Epidemics 55 (2026) 100919

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Forecasting seasonal influenza epidemics with physics-informed neural networks☆

Martina Rama [a,b,*], Gabriele Santin [c], Giulia Cencetti [d], Michele Tizzoni [e], Bruno Lepri [b]

[a] Department of Information Engineering and Computer Science, University of Trento, Trento, Italy
[b] Mobile and Social Computing (MobS) Lab, Fondazione Bruno Kessler (FBK), Trento, Italy
[c] Department of Environmental Sciences, Informatics and Statistics, University of Venice, Venice, Italy
[d] Centre de Physique Théorique, CNRS, Marseille, France
[e] Department of Sociology and Social Research, University of Trento, Trento, Italy

ARTICLE INFO

Keywords:
Physics-informed neural network
Epidemic modeling
Epidemic forecasting
Seasonal influenza

ABSTRACT

Accurate epidemic forecasting is critical for informing public health decisions and timely interventions. While physics-informed neural networks have shown promise in various scientific domains, their potential application to real-time epidemic forecasting remains underexplored. Here, we present SIR-INN, a hybrid forecasting framework that integrates the mechanistic structure of the classical Susceptible–Infectious–Recovered (SIR) model into a neural network architecture. Trained once on synthetic epidemic scenarios, the model is able to generalize across epidemic conditions without retraining. From limited and noisy observations, SIR-INN infers key transmission parameters via Markov chain Monte Carlo, generating probabilistic short- and long-term forecasts. We validate SIR-INN using national influenza data from the Italian National Institute of Health in the 2023–2024 and 2024–2025 seasons. The model performs competitively with current state-of-the-art approaches, particularly in terms of weighted interval score. It shows accurate predictive performance in nearly all phases of the outbreak, with improved accuracy observed for the 2024–2025 influenza season. Credible uncertainty intervals are consistently maintained, while coverage metrics highlight room for improvement in uncertainty calibration. SIR-INN offers a computationally efficient, transparent, and generalizable solution for epidemic forecasting, appropriately leveraging the framework's hybrid design. Its ability to provide real-time predictions of epidemic dynamics, together with uncertainty quantification, makes it a promising tool for real-world epidemic forecasting.

1. Introduction

In recent years, numerical modeling for epidemic forecasting has become a key tool for public health. Short-term forecasts of the epidemic burden, such as the expected number of cases, hospitalizations, or deaths, can inform policymaking and provide situational awareness to respond effectively with targeted intervention strategies (Desai et al., 2019; Lauer et al., 2021). Over the years, different epidemic forecasting approaches have been adopted to predict the temporal dynamics of dengue (Johansson et al., 2019), Ebola (Viboud et al., 2018), seasonal influenza (Biggerstaff et al., 2016; Brownstein et al., 2017), COVID-19 (Sherratt et al., 2023; Wolffram et al., 2023) and many other diseases (Del Valle et al., 2018; Holcomb et al., 2023).

On the one hand, recent developments in the field have highlighted the numerous advantages of using ensemble modeling strategies based on large collaborative efforts (Reich et al., 2019; Fox et al., 2024; Fiandrino et al., 2025). At the same time, the most effective standalone approaches generally combine statistical or machine learning techniques with conceptual models that incorporate infectious disease transmission dynamics, either explicitly through compartmental models or implicitly through factors such as seasonality or recent trends in observed data (Lopez et al., 2024; Ray et al., 2025). However, the use of such hybrid models is still limited in real-world applications, as demonstrated by a recent review of COVID-19 modeling studies in the United States, which showed that only 13% of the research teams adopted a hybrid methodology for pandemic forecasting (Nixon et al., 2022).

In general, hybrid methods have been developed to preserve the knowledge of the spreading mechanism, derived from mechanistic modelization, while simultaneously leveraging the data-driven learning capabilities and flexibility of machine learning models (Karniadakis

☆ This article is part of a Special issue entitled: 'AI for ID modeling' published in Epidemics.

* Corresponding author at: Department of Information Engineering and Computer Science, University of Trento, Trento, Italy.
E-mail address: mrama@fbk.eu (M. Rama).

https://doi.org/10.1016/j.epidem.2026.100919
Received 22 May 2025; Received in revised form 7 May 2026; Accepted 13 May 2026
Available online 28 May 2026

et al., 2021; O'Dea and Drake, 2022; Ye et al., 2025). Among hybrid models, a promising framework is emerging rapidly within the field of epidemiology: the Physics-Informed Neural Network (PINN) (Raissi et al., 2019). PINNs are capable of incorporating the observed data and mathematical models described by Partial Differential Equations (PDEs) through a regularization mechanism endowed with the neural network. Furthermore, compared with purely data-driven approaches, a PINN can take advantage of few and noisy data while still ensuring robustness and generalization capabilities for its physically consistent predictions (Raissi et al., 2019; Karniadakis et al., 2021; Cuomo et al., 2022).

For these reasons, PINNs are successfully used for a wide variety of applications such as aerodynamics, fluid mechanics, biology, and epidemiology (Lagergren et al., 2020; Mao et al., 2020; Cai et al., 2021). Specifically, within the epidemiological context, the integration of prior epidemiological knowledge (derived from mechanistic models) enables the neural network to serve as an efficient surrogate, accurately learning the dynamics of disease spread and inferring disease-related parameters (Shaier et al., 2021; Bertaglia et al., 2022; Ning et al., 2023a,b; Qian et al., 2025). However, despite the fact that the literature on epidemiological PINNs is becoming vast and ubiquitous, the prediction of real-time epidemic scenarios with PINNs has not been extensively explored. Some recent efforts have begun to address this gap. Madden et al. (2024) has proposed integrating a time series SIR model (TSIR) with PINNs that improves both forecast and parameter inferences of measles dynamics. Berkhahn and Ehrhardt (2022) implemented a PINN for COVID-19 modelization and future outbreak scenario generation. Millevoi et al. (2024) implemented a multiple PINNs framework investigating joint and split approaches to estimate temporal changes in model parameters and state variables, also providing a short-term forecast application on Italian COVID-19 data. The Epidemiologically-Informed Neural Networks (EINNs) framework, introduced by Rodríguez et al. (2023), combines a time module and a feature module to leverage multiple data sources for both short-term and long-term forecasting. Finally, Kharazmi et al. (2021) investigated several epidemiological models adopting the PINNs approach that uses multiple neural networks, identifying time-dependent parameters and forecasting with uncertainty quantification.

Most of these works adopt a framework that simultaneously addresses both forward and inverse problems, fully exploiting the potential of PINNs. Alternatively or jointly, some of the aforementioned approaches rely on extensive use of different neural networks, employing them independently to solve specific tasks (e.g., learning the compartments of the PDEs model, finding the unknown parameters) and subsequently combining their results. In either case, when aiming for real-time forecasting, a weakness emerges: the designed framework requires neural network retraining whenever new observations (or new epidemic scenarios) become available, thus increasing both the computational cost and the delay in receiving results. This also restricts the generalizability of a single pre-trained neural network. Moreover, by design, the neural network evaluation produces one single deterministic output as forecast trajectories. As a drawback, in almost all of these cases, there is no uncertainty analysis with respect to the forecasting process, which, instead, is provided by statistical inferential approaches and stochastic simulation studies (Bracher et al., 2021; Sherratt et al., 2023).

To address these limitations while preserving the known benefits and potential of a hybrid approach, we propose here a novel framework for epidemic forecasting, namely Susceptible–Infectious–Recovered–Informed Neural Network (SIR-INN). In particular, we construct a single PINN that takes advantage of the simplicity and modeling ability in describing different epidemic dynamics of the SIR model. The prior epidemiological knowledge embedded in the neural network architecture is directly derived from the differential equations underlying the SIR compartmental model, with constant transition rates. Our PINN is trained once on synthetic observations of epidemic scenarios that are close to the real ones in a temporal-epidemic domain that considers both the epidemic parameters and time as variables. This enables our model to efficiently learn a wide range of disease dynamics, generalizing its understanding to the transmission patterns of other infectious diseases. Thus, starting from a limited and noisy set of observations, the pre-trained SIR-INN is able to estimate, via the Markov Chain Monte Carlo (MCMC) method, the parameters that characterize the SIR-based epidemic dynamics. Finally, by exploiting these estimated parameters, our model can efficiently forecast future time windows. In particular, our SIR-INN solution provides (i) a single neural network trained on synthetic data, eliminating the need for retraining or relying on multiple networks and significantly limiting the computational cost of the method, while preserving the mechanistic structure of epidemic spreading; (ii) an efficient inference model for epidemiological data, combining a deterministic neural network as the underlying model with MCMC; (iii) explicit uncertainty quantification of the estimated epidemiological parameters, finally ensuring probabilistic forecasting.

