After training, the optimized surrogate model $\tilde{\phi}$ predicts future states beyond the training period by iteratively propagating the system forward. Starting from $\hat{\boldsymbol{x}}^{M\Delta} = \boldsymbol{x}^{M\Delta}$, future states are computed as

$$\hat{\boldsymbol{x}}^{(s+1)\Delta} = \text{Integrator}(\tilde{\phi}, \hat{\boldsymbol{x}}^{s\Delta}, \Delta), s \geq M.$$

In epidemiological applications, the trained surrogate model can forecast disease spread and evaluate intervention strategies, such as vaccination or social distancing.

FEX methodology. FEX aims to approximate $f(\boldsymbol{x})$ with a surrogate function:

$$\Phi_{\text{FEX}}(\boldsymbol{x}) := \left[\phi_{\text{FEX}}^{(1)}(\boldsymbol{x}), \phi_{\text{FEX}}^{(2)}(\boldsymbol{x}), \cdots, \phi_{\text{FEX}}^{(d)}(\boldsymbol{x}) \right]^{\top} \approx f(\boldsymbol{x}),$$

where $\phi_{\text{FEX}}^{(i)}(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}^1$ for $i = 1, \cdots, d$.

Given historical data $\{\boldsymbol{x}^{s\Delta}\}_{s=0}^{M}$, with $\boldsymbol{x}^{s\Delta} = [x_1^{s\Delta}, \cdots, x_d^{s\Delta}]$, the loss function for each component is

$$\min_{\boldsymbol{e},\theta} \mathcal{L}\left(\phi_{\text{FEX}}^{(i)}(\boldsymbol{x}; \boldsymbol{e}, \theta) \right) := \frac{1}{M} \sum_{s=0}^{M-1} \left\| x_i^{(s+1)\Delta} - \hat{x}_i^{(s+1)\Delta} \right\|_2^2,$$

where $\hat{x}_i^{(s+1)\Delta}$ is predicted with the surrogate model starting from $\boldsymbol{x}^{(s+1)\Delta}$.

Using an Euler scheme, the loss function for $\phi_{\text{FEX}}^{(i)}(\boldsymbol{x})$ becomes

$$\mathcal{L}\left(\phi_{\text{FEX}}^{(i)}(\boldsymbol{x}; \boldsymbol{e}, \theta) \right) := \frac{1}{M} \sum_{s=0}^{M-1} \left\| x_i^{(s+1)\Delta} - x_i^{s\Delta} - \phi_{\text{FEX}}^{(i)}(\boldsymbol{x}^{s\Delta})\Delta \right\|_2^2. \tag{3.5}$$

This loss function allows independent learning of each component of the dynamics using the RL-based optimization approach detailed in Section 2.

4. NUMERICAL RESULTS

In this section, we demonstrate the effectiveness of the FEX in learning epidemiological dynamics using both synthetic and real-world data.

4.1 Synthetic Epidemiological Data

Synthetic data are generated based on the SIR [Eq. (3.1)], SEIR [Eq. (3.2)], and SEIRD [Eq. (3.3)] models. The performance of the FEX is evaluated against NN-based approaches, specifically the NN and RNN methods (Goodfellow, 2016; LeCun et al., 2015; Sutton and Barto, 2018). The NN method uses a NN as a surrogate to approximate the dynamics, while the RNN method leverages a recurrent NN for modeling time-series data. Model performance is quantified using the mean squared error (MSE) at each time step, calculated between the predicted and true trajectories generated from varying initial conditions.

Data Generation. The parameter values in the SIR model [Eq. (3.1)], the SEIR model [Eq. (3.2)], and the SEIRD model [Eq. (3.3)] are set as follows: $\beta = 0.9$, $\gamma = 0.2$, $\mu = 0.3$, $\sigma_1 = 0.6$ (for SEIR), $\sigma_2 = 0.5$ (for SEIRD), $\nu = 0.2$, and $\delta = 0.05$. Data for the three models are generated using Euler's method. For each model, 200 simulated trajectories are generated, each containing $M = 250$ time steps with a time step size of $\Delta = 0.2$. These trajectories are divided evenly into training and testing datasets. Initial conditions for each trajectory are sampled from a uniform distribution $\mathcal{U}(0, 1)$.