To validate our proposed framework in a real-world epidemic scenario, we use national data from the Italian seasonal influenza surveillance system, provided by the Italian National Institute of Health (ISS) (Istituto Superiore di Sanità, 2020). In particular, we perform forecasting simulations four weeks and ten weeks ahead for both the 2023–2024 and 2024–2025 influenza seasons. Furthermore, we compare our results with state-of-the-art approaches for seasonal influenza forecasting, presented in the Influcast Hub (InfluCast, 2025), and adopt the same evaluation metrics as those used by Fiandrino et al. (2025). Such results and comparisons suggest that SIR-INN offers an efficient, cost-effective, and accurate hybrid alternative for real-world epidemic forecasting. It maintains high predictive accuracy even over long-term horizons and across multiple epidemic scenarios, while ensuring both generalizability and reliable uncertainty quantification.

The remainder of the paper is structured as follows. In Section 2, we present the structure of the SIR-INN methodology, highlighting the three phases of model approximation, parameter inference, and forecasting. The details of the implementation of all components of the methodology are reported in Section 3. The numerical results of the application of the proposed framework in seasonal influenza scenarios are provided in Section 4. Finally, we summarize our work and derive our conclusions in Section 5.

2. Methods

The SIR-INN framework presented in this work mainly consists of three consecutive and related steps: the PINN training, the inference of the disease parameters, and, finally, the forecasting of the future disease states.

In the first phase, described in detail in Section 2.1, we implement a SIR-based PINN. Specifically, we insert the ordinary differential equations of a SIR epidemic model as a regularization term into the loss function of a neural network. Then, we train the PINN on synthetic data close to real disease scenarios, as explained in detail in Section 3.1. The following steps leverage this pre-trained PINN with the aim of forecasting the disease spreading process starting from a few noisy observations. In particular, Section 2.2 (and Section 3.2) describes how, from a selected time window of observations, we infer the parameters that characterize the disease-spreading dynamics via the MCMC method. Then, by evaluating our pre-trained PINN model on these estimated parameters and on a future time window, we can perform both short-term and long-term disease forecasting. This last step is described in detail in Section 2.3.

2.1. SIR model approximation via PINN

PINNs are universal function approximators efficiently endowed with physical knowledge described in terms of systems of either Ordinary Differential Equations (ODEs) or PDEs (Raissi et al., 2019).

Here, we focus on a PINN that incorporates an epidemic spreading dynamic as physical knowledge, while fitting some synthetic data that describe plausible epidemic scenarios (see also Section 3.1 for details). In this setting, a well-known and widely used compartmental model is the SIR epidemic model (Kermack and McKendrick, 1927), defined by

$$\begin{cases} \frac{dS(t)}{dt} &= -\beta S(t) I(t), \\ \frac{dI(t)}{dt} &= \beta S(t) I(t) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t), \end{cases} \tag{1}$$

where $S : [t_0, T] \to [0,1]$ represents the proportion of susceptibles, $I : [t_0, T] \to [0,1]$ the proportion of infected individuals, and $R : [t_0, T] \to [0,1]$ the proportion of removed individuals, with respect to the total population size N. Individuals are transferred between compartments through two transition rates: $\beta \in \mathbb{R}_{\geq 0}$, the epidemic contact or transmission rate, and $\gamma \in \mathbb{R}_{\geq 0}$, the removal rate (recovered or mortality). A fundamental threshold quantity for the analysis of disease spread is the so-called basic reproduction number $R_0 = \beta/\gamma$, which represents the expected number of new infections produced by a single infected individual during their infectious period when introduced into a population where all subjects are susceptible (Heffernan et al., 2005). The effective reproduction number, R_t, is a time-dependent quantity that measures the number of new infections caused by an infected individual at any point in time of the outbreak. It can be estimated based on the basic reproduction number and on the proportion of susceptible in the population as $R_t = R_0 \cdot S(t)$, by assuming constant β and γ (Nishiura and Chowell, 2009).

Given the initial conditions $S(0), I(0), R(0)$ the system is fully specified. Summing the equations in (1), we obtain $\frac{d}{dt}(S(t) + I(t) + R(t)) = 0$, which implies that the total population is constant over time. Since the model is expressed in normalized form, it follows that $S(t) + I(t) + R(t) = 1 \ \ \forall t \in [t_0, T]$. Note also that the epidemic dynamics do not allow any effects of births and deaths on the populations since the time scale of the epidemic is assumed to be shorter with respect to the vital dynamics mechanism.

The first step, illustrated in Fig. 1, is to implement a PINN that learns the SIR model defined by Eq. (1). This ODE system can be formulated as a parametrized non-linear ODEs system

$$\frac{du(t;\lambda)}{dt} + F[u(t;\lambda);\lambda] = 0, \qquad t \in [t_0, T], \tag{2}$$

where $\lambda := (\beta, \gamma) \in \mathbb{R}^2_{>0}$ are the parameters of (1), $F[u;\lambda]$ is the parametric non-linear differential operator, and $u : [t_0, T] \times \mathbb{R}^2 \to [0,1]^3$, $u(t;\lambda) := (S(t), I(t), R(t)) \in [0,1]^3$ is the solution map (Folland, 1995; Evans, 2022).

With the framework defined by Eq. (2), the objective of the PINN approach is as follows. Given fixed model parameters $\lambda \in \mathbb{R}^m$, find a neural network $u_N(t;\lambda)$ that approximates the solution $u(t;\lambda)$ of the ODEs system. We recall that a neural network u_N is a parametric function depending on a vector θ of parameters learnable via an optimization procedure, which will be explained in the following, that finds the values of θ which steer the network output towards some desired values. Since the specific inner working of these parameters is not the focus of our paper, we omit the dependence on θ in our notation, and we refer to LeCun et al. (2015) and to Goodfellow et al. (2016) for further details. To ensure that the neural network that approximates the solution $u(t;\lambda)$ efficiently learns a variety of epidemic scenarios, we extend the temporal domain of the function approximator to a temporal-epidemic domain that considers the epidemic parameters as variables in addition to the time variable. Hence, we construct the PINN as a function $u_N : \mathcal{T} \times \mathcal{B} \times \Gamma \to [0,1]^3$, namely $u_N(t, \beta, \gamma)$, with the aim of approximate the solution $u(t;\beta,\gamma)$. Although u_N is in principle defined for more general inputs, we aim at an accurate model for values in $\mathcal{T} \times \mathcal{B} \times \Gamma$, with $\mathcal{T} = [t_{min}, t_{max}]$, $\mathcal{B} = [\beta_{min}, \beta_{max}]$, and $\Gamma = [\gamma_{min}, \gamma_{max}]$, where $0 \leq t_{min} < t_{max}$, $0 \leq \beta_{min} < \beta_{max}$, $0 \leq \gamma_{min} < \gamma_{max}$ are suitable values that are specified in the following.

The methodology setting is data-driven, i.e., we suppose we have knowledge of a number of samples of the unknown solution, namely

$$\mathcal{X}_{tr} = \{(t_i^u; \beta_j^u, \gamma_k^u),\ i = 1, \dots, N_t, j = 1, \dots, N_\beta, k = 1, \dots, N_\gamma\} \subset \mathcal{T} \times \mathcal{B} \times \Gamma,$$
$$\mathcal{Y}_{tr} = \{y_i = u(t_i^u; \beta_j^u, \gamma_k^u), \quad i = 1, \dots, N_t, j = 1, \dots, N_\beta, k = 1, \dots, N_\gamma\}, \tag{3}$$

where N_t, N_β and N_γ are the total number of sampled parameters t_i^u, β_j^u and γ_k^u, respectively. Note that these parameters are chosen in the same intervals $\mathcal{T}, \mathcal{B}, \Gamma$ defined before. The values y_i are collected by numerical solutions of the system (2) for the corresponding parameters.

The neural network's parameters are then optimized to minimize a certain loss function, which is a measure of the network's accuracy. In particular, PINNs use a model-dependent, customized loss that integrates the governing equations of the ODEs system (2) while fitting the data (Raissi et al., 2019). More precisely, the loss has two components

$$\mathcal{L} = \mathcal{L}_{data} + \mathcal{L}_{ODE}, \tag{4}$$

where

$$\mathcal{L}_{data} = \mathrm{MSE}(u_N, \mathcal{X}_{tr}, \mathcal{Y}_{tr}) \tag{5}$$

is the standard Mean Squared Error (MSE) loss, i.e., the approximation error computed considering the training dataset (3) of $N_t \cdot N_\beta \cdot N_\gamma$ points, while the physics-endowed loss $\mathcal{L}_{ODE}$ is the component that forces the neural network to approximate the ODEs. This is done by defining a so-called residual function $R(u, t; \lambda)$ associated to the ODEs (2), i.e.,

$$R(u, t; \lambda) := \frac{du(t;\lambda)}{dt} + F[u(t;\lambda);\lambda], \qquad t \in [t_0, T], \qquad \lambda \in \mathbb{R}^2_{>0}, \tag{6}$$

in such a way that the physics-endowed loss penalizes the solutions with a large residual, i.e.

$$\mathcal{L}_{ODE} = \frac{1}{N_t^R \cdot N_\beta^R \cdot N_\gamma^R} \sum_{i,j,k} R\left(u_N, t_i^R; \beta_j^R, \gamma_k^R\right)^2, \tag{7}$$

where the points are taken in the so-called collocation set

$$\mathcal{X}_{coll} = \{(t_i^R; \beta_j^R, \gamma_k^R),\ i = 1, \dots, N_t^R, j = 1, \dots, N_\beta^R, k = 1, \dots, N_\gamma^R\}. \tag{8}$$

The construction of this physics-informed loss function enables the addition of a regularization term that drives the network to learn the underlying model structure, simultaneously fitting the data, and preserving the initial conditions if they are specified. Compared to purely data-driven training, this mechanism allows us to benefit from one simple feedforward neural network, with few layers and neurons, and to train it on small amounts of data (Raissi et al., 2019; Karniadakis et al., 2021). Although the two loss terms in (4) are sometimes weighted with different coefficients in order to balance their relative importance, we found in our results that no noticeable difference can be obtained in this way (see Appendix D.3). We thus keep an unweighted definition for the loss in our framework.

2.2. Parameters inference

As a second step of our methodology, we employ an MCMC (Neal, 1993; Richey, 2010) method to extract, from a few noisy observations, the epidemic features that describe the underlined disease spreading process.

In particular, once our SIR-INN is trained on synthetic data approximating credible epidemic scenarios as described in Section 2.1 (and in Section 3.1), we use MCMC to estimate the epidemic parameters that drive the PINN solution to be as close as possible to the observations. As fundamental parameters that characterize the outbreak dynamics, we choose the initial time $\tau_0 \in \mathbb{R}_{\geq 0}$ of the observed spreading process, that is the time at which we start to observe the data, and the two transition rates of the SIR-based model: the transmission rate and the removal rate, $\beta \in \mathbb{R}_{\geq 0}$ and $\gamma \in \mathbb{R}_{\geq 0}$, respectively. It is important to note that at this stage the weights of the neural network are fixed, and the trained SIR-INN represents a function $u_N : \mathcal{T} \times \mathcal{B} \times \Gamma \to$

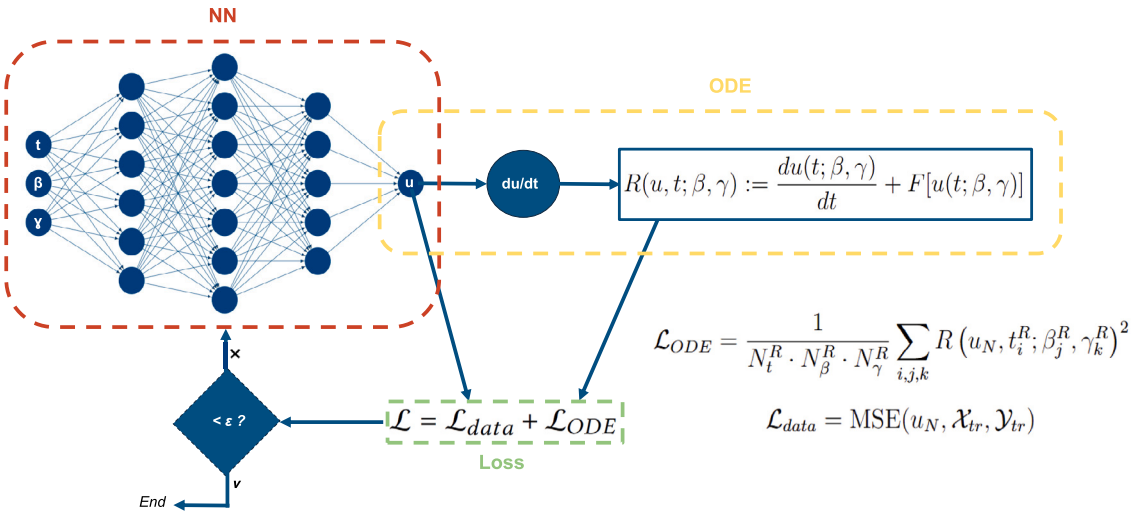

Fig. 1. Diagram of the PINN architecture for the SIR model learning. The red-dashed frame represents the neural network which takes as inputs the time $t \in \mathcal{T}$, the disease transmission rate, $\beta \in \mathcal{B}$, the recovery rate, $\gamma \in \Gamma$, and outputs $u \in [0,1]^3$, i.e., the three normalized components of the SIR epidemic model: susceptibles, infectious, and recovered. Furthermore, the neural network is integrated with the physical knowledge of the SIR epidemic model, represented by the yellow-dashed frame, through the insertion of the ODEs system into the loss function of the neural network, becoming a physics-informed neural network. PINN parameters are obtained by minimizing the total loss function, represented by the green-dashed frame, on the training data and satisfying a SIR-based dynamics as well. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$[0,1]^3$. Then, the MCMC method is implemented assuming, as priors, uniform distributions for all the parameters and, as reference model, our pre-trained PINN. Furthermore, we chose a likelihood function that reflects the nature of epidemiological data: a Poisson distribution with a function of the model output serving as its mean. For details about this step, please refer to Section 3.2.

Once we run the MCMC algorithm for a selected time window of observations, we obtain a Markov chain of the estimated parameters. We then select 1000 posterior samples of the tails, namely: $\hat{\tau}_0, \hat{\beta}, \hat{\gamma}$, from which we can extract summary statistics and perform probabilistic forecasting.

2.3. Forecasting

The last step of our methodology, that is forecasting the future state of the epidemic, is the most crucial but perhaps the simplest to implement. Indeed, it only amounts at evaluating the pre-trained SIR-INN model with the parameters estimated via MCMC, as explained in the previous section.

Namely, for each observation window we obtain $\hat{\tau}_0, \hat{\beta}, \hat{\gamma}$, as in Section 2.2, and evaluate the SIR-INN model on $\hat{\beta}, \hat{\gamma}$ parameters (with their uncertainty) and on the full time window, starting at time 0. We then translate time 0 to the estimated initial time $\hat{\tau}_0$ to obtain the predicted S, I, and R components of interest. The resulting values can be used both for short- and long-term probabilistic forecasting, just selecting a forecasting window size.

Iterating this procedure over each time rolling-window of observations, it provides forecasts over the entire time period of interest.

3. Implementation and evaluation details

All simulations were coded in the Python programming language using the libraries `PyTorch` version 2.6.0 (Paszke et al., 2019), `Pymcmcstat` package version 1.9.1 (Miles, 2019), `seaborn` library version 0.13.2 (Waskom, 2021). Validation metrics were computed using the R programming language, specifically with the `scoringutils` package version 1.2.2 (Bosse et al., 2022).

3.1. PINN architecture and training

Regarding the neural network architecture, we leverage the regularization mechanism of the PINNs to benefit from simple feedforward neural networks, minimizing as much as possible the computational cost derived from the usual deep neural networks (Raissi et al., 2019; Karniadakis et al., 2021). We implemented a single PINN consisting of only three hidden layers, and an overall architecture of $[3, 16, 32, 16, 3]$, where each value denotes the number of neurons per layer, from input to output. As activation functions, we use smooth differentiable functions commonly used for PINNs (Cuomo et al., 2022): hyperbolic tangent for the hidden layers and a sigmoid activation function for the output layer, to have outputs in the range $[0,1]$. This data normalization is commonly adopted for training neural networks: in our case, this requires normalizing the data by the total population size N, as $(S/N, I/N, R/N)$ components, during training. This normalization is consistent both with the ODE system (1) embedded in the neural network via loss function (7), and with the second step of our framework (i.e., parameter inference), where the model outputs are compared with incidence data, which are inherently expressed as proportions with respect to the population size.