4.1.1 Training Procedures

The FEX. Two types of tree structures are considered (see Fig. 3). Type 1 consists of three layers with one binary operator and three unary operators, while Type 2 also has three layers but includes two binary operators and three unary operators. Type 2 is used to approximate differential equations with nonlinear terms, while Type 1 is employed to approximate differential equations without nonlinear terms aiming to achieve the desired form with fewer training iterations. Detailed information of the tree implementation for each of the three models is summarized in Table A1 in Appendix A. The FEX model is trained for 100 epochs with a batch size of 10. A greedy search strategy (Feo and Resende, 1995; Liang and Yang, 2022; Wilt et al., 2010) with a probability of 0.1 is adopted, and the learning rate for optimizing the controller is set to 0.002. The candidate unary operators include 0, 1, x, x^2, x^3, x^4, $\sin(x)$, $\cos(x)$, $\exp(x)$, while the candidate binary operators are $+$, $-$, $\times$. Euler's method is applied as the integrator in the loss function (3.5). The generalization capabilities of identified expressions using the FEX for the three models are showcased in Fig. A1 in Appendix A.

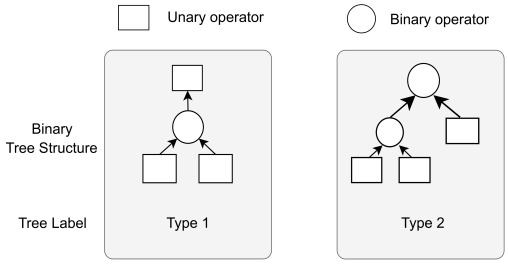


FIG. 3: Illustration of two tree structures used in the FEX implementation

The NN method. For the NN method as described in Eq. (3.4), a network with three linear layers and two ReLU activation layers is employed. The NN model is trained using the Adam optimizer with a learning rate of 0.001 for 100 epochs and a batch size of 32. Euler's method (Burden and Faires, 1997; Shen et al., 2020) is used as the integrator in the loss function (3.5).

The RNN method. The RNN model consists of two long short-term memory (LSTM) layers and a linear output layer for predicting time-series data of the variables in the epidemiological models. The input to the RNN model comprises the variables from the training dataset, where the input dimension corresponds to the number of variables in the specific epidemiology model (e.g., 3 for SIR, 4 for SEIR, and 5 for SEIRD). Each LSTM layer has a hidden size of 51, and the final linear layer maps the 51-dimensional hidden state to an output dimension matching the number of variables in the corresponding model. At the beginning of each forward pass, the hidden and cell states of both LSTM layers are initialized to zero tensors. The model is trained using the limited-memory BFGS (LBFGS) optimizer with the MSE loss function.

Training effectiveness. Figure 4 illustrates the training MSE loss for the various methods applied to different epidemiological models: (a) SIR, (b) SEIR, and (c) SEIRD. From these plots, despite using different numbers of training epochs, the training MSE for all methods converges to a loss value close to zero. This demonstrates the effectiveness of the training process.

Testing. After identifying the expressions that best fit the training data, predicted trajectories are generated by applying Euler's method to the initial state of each testing trajectory. The testing MSE is calculated by comparing the predicted trajectories with the actual testing data.

Numerical Results. Figure 5 presents the testing MSE with 95% confidence interval over time, comparing the performance of FEX with the NN and RNN methods on the SIR, SEIR, and SEIRD models, respectively. The results show that the FEX achieves the smallest MSE on the