As described in Section 2.1, to train our SIR-INN, we chose synthetic data for the training set and the collocation set. In particular, since we generated them to construct the training set (3), we used the same samples for the collocation set (8). Hence, following the same notation of Section 2.1, we considered $\mathcal{X}_{tr} \equiv \mathcal{X}_{coll}$. For the time component, t_i^u, we used weekly time data in a time window of 600 days; from $t_1^u = 0$ to $t_{86}^u = 595$. Regarding the ranges of parameters, we define a uniform grid of size 50 for the removal rate $\gamma \in [1/12, 1/2.5]$, and the transmission rate $\beta \in [0.12, 0.45]$. Then we restricted the values with linear constraints around the diagonal of the grid (that is, around $R_0 = \beta/\gamma = 1$) obtaining 1068 pairs of values (β_j^u, γ_k^u), $j = 1, \dots, N_\beta, k = 1, \dots, N_\gamma$, where $N_\beta \cdot N_\gamma = 1068$. These intervals were selected to represent plausible epidemiological conditions for seasonal influenza (Carrat et al., 2008; Biggerstaff et al., 2014). In particular, the corresponding infectious period $1/\gamma$ ranges from approximately 2 to 12 days, while the basic reproduction number $R_0 = \beta/\gamma$ spans the interval $(0.75, 2.5)$. These values were chosen with the aim of covering a

sufficiently broad domain that could both plausibly contain parameter values inferred from real influenza surveillance data and maintain the neural network generalization capabilities. Indeed, the training domain contains both subcritical regimes (i.e., $R_0 < 1$) in which epidemics do not occur, and regimes where $R_0 > 1$ characterized by epidemic growth that allows the framework to deal with different influenza seasons. A visualization of all the parameter pairs included in the training set is provided in Appendix C (Figure C.11).

To construct the $\mathcal{Y}_{tr}$ training set (3), for each pair of parameters, (β_j^u, γ_k^u), we first generated the SIR components via *odeint*. This is done considering a daily time window of 600 days for t_i^u, a population size $N = 10^6$, and the SIR initial conditions $(S(0), I(0), R(0)) = (N-1, 1, 0)$. Second, we normalized all the values by the total population size, which correspond to obtain the solutions $u(t_i^u; \beta_j^u, \gamma_k^u)$, defined in Section 2.1. Third, we restricted the generated samples to the weekly time window, coherently with the $\mathcal{X}_{tr}$ training set (3). We train our PINN for 7500 epochs using the Adamax optimization algorithm, with a learning rate of 0.001. The neural network is trained with the same batches of 100 training, validation and collocation points, with train size = 0.9.

All of these parameters and synthetic data choices are chosen according to plausible scenarios of rapidly transmitted respiratory infection spreading, while trying to generalize them to better train our neural network, increasing its generalization capabilities. For more details about the approximation capabilities of the network, including an out-of-distribution analysis, and the suitability of the training set showing that the epidemiological parameters inferred from real influenza data lie well within the training domain, we refer the reader to Appendix C. Regarding the choice of N, it is intentional and does not affect the general results of the model, since the SIR dynamics is scale-invariant for sufficiently large populations. Moreover, we consider N as an effective interacting population, since assuming the full Italian population to be susceptible and well mixed would be unrealistic (see also Appendix D.1 for a sensitivity analysis on N). Similarly, although the PINN is trained on deterministic synthetic data, with stochasticity incorporated at the parameter inference stage rather than at training time, we verify in Appendix D.2 that introducing stochastic perturbations during training results in no significant improvement across all evaluation metrics, confirming that the deterministic training strategy is sufficient. Furthermore, although population conservation $S(t) + I(t) + R(t) = 1$ is not explicitly enforced during training, we verify in Appendix B that the trained PINN satisfies it within a Mean Squared Error (MSE) of 10^{-4}, confirming that the network implicitly learns to respect this physical constraint.

3.2. Inference of disease parameters

Parameter estimation is obtained via MCMC, as described in Section 2.2. The reference data are Influenza-Like-Illness (ILI) incidence observations, which are calculated as weekly reported cases per 1000 patients. We rescale it to our reference population of $N = 10^6$ (i.e., we multiply the measured incidence by $N/1000$).

As the network $u_N(t, \beta, \gamma)$ outputs normalized S, I, R values, we first define a function G extracting the corresponding estimated weekly incidence. We fix a time grid $\tau_j = \tau_0 + 7j$, $j = 0, 1, \dots$ depending on an unknown alignment time $\tau_0 \in [0, 600]$, and define

$$G(j; \tau_0, \beta, \gamma) = \left(u_N^{(0)}(\tau_{j-1}, \beta, \gamma) - u_N^{(0)}(\tau_j, \beta, \gamma) \right) \cdot N, \tag{9}$$

where the superscript $u_N^{(0)}$ denotes the first output component, i.e., the estimated value of S.

Parameter estimation now amounts at estimating τ_0, β, γ in G from the measured incidence data. For this, we consider a rolling time window of $M = 5$ weeks of observations, let $y_1, \dots, y_5$ be the rescaled incidence per N people measured in these weeks, and assume

$$y_j \sim \text{Poisson}(\mu_j), \quad \mu_j = G(j; \tau_0, \beta, \gamma), \tag{10}$$

i.e., a Poisson distribution which reflects the counting nature of the epidemiological data, with mean modeled by the PINN-estimated incidence.

We note that if τ_0 was kept fixed over the entire estimation period it would correspond to an estimated elapsed time from the epidemic start, but this would assume an exact, fixed-parameter SIR to exactly reproduce the whole sequence of data. Since this is an unrealistic assumption, and since τ_0 is re-estimated over each $M = 5$ time-window, it rather plays the role of a time-alignment factor.

The MCMC algorithm now samples from the posterior distribution

$$p(\tau_0, \beta, \gamma \mid \mathbf{y}) \propto p(\mathbf{y} \mid \tau_0, \beta, \gamma) \cdot p(\tau_0, \beta, \gamma), \tag{11}$$

where $\mathbf{y} = (y_1, \dots, y_5)$ are the observations, $p(\mathbf{y} \mid \tau_0, \beta, \gamma)$ is the Poisson likelihood, and $p(\tau_0, \beta, \gamma)$ is the prior distribution. The corresponding log-likelihood is

$$\log p(\mathbf{y} \mid \tau_0, \beta, \gamma) = \sum_{j=1}^{M} \left[y_j \log \mu_j - \mu_j - \log \Gamma(y_j + 1) \right], \tag{12}$$

where $\Gamma(\cdot)$ is the gamma function. As mentioned in Section 2.2, we use uniform and static priors for all parameters, i.e.

$$\beta \sim \mathcal{U}(0.12, 0.45), \quad \gamma \sim \mathcal{U}(1/12, 1/2.5), \quad \tau_0 \sim \mathcal{U}(0, 400), \tag{13}$$

where the bounds are chosen to encompass biologically plausible ranges for influenza transmission, in line with the training set choice (Section 3.1). This choice reflects the rolling-window nature of our inference framework: parameters are independently estimated at each forecast round without assuming temporal continuity, and we lack reliable prior information about the parameter distributions for each weekly window. In Appendix E.2, we additionally report an analysis using weakly informative Gaussian priors that lead to quantitatively similar results.

For each observation window (i.e., round), the MCMC sampling is performed for 10,000 iterations with initial parameter values set to $(\tau_0, \beta, \gamma)|_{\text{init}} = (200, 0.28, 0.24)$, namely, the midpoints of the intervals in (13). For each round we obtain a posterior from which we discard the first 9000 iterations as burn-in. We obtain 1000 posterior samples (i.e., $\hat{\tau}_0, \hat{\beta}, \hat{\gamma}$) from which we extract summary statistics and do probabilistic forecasting (Section 2.3) by finally converting SIR-INN outputs — in terms of estimated S, I, and R components — to incidence values, coherently with the data.

We further remark that the choice of 10,000 iterations is robust: extending to 50,000 iterations does not substantially affect the results (Appendix E.1).

All these steps are implemented using the Python package `Pymcmcstat`, version 1.9.1 (Miles, 2019), that can run MCMC on any model implemented by a Python function, which in our case is simply the pre-trained physics-informed neural network implemented in `PyTorch` (version: 2.6.0).

3.3. Epidemiological data

We used seasonal influenza national data provided by the Italian National Institute of Health (ISS) (Istituto Superiore di Sanità, 2020) through the RespiVirNet surveillance system. These consist of weekly ILI incidence observations reported by sentinel doctors at both national and regional levels (with the exception of the Valle d'Aosta and Calabria regions for the season 2023–2024), starting from the season 2003–2004 to the most recent, 2024–2025. In particular, we perform simulations for both the seasons 2023–2024 and 2024–2025, with influenza surveillance running from November to May, starting from week 42 in a given year and continuing until week 17 of the following year. All data is publicly available on GitHub at: RespiVirNet, 2020 (2020).

3.4. Forecasting evaluation metrics

We evaluate our SIR-INN framework forecasting performances using several metrics, employed and described in detail in Fiandrino et al. (2025). Further details are provided in Appendix G.

Absolute Error (AE). It is computed as the absolute value of the difference between the median forecast and the actual corresponding value. The Mean Absolute Error (MAE) is the average of the AE over different time steps. We also implemented it by averaging for all the forecasting time windows.

Weighted Interval Score (WIS). It is an approximation of the Continuous Ranked Probability Score (CRPS) (Bracher et al., 2021) that evaluates both the accuracy of the forecast median solution and of the prediction intervals in containing the actual observations. The metric is defined for a given prediction interval of a model forecast, generalizing the interval score to multiple prediction intervals. We refer to Appendix G for the exact choice of the corresponding parameters.

Coverage. It measures the calibration of a model and is defined as the fraction of times a prediction interval contains the actual data. Notice that in a well-calibrated model, the coverage values should closely match the nominal levels of the predictive intervals. For example, observations will be included exactly 90% of the time in the prediction interval 90%. We also computed the coverage for both the 50% and 90% prediction intervals of our forecasts.

4. Results

As discussed in Section 3.3, to validate our methodology we focus on disease forecasting using national data from the Italian influenza surveillance. In particular, to increase both the complexity and the relevance of the task, we put ourselves under the same conditions as the teams that participated in the Influcast Hub challenge 2023–2024 (InfluCast, 2025). Specifically, we perform four weeks ahead forecasting simulations, as outlined in Section 4.2.1, and we employ the same validation metrics adopted in Fiandrino et al. (2025) to ensure a fair comparison between our model and all the models of the participating teams. For the corresponding quantitative results, we refer to Section 4.2.2. Moreover, in Section 4.1 we show the results on parameter inference obtained by our model using a methodology MCMC. In addition, in Section 4.2.3 we analyze the performance of our approach in a long-term forecasting task. All experiments were carried out using data from two different influenza seasons (years 2023–2024 and 2024–2025), to illustrate the ability of our SIR-INN hybrid methodology to generalize across different epidemic scenarios without retraining the neural network.

4.1. Parameters inference results

As explained in Section 3.2, after we have selected an ideal size for the observation time window, $M = 5$, we perform an MCMC simulation (i.e., round) every time we observe new data, in a rolling-window manner. For example, at week 51, we use five data points from the weeks 47-51 to estimate the parameters that best align the SIR-based PINN output to the observed ILI incidence in this window. In line with the Influcast Hub, we performed simulations starting with four weeks of observation (also for the 2023–2024 influenza season) and considering retrospective weekly updates of past observations, as the system provides new surveillance data. In the end, the total number of simulations is equal to the total number of observations (i.e., the total number of weeks in a season) minus four, which is 24 for both seasons. Note that for the first simulation we used only four observations, as they are the only available.

Figs. 2 and 3 show the posterior distributions (specifically, the 1000 posterior samples of the chain tails) of the estimated parameters $\hat{\beta}, \hat{\gamma}, \hat{\tau}_0$ and the posterior distributions of $\hat{R}_t$ for all simulations (that is, weeks), for the season 2023–2024 and 2024–2025, respectively. From the plots, we observe that the overall median values of the parameters are similar across the two seasons, and their values remained stable as the seasons progressed. The median disease transmission rates, β, are close to 0.30 and the median removal rates, γ, are (slightly lower) close to 0.27 day^{-1}, leading to an average R_0 equal to 1.13, for the 2023–2024 influenza season, and to 1.14, for the 2024–2025 (see Figure A.7 and the relative section for additional information). Regarding the observation alignment time, $\hat{\tau}_0$, its median throughout the weeks is close to day 300, both for the 2023–2024 and 2024–2025 seasons.

For each epidemic week, fixing a posterior sample, the effective reproduction number $\hat{R}_t$, coherently with the definition given in Section 2.1, is computed as:

$$\hat{R}_t = \hat{R}_0 \cdot u_N^{(0)}(\hat{\tau}_{M-1}, \hat{\beta}, \hat{\gamma}) \tag{14}$$

where $\hat{R}_0 = \hat{\beta}/\hat{\gamma}$ is the estimated basic reproduction number, and $u_N^{(0)}(\hat{\tau}_{M-1}, \hat{\beta}, \hat{\gamma})$ is the SIR-INN predicted susceptible fraction of the population at the current time $\hat{\tau}_{M-1} = \hat{\tau}_0 + 7 \cdot (M-1)$ (see Section 3.2), depending on the observation alignment time $\hat{\tau}_0$ estimated from the corresponding posterior sample.

Regarding $\hat{R}_t$, the results are consistent with the expected epidemic dynamics for both seasons. While in 2023–2024 the ILI incidence peaked at weeks 51–52, in 2024–2025 the peak arrived a month later, at weeks 4 and 5. For the first season (Fig. 2), both the posterior median and the interquartile range (IQR) remain above the epidemic threshold $\hat{R}_t = 1$ up to week 52, indicating sustained transmission, while at week 1 the median reaches $\hat{R}_t \approx 1$ and from week 2 onwards both the median and the interquartile range fall below the threshold, suggesting that the outbreak had entered the decline phase. A similar pattern is observed for the 2024–2025 season (Fig. 3), although the growth phase is more prolonged: the posterior median stays above $\hat{R}_t = 1$ up to week 4, the median reaches the threshold at week 5, and from week 6 onwards both the median and the interquartile range lie below 1. However, during the early weeks of this season, the posterior distributions exhibit greater uncertainty, with the interquartile range crossing the threshold $\hat{R}_t = 1$ from below, reflecting a less definitive signal of epidemic growth compared to the 2023–2024 season. Overall, the temporal evolution of $\hat{R}_t$ across both seasons is coherent with the observed ILI incidence curves, capturing the transition from the epidemic growth phase to the declining phase with epidemiologically plausible timing.

For the same analysis performed with a chain of length 50,000, please refer to Appendix E.1.

Even if the best parameter values are re-estimated every week, as new surveillance data are reported and previously published data are updated, our model consistently identifies a narrow range of plausible values that remain almost constant across the season. The largest fluctuations are observed for the estimated $\hat{\tau}_0$, which the model adjusts to capture the temporal dynamics of the outbreak, which were different in the two seasons, as we highlighted before.

Figure C.11 shows that the posterior medians of these inferred parameter pairs $(\hat{\beta}, \hat{\gamma})$ fall well inside the training parameter area, demonstrating that our choice for the ranges of the training parameters was appropriate. Furthermore, the parameter values of $\hat{\beta}$ and $\hat{\gamma}$ fall in line with the typical values measured for seasonal influenza. In particular, an average infectious period of about 4 days ($\gamma = 0.25$ day^{-1}), and a basic reproductive number of $R_0 \sim 1.2$ correspond to the traditional estimates of these parameters for seasonal flu (Carrat et al., 2008; Biggerstaff et al., 2014). The values of the estimated parameters and their stability between weeks are encouraging, even if this does not necessarily imply a global epidemiological identifiability and interpretability. Thus, although this is not explicitly enforced, we see that the data and the structure based on SIR allow the model to effectively learn the epidemiological role of the two parameters, β and γ.