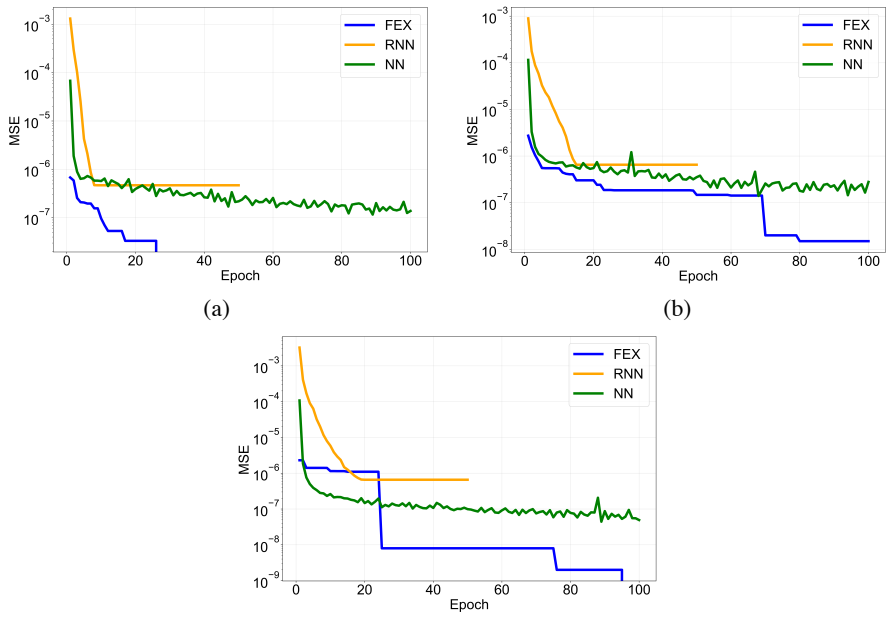


FIG. 4: Comparison of the training loss of three methods for the three models: (a) SIR, (b) SEIR, and (c) SEIRD

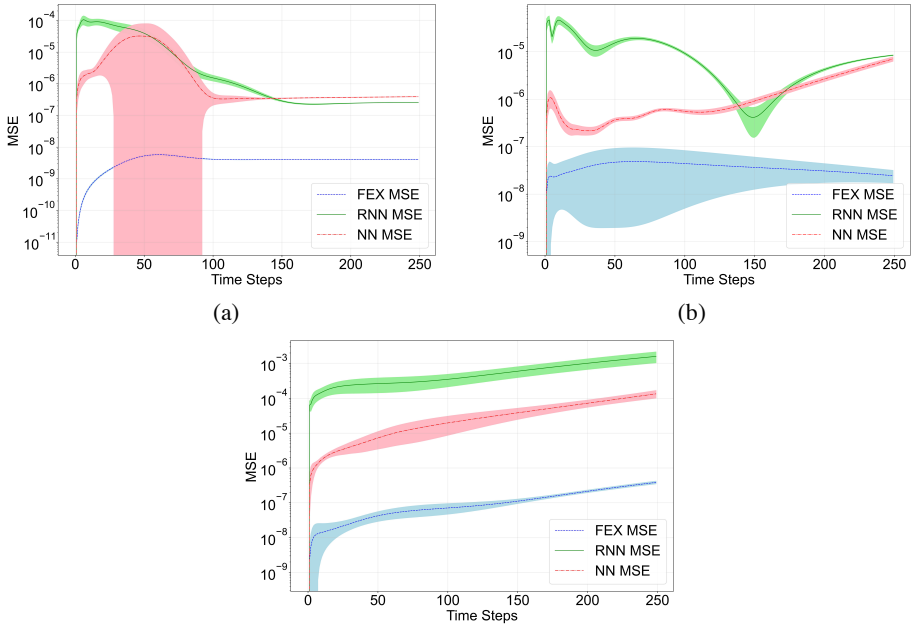


FIG. 5: Comparison of MSE with 95% confidence interval over time for three methods (FEX, RNN, NN) on (a) SIR model, (b) SEIR model, and (c) SEIRD model over 250 time steps

order of 10^{-9}–10^{-7}, significantly outperforming the NN and RNN models, which exhibit MSEs on the order of 10^{-4}–10^{-2} and 10^{-6}–10^{-4}, respectively. Additionally, the FEX maintains consistently small errors over time, whereas the error trajectories predicted by the NN method increase as time progresses. These findings highlight the superior predictive accuracy and robustness of the FEX.

4.2 Real-World Epidemiological Data

To validate the FEX on real-world data, we use the publicly available COVID-19 dataset from Our World in Data (JHU CSSE, 2024), which aggregates detailed global case records from the Johns Hopkins University COVID-19 Data Repository. We compare the FEX to the fractional-order SEIQRDP model (Bahloul et al., 2020), a well-regarded approach for modeling COVID-19 dynamics using fractional-order derivatives to capture complex epidemic patterns.

COVID-19, caused by the highly transmissible SARS-CoV-2 virus, emerged in late 2019 and rapidly escalated into a global pandemic. Understanding its transmission dynamics, influenced by interactions among susceptible, exposed, infected, and recovered individuals, is essential for devising public health interventions. By applying FEX to real-world COVID-19 data, we demonstrate its effectiveness and interpretability compared to established methods.