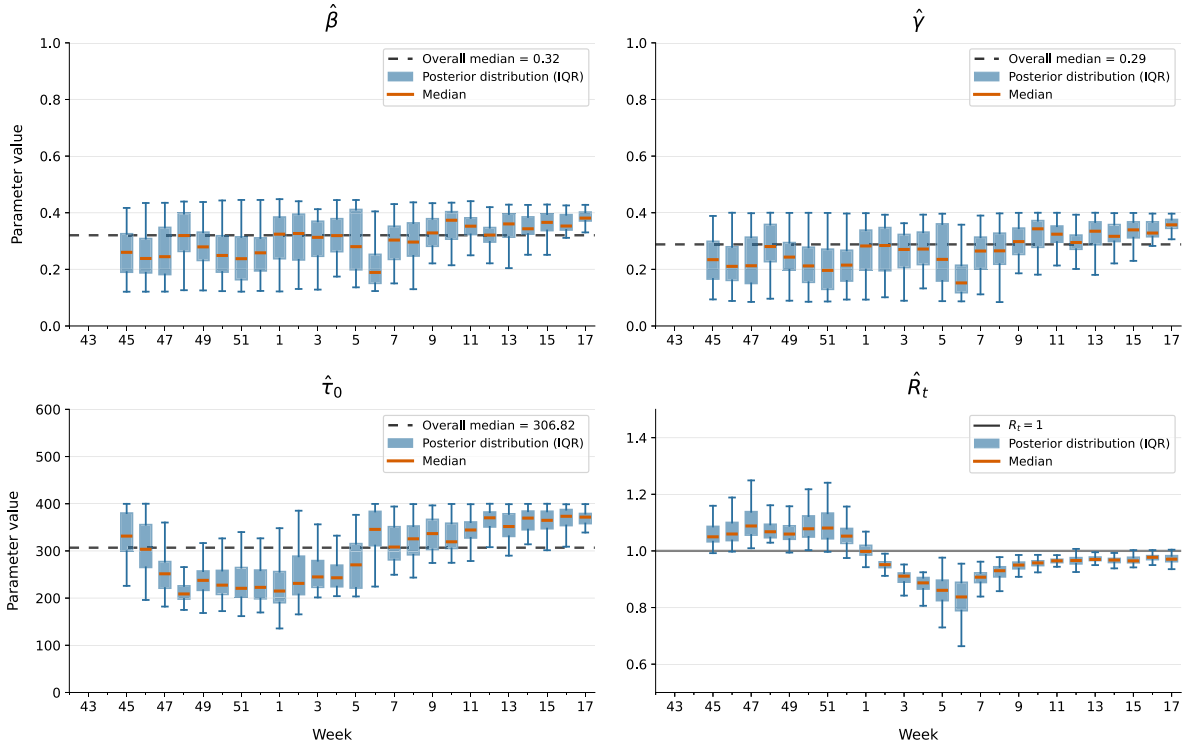

Fig. 2. Behavior of the SIR-based model parameters estimated via MCMC — 2023–2024 Italian seasonal influenza. Posterior distributions of the inferred parameters $\hat{\beta}$, transmission rate, $\hat{\gamma}$, recovery rate, $\hat{\tau}_0$, observation alignment time, and $\hat{R}_t$, effective reproduction number, across epidemic weeks. Each box represents the interquartile range (IQR, 25th–75th percentile), with whiskers extending to the 5th and 95th percentiles; the orange horizontal line within each box denotes the weekly posterior median. The dashed line in the panels for $\hat{\beta}, \hat{\gamma}, \hat{\tau}_0$ indicates the overall posterior median across all weeks. For $\hat{R}_t$, the solid horizontal line marks the threshold $R_t = 1$, separating epidemic growth ($R_t > 1$) from decline ($R_t < 1$).

4.2. Forecasting results

From the results derived from the estimation of β, γ, and τ_0 parameters via MCMC, we obtain Markov chains, referred to each time window of observations (i.e., round). Hence, we perform probabilistic forecasting in the form of quantiles with prediction intervals based on model outcomes, on each evaluation window, defined as the union of the observations time window and the forecast time window. In particular, for each MCMC chain we select 1000 tail samples to draw from the posterior, then we evaluate our model on these samples, computing the quantiles as measures of uncertainty.

4.2.1. Four-weeks-ahead forecasting

In a first set of experiments, we generate forecasts four weeks ahead, according to the Influcast Hub forecasting standards (InfluCast, 2025). This means that, similarly to the parameter inference experiments, we update our forecasts each time the epidemic curve is updated with a new observation, removing the oldest one within a time window of five data points. We consider a fixed size of four weeks for the forecast time window. Note that for the last four simulations, we forecast on a shorter time window since we do not have any other future observations to compare with.

Figs. 4 and 5 show the results for nine representative time windows and for the 2023–2024 and 2024–2025 seasons, respectively. Note that the error bars and the mean results of SIR-INN start from the last reported incidence observation since our methodology based on PINN produces uncertainty throughout the evaluation window. The complete results obtained considering all the time windows are shown in Appendix A, Figure A.8, for the 2023–2024 season, and in Figure A.9, for the 2024–2025 season.

In the 2023–2024 influenza season (Fig. 4), we can see that SIR-INN is fairly accurate in the initial phase of the outbreak, despite the inherent challenges associated with the limited number of observations available at that stage. However, the forecast tends to underestimate the observed incidence in this initial period, as shown by the forecasts at weeks 46 and 48. As the peak time approaches in weeks 51 and 52, we can see that the error bars become larger and the 90% percentile interval contains the current data and almost all of the following four observations. In particular, at week 50 the model almost correctly identifies the occurrence of the peak, which is the most critical and challenging phase, even if the actual incidence value is slightly underestimated. The declining epidemic phase that follows the peak is well predicted by SIR-INN, in its initial phase (week 2). However, as the outbreak fades, our model rapidly begins to underestimate the real observations, forecasting a swift extinction of the newly reported cases. The forecasting accuracy increases again in the final phase of the season, as the curve approaches the baseline values once more (weeks 8–10).

The results for the 2024–2025 season (Fig. 5) show an increased forecasting accuracy, compared to the previous year. The early outbreak phase, the peak time, and the decreasing phase are consistently estimated with good accuracy. In particular, the crucial forecast of the peak time is improved with respect to the results of the 2023–2024 season. The peak incidence is slightly underestimated when projected from week 2, and slightly overestimated at week 4. The predicted peak time in both cases is nearly identical to the real one, and the general trend is well captured by our model. However, at week 50, we can see that our model is unable to accurately predict the later increase in the ILI incidence curve. Nevertheless, it seems to be able to predict the constant trend of the subsequent two weeks and with 90% interval a possible increasing of the incidence function. Indeed, by week 52, the model appears surprisingly to infer that the current peak may represent only a relative infection peak, thus anticipating a potential subsequent rapid increase in incidence. For the peculiarity of