Data Acquisition. We analyze daily COVID-19 data from Hubei, China, focusing on active cases (Q), deceased cases (D), and recovered cases (R). The dataset spans 1147 days and is publicly accessible in Bahloul et al. (2020). For this study, we use data from the first 100 days (January 22 to April 30, 2020), as values stabilize beyond this period. The dataset is split into training (first 85 days) and testing (remaining 15 days) subsets.

Training. Using the FEX framework, we model COVID-19 dynamics with a Type 2 tree structure (Fig. 3). Training is conducted over 100 epochs with a batch size of 10, employing a greedy search strategy with a 0.1 exploration probability and a controller learning rate of 0.002. For comparison, the fractional-order SEIQRDP method is implemented using the MATLAB code provided in Bahloul et al. (2020). Both methods are trained on the first 85 days of data and evaluated on the subsequent 15 days. However, there are two implementation differences between the two methods:

1. The number of training data: The original SEIQRDP used data between *February 1 and April 1*, 2020, while ours includes more of the earlier data (*January 22–April 15*).
2. Temporal resolution: The original SEIQRDP used $dt = 1/24$ (hourly) while we used $dt = 1$ (daily).

Testing. To assess the FEX model's performance, we use the one-step Euler method (Burden and Faires, 1997) to simulate both training and prediction trajectories:

- Training trajectory ($t = 0$ to $t = 84$): Starting with the ground truth at $t = 0$, the model computes subsequent values using the one-step Euler method and ground truth inputs from the previous time step.
- Prediction trajectory ($t = 85$ to $t = 99$): At $t = 85$, predictions are initiated using the one-step Euler method and the ground truth from $t = 84$. For later steps, predictions are generated iteratively using outputs from the previous time step as inputs.

This approach evaluates the FEX's ability to generalize beyond the training phase to unseen data.

Numerical Results. The FEX identifies the governing dynamics of R, D, and Q during the training phase, effectively capturing interactions among these critical epidemiological variables. The learned equations are

$$\begin{aligned}
\frac{dR}{dt} &= \left(-0.9030R^3 + 2.4025D^3 - 0.0262Q^3 + 0.0311\right) \\
&\quad \times \left(-0.1840R^3 - 0.0432D^3 - 2.5147Q^3 - 0.0181\right) \\
&\quad \times \left(0.1919\sin(R) + 0.1812\sin(D) + 0.7006\sin(Q) - 0.7283\right), \\
\frac{dD}{dt} &= \left(-1.5383e^R + 1.3790e^D - 0.0797e^Q - 0.9186\right) \\
&\quad \times (-0.4292R + 0.9682D - 0.2423Q - 0.2602) \\
&\quad \times (-0.8972R + 1.4067D - 0.2637Q + 0.0054), \\
\frac{dQ}{dt} &= \left(1.6211\sin(R) - 2.1673\sin(D) + 0.6259\sin(Q) - 0.0008\right) \\
&\quad \times \left(4.3940R^3 + 1.6576D^3 + 0.5903Q^3 - 0.6928\right) \\
&\quad \times \left(0.1196R^2 - 3.7194D^2 + 1.7070Q^2 + 1.7746\right).
\end{aligned}$$

Figure 6 compares FEX and the SEIQRDP method (Bahloul et al., 2020) on Q, D, and R. During training, FEX achieves superior data fitting. In the prediction phase, FEX closely aligns with observed data, demonstrating robust generalization. Furthermore, FEX uses only three input variables (Q, R, and D) to derive governing equations, while the SEIQRDP method requires five (E, I, Q, R, and D) in general, with assumptions $E = I$ and $I = Q + R + D$ for current data. This highlights FEX's simplicity and efficiency in modeling complex epidemiological dynamics.

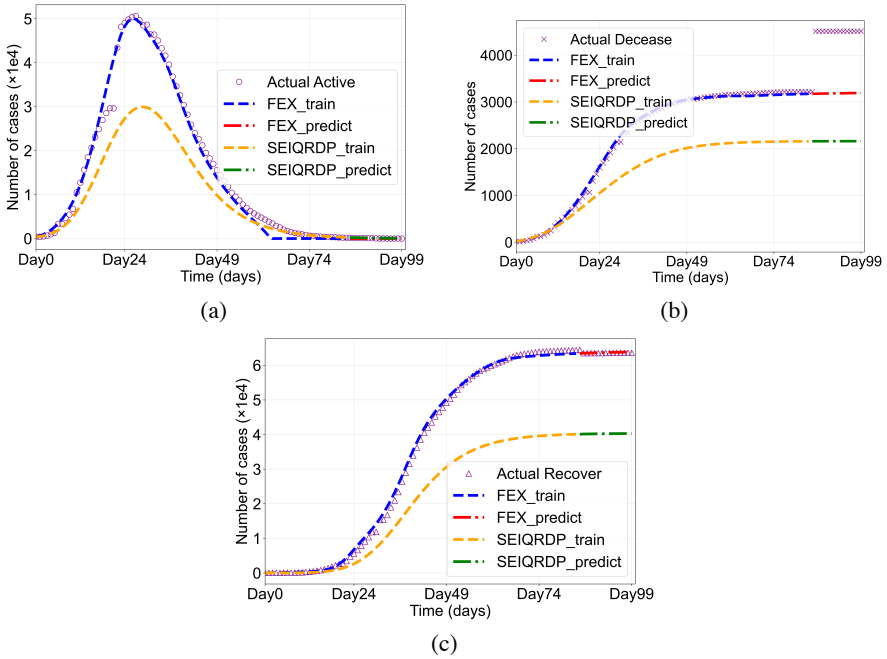


FIG. 6: Comparison of actual and predicted COVID-19 cases using the FEX and SEIQRDP: (a) active cases (Q), (b) deceased cases (D), and (c) recovered cases (R). FEX demonstrates superior data fitting during the training phase and accurate predictions in the testing phase, compared to SEIQRDP.

5. CONCLUSIONS AND FUTURE WORK

This paper introduces the FEX for learning epidemiological dynamics directly from data. The effectiveness of FEX is demonstrated through extensive experiments on both synthetic datasets—generated from SIR, SEIR, and SEIRD models—and real-world COVID-19 data. The results show that FEX not only achieves high accuracy in modeling dynamics but also produces interpretable mathematical expressions that approximate the underlying relationships. These expressions provide valuable insights into the interactions among different population groups in epidemiological data, enabling a deeper understanding of disease spread.

Despite its promising results, FEX faces several challenges when applied to learning epidemiological dynamics:

- *Computational cost*. The primary computational burden stems from the search process. Specifically, evaluating the score of each candidate operator sequence requires performing optimization from scratch, which is computationally intensive. Table 1 compares the training time, test time, and error rates of various methods used to learn the SIR model. Although FEX incurred higher training time costs compared to NN and RNN, this additional cost is justified by the improved interpretability and accuracy of the resulting finite expressions. We note that the current implementation processes candidate operator sequences sequentially during score computation and fine-tuning. Significant runtime improvements could be achieved through parallelization.

  Furthermore, FEX benefits from faster inference due to its compact representation, which uses significantly fewer parameters. For instance, in learning the SIR model, the FEX

TABLE 1: Comparison of training time, testing time, and errors across various methods for learning the SIR model

Methods	Training (sec)	Testing (sec)	Testing error
NN	144.1538	0.1738	6.6×10^{-6}
RNN	175.8494	0.1692	1.7×10^{-5}
FEX	4844.7300	0.1535	3.9×10^{-9}

involves only 31 parameters, whereas NN and RNN require 2435 and 32,796 parameters, respectively.

- *Nonuniqueness of solutions*. FEX may generate multiple valid finite expressions for the same dataset. While these solutions can still be meaningful, the lack of uniqueness complicates interpretability and may pose challenges for drawing definitive conclusions. Addressing this issue could involve incorporating improved regularization techniques and increasing the size and diversity of training data to enhance the consistency and reliability of the generated solutions.

Future work. We will focus on addressing these challenges, including the development of more efficient search algorithms and techniques to promote solution uniqueness. These advancements aim to enhance the utility and robustness of the FEX in practical epidemiological applications.

ACKNOWLEDGMENT

The research of Chunmei Wang was partially supported by National Science Foundation Grant DMS-2206332.

DATA AND CODE AVAILABILITY

The dataset used in this study is retrieved from Our World in Data at https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/jhu/full_data.csv (accessed on August 21, 2024).