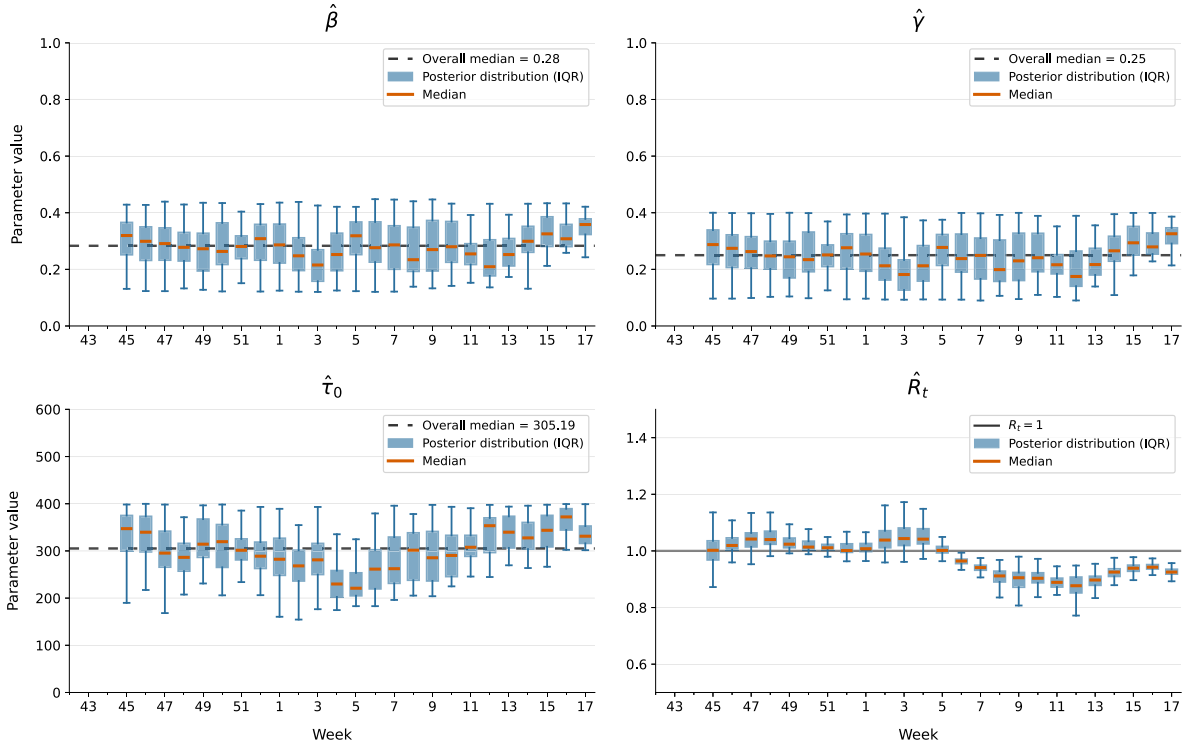

Fig. 3. Behavior of the SIR-based model parameters estimated via MCMC — 2024–2025 Italian seasonal influenza. Posterior distributions of the inferred parameters $\hat{\beta}$, transmission rate, $\hat{\gamma}$, recovery rate, $\hat{\tau}_0$, observation alignment time, and $\hat{R}_t$, effective reproduction number, across epidemic weeks. Each box represents the interquartile range (IQR, 25th–75th percentile), with whiskers extending to the 5th and 95th percentiles; the orange horizontal line within each box denotes the weekly posterior median. The dashed line in the panels for $\hat{\beta}, \hat{\gamma}, \hat{\tau}_0$ indicates the overall posterior median across all weeks. For $\hat{R}_t$, the solid horizontal line marks the threshold $R_t = 1$, separating epidemic growth ($R_t > 1$) from decline ($R_t < 1$).

this first less pronounced peak of infection, weeks 49–52 are the most challenging predictions that are still in line with most of the forecasts reported on the Influcast platform (InfluCast, 2025). Finally, our model effectively captures the decreasing phase of the outbreak. This clearly emerges from the forecasts at weeks 6 and 8, where the SIR-INNN mean value, represented by the green line, almost perfectly fits the unseen observations, and the 50% percentile intervals contain almost all the data. Only the latest observations are slightly underestimated by our model, as shown by the results at week 10.

Table 1 describes the forecast performance of our model compared to the baseline, which is used as a reference model in Fiandrino et al. (2025), for both seasons. The results are obtained from national-level four-weeks-ahead forecasting using several metrics, aggregating forecasts from all rounds and all the horizons. The number of rounds represents the number of simulations considered for evaluation. In line with the Influcast hub challenge (InfluCast, 2025), we considered a total number of 20 rounds for the 2023–2024 season, starting the first four-weeks-ahead forecasting from the 46th week of 2023 and finishing with the last forecast simulation at week 13th of 2024. For the 2024–2025 season, we began to evaluate the performances of the simulation whose last observation corresponds to the 45th week of 2024. The number of horizons considered for the evaluation is four, corresponding to the size of the forecast time window. A detailed description of the evaluation metrics is reported in Section 3.4.

From the results shown in Table 1 we can see that, in both seasons, SIR-INN consistently outperforms the baseline in terms of point forecast accuracy, as confirmed by lower MAE and WIS values. Specifically, the MAE of the median forecast decreases from 2.68 to 2.28 in 2023–2024 and from 2.54 to 1.98 in 2024–2025, confirming an improvement in forecast precision for the 2024–2025 season compared to 2023–2024. Furthermore, the MAE of the median forecast, the WIS values, and the coverages of the 2024–2025 season are notably optimal with respect to the baseline as well as the previous season, meaning that the accuracy of the median forecast and the prediction intervals in bounding the actual observations is also greater compared to the baseline model. Moreover, although the coverage values remain somewhat below the ideal nominal levels, their increase suggests that the model exhibited improved calibration by transitioning from the previous influenza season to the most recent.

Table 1
Forecasting performance of the SIR-INN model, compared to the baseline. The results are obtained from the four-weeks-ahead forecasting at national level, across the two influenza seasons, in terms of MAE of the median forecast, WIS, 50% and 90% coverage. Bold values indicate the best performance for each metric within the corresponding season.

Model	Influenza season	N. of rounds	MAE	WIS	Coverage 50%	Coverage 90%
Baseline	2023/2024	20	2.68	1.90	**0.49**	**0.66**
SIR-INN	2023/2024	20	**2.28**	**1.63**	0.11	0.39
Baseline	2024/2025	21	2.54	1.71	0.14	0.52
SIR-INN	2024/2025	21	**1.98**	**1.36**	**0.24**	**0.58**

4.2.2. *Comparison with Influcast models*

Since we performed four-week forecast simulations of Italian seasonal influenza 2023–2024 according to the requirements of the Influcast challenge (InfluCast, 2025), we implemented the same validation metrics adopted in Fiandrino et al. (2025) to evaluate the accuracy of our model relative to those of the other teams. Table 2 presents the aggregated performance, for all rounds and horizons, of all models under consideration, as in Fiandrino et al. (2025), with the addition of our model's performance (last row). Regarding the performance evaluation of all the Influcast models, the relative values of the MAE

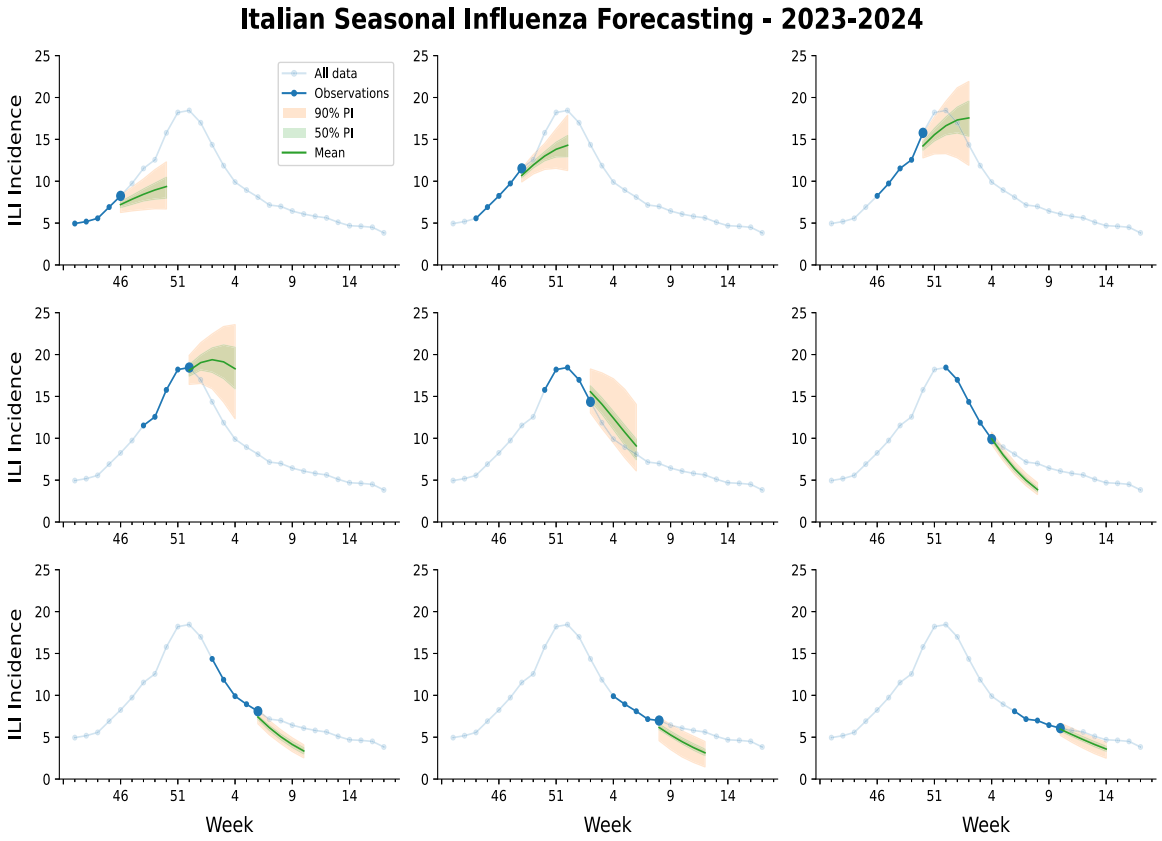

Fig. 4. Seasonal influenza forecasting four weeks ahead via SIR-INN. The ground truth data, in blue, represent the weekly observations at national level of the 2023–2024 seasonal influenza. The highlighted blue data represent the time window of observations, with fixed length of 5, from which the SIR-based PINN infers the epidemic parameters and then forecasts on the following four weeks ahead. The mean-model forecasts are represented with green lines and the 50% and 90% percentile intervals are shaded, respectively, in green and orange error bars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the median forecast and the WIS are computed by comparing pairs of models that participate in the same forecast round. These values are then normalized by dividing by the corresponding value of the baseline model. In addition, these values incorporate both national-level seasonal influenza predictions and aggregated regional forecasts, in contrast to national-level results reported in Section 4.2.1. In this way, each individual model's result accounts for both regional and national aggregated outcomes of all other models, and it is expressed relative to the baseline. Consequently, relative values below 1 indicate better performance than the baseline, while values above 1 indicate worse performance. The ensemble model is obtained by combining all the model forecasts, and it consistently outperforms the baseline model on all results, as detailed in Fiandrino et al. (2025), which also provides further information on the single models and evaluation metrics. Regarding our model results, the relative MAE of the median forecast and the WIS are computed by comparison with the results of all other models and normalized by the corresponding baseline values. Note that SIR-INN's relative results are evaluated with respect to the other models, whereas the reverse does not hold, as we did not officially participate in the challenge. Consequently, the ensemble model also does not include our model's predictions. Moreover, the values related to our model are all derived from national-level predictions, rather than regional ones.

Taking into account the results of our model, outlined in the last row of Table 2, the relative MAE and WIS values are close to the baseline values (1.03 and 0.98, respectively), resulting slightly better than the baseline in terms of relative WIS, and the overall performance remains comparable to that of the other models. Specifically, in terms of relative MAE and WIS, our model outperforms two other models: *Mechanistic-2* (1.86, 1.64), *Semi-mechanistic-1* (1.80, 2.14). Additionally, it outperforms *Statistical-1* (1.09) and *Semi-mechanistic-2* (1.04) in term of WIS. Regarding the coverage, although the values are below the ideal nominal levels, our model still compares well with the others. It performs better than *Mechanistic-2* (0.04), *Semi-mechanistic-1* (0.06), and similarly to *Statistical-2* (0.11), for the coverage 50%. Furthermore, in terms of coverage 90%, SIR-INN improves over *Mechanistic-4* (0.37), *Semi-mechanistic-1* (0.27), *Semi-mechanistic-2* (0.30), and *Statistical-2* (0.31). In summary, our model shows overall performance in line with the participating Influcast models (InfluCast, 2025), usually ranking above the bottom three. This suggests that its implementation could positively contribute to the prediction of future seasonal influenza epidemics when included in the ensemble forecasts.

4.2.3. *Ten-weeks-ahead forecasting*

In addition, we analyze the performances of our SIR-INN on long-term forecasting. The experimental setup is identical to the one implemented for the short-term forecasts, with the only difference being a different size for the forecasting time window. In this case, we select a range of ten weeks instead of four weeks.

The results for the 2023–2024 season obtained considering four consecutive time windows of observations during the peak phase are shown in Fig. 6. For the results of the 2024–2025 season, see Figure A.10, in Appendix A.

The results of the 2023–2024 influenza season, shown in Fig. 6, illustrate how our model is able to partially capture the epidemic trend beyond the short-term horizon in this delicate phase, although without providing a fully accurate estimate. In particular, at week 50, two weeks before the peak, all the future ten unknown incidence values

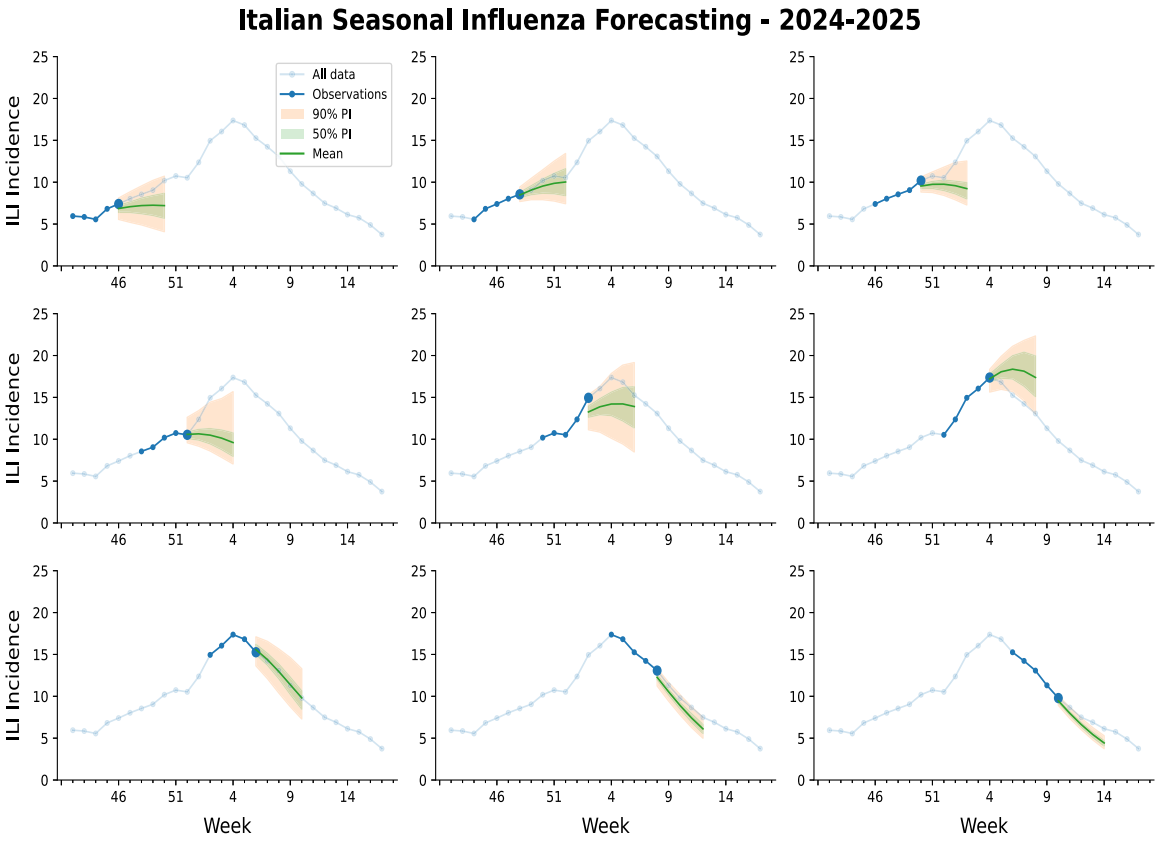

Fig. 5. Seasonal influenza forecasting four weeks ahead via SIR-INN. The ground truth data, in blue, represent the weekly observations at national level of the 2024–2025 seasonal influenza. The highlighted blue data represent the time window of observations, with fixed length of 5, from which the SIR-based PINN infers the epidemic parameters and then forecasts on the following four weeks ahead. The mean-model forecasts are represented with green lines and the 50% and 90% percentile intervals are shaded, respectively, in green and orange error bars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Forecasting performance of the SIR-INN model compared to Influcast models. The results are obtained from the four-weeks-ahead forecasting at national level in terms of relative MAE of the median forecast, relative WIS, 50% and 90% coverage.

Model	N. of rounds	Relative MAE	Relative WIS	Coverage 50%	Coverage 90%
Mechanistic-1	19	0.56 (1st)	0.54 (1st)	0.75	0.89 (1st)
Mechanistic-2	20	1.86	1.64	0.04	0.49
Mechanistic-3	20	0.58 (2nd)	0.58 (3rd)	0.47 (3rd)	0.70 (3rd)
Mechanistic-4	19	0.73	0.73	0.16	0.37
Semi-mechanistic-1	16	1.80	2.14	0.06	0.27
Semi-mechanistic-2	14	0.84	1.04	0.19	0.30
Statistical-1	20	0.99	1.09	0.17	0.43
Statistical-2	19	0.77	0.85	0.11	0.31
Ensemble	20	0.59 (3rd)	0.57 (2nd)	0.51 (1st)	0.80 (2nd)
Baseline	20	1.00	1.00	0.49 (2nd)	0.66
SIR-INN	20	1.03	0.98	0.11	0.39

fall