All code, including implementations of the compared methods, is available on GitHub: https://github.com/Daviddjddu/FEX_Ep.git.

REFERENCES

Anderson, R.M. and May, R.M., *Infectious Diseases of Humans: Dynamics and Control*, Oxford, UK: Oxford University Press, 1991.

Bahloul, M.A., Chahid, A., and Laleg-Kirati, T.-M., Fractional-Order SEIQRDP Model for Simulating the Dynamics of COVID-19 Epidemic, *IEEE Open J. Eng. Med. Biol.*, vol. **1**, pp. 249–256, 2020.

Brauer, F., Castillo-Chavez, C., and Feng, Z., *Mathematical Models in Epidemiology*, New York: Springer, 2019.

Burden, R.L. and Faires, J.D., *Numerical Analysis*, Pacific Grove, CA: Cole Publishing Company, 1997.

Cranmer, M., Sanchez Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S., Discovering Symbolic Models from Deep Learning with Inductive Biases, *Adv. Neural Inf. Process. Syst.*, vol. **33**, pp. 17429–17442, 2020.

Feo, T.A. and Resende, M.G.C., Greedy Randomized Adaptive Search Procedures, *J. Global Optim.*, vol. **6**, pp. 109–133, 1995.

Fletcher, R., *Practical Methods of Optimization*, New York: John Wiley and Sons, 2000.

Goodfellow, I., *Deep Learning*, Cambridge, MA: MIT Press, 2016.

Grassly, N.C. and Fraser, C., Mathematical Models of Infectious Disease Transmission, *Nat. Rev. Microbiol.*, vol. **6**, no. 6, pp. 477–487, 2008.

Hethcote, H.W., The Mathematics of Infectious Diseases, *SIAM Rev.*, vol. **42**, no. 4, pp. 599–653, 2000.

Hochreiter, S. and Schmidhuber, J., Long Short-Term Memory, *Neural Comput.*, vol. **9**, pp. 1735–1780, 1997.

Huang, Z., Liang, M., Zhong, S., and Lin, L., AttNS: Attention-Inspired Numerical Solving for Limited Data Scenarios, in *Proc. of the 41st Int. Conf. on Machine Learning*, Vienna, Austria, Article No. 802, pp. 19929–19948, 2024.

Jiang, Z., Wang, C., and Yang, H., Finite Expression Methods for Discovering Physical Laws from Data, arXiv Preprint arXiv:2305.08342, 2023.

Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE), COVID-19 Data Repository, accessed August 21, 2024, from https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/jhu/full_data.csv, 2024.

Keeling, M.J. and Rohani, P., *Modeling Infectious Diseases in Humans and Animals*, Princeton, NJ: Princeton University Press, 2008.

Kermack, W.O. and McKendrick, A.G., A Contribution to the Mathematical Theory of Epidemics, *Proc. R. Soc. London, Ser. A: Math. Phys. Char.*, vol. **115**, no. 772, pp. 700–721, 1927.

LeCun, Y., Bengio, Y., and Hinton, G., Deep Learning, *Nature*, vol. **521**, no. 7553, pp. 436–444, 2015.

Liang, S., Huang, Z., and Zhang, H., Stiffness-Aware Neural Network for Learning Hamiltonian Systems, *International Conference on Learning Representations*, Virtual, 2022.

Liang, S., Lyu, L., Wang, C., and Yang, H., Reproducing Activation Function for Deep Learning, *Commun. Math. Sci.*, vol. **22**, no. 2, pp. 285–314, 2024.

Liang, S. and Yang, H., Finite Expression Method for Solving High-Dimensional Partial Differential Equations, arXiv Preprint arXiv:2206.10121, 2022.

Liu, Y., Chen, Y., Xiu, D., and Zhang, G., A Training-Free Conditional Diffusion Model for Learning Stochastic Dynamical Systems, arXiv Preprint arXiv:2410.03108, 2024.

Ma, C., Wang, J., and Weinan, E., Model Reduction with Memory and the Machine Learning of Dynamical Systems, *Commun. Comput. Phys.*, vol. **25**, no. 4, pp. 947–962, 2019.

Petersen, B.K., Landajuela, M., Mundhenk, T.N., Santiago, C.P., Kim, S.K., and Kim, J.T., Deep Symbolic Regression: Recovering Mathematical Expressions from Data via Risk-Seeking Policy Gradients, arXiv Preprint arXiv:1912.04871, 2019.

Raissi, M., Perdikaris, P., and Karniadakis, G.E., Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations, *J. Comput. Phys.*, vol. **378**, pp. 686–707, 2019.

Ruder, S., An Overview of Gradient Descent Optimization Algorithms, arXiv Preprint arXiv:1609.04747, 2016.

Rudin, C., Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, *Nat. Mach. Intell.*, vol. **1**, no. 5, pp. 206–215, 2019.

Schmidt, M. and Lipson, H., Distilling Free-Form Natural Laws from Experimental Data, *Science*, vol. **324**, no. 5923, pp. 81–85, 2009.

Shen, X., Cheng, X., and Liang, K., Deep Euler Method: Solving ODEs by Approximating the Local Truncation Error of the Euler Method, arXiv Preprint arXiv:2003.09573, 2020.

Song, Z., Wang, C., and Yang, H., Finite Expression Method for Learning Dynamics on Complex Networks, arXiv Preprint arXiv:2401.03092, 2024.

Su, W., Chou, C., and Xiu, D., Deep Learning of Biological Models from Data: Applications to ODE Models, *Bull. Math. Biol.*, vol. **83**, pp. 1–19, 2021.

Sutton, R.S. and Barto, A.G., *Reinforcement Learning: An Introduction*, Cambridge, MA: MIT Press, 2018.

Wilt, C., Thayer, J. and Ruml, W., A Comparison of Greedy Search Algorithms, *Proc. Int. Symp. Combinatorial Search*, vol. **1**, pp. 129–136, 2010.

APPENDIX A. SUPPLEMENTARY MATERIALS

APPENDIX A.1 Tree Types

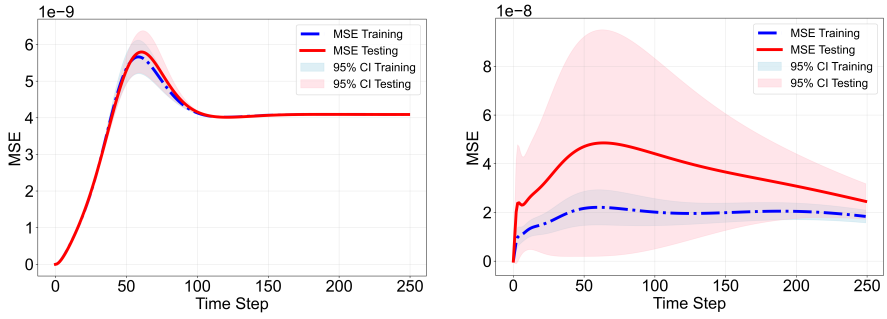
Table A1 summarizes the selected trees used to generate our results using the FEX. The tree selection is guided by epidemiological knowledge while balancing model flexibility and parsimony.

APPENDIX A.2 Generalization Capability of FEX

Figure A1 displays the MSE between the predicted trajectories and the true trajectory over time, using initializations from both the training and testing datasets. These predictions are derived

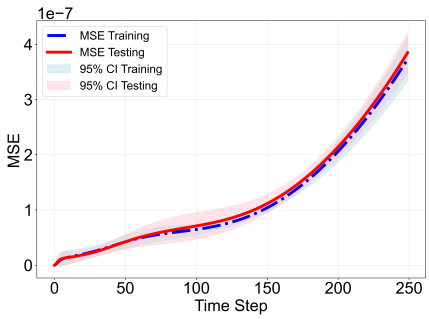
TABLE A1: Tree types used for learning different variables

	S	I	R	E	D	COVID
SIR model	Type 2	Type 2	Type 1	—	—	—
SEIR model	Type 2	Type 1	Type 1	Type 2	—	—
SEIRD model	Type 2	Type 1	Type 1	Type 2	Type 1	—
COVID	—	—	—	—	—	Type 2



(a) (b)

FIG. A1.



(c)

FIG. A1: Comparison of training MSE and testing MSE of FEX with 95% confidence interval for the three models: (a) SIR, (b) SEIR, and (c) SEIRD

from the identified FEXs generated by FEX for different epidemiological models: (a) SIR, (b) SEIR, and (c) SEIRD. As observed, the testing curve aligns closely with the training curve, demonstrating the generalization capability of the identified FEX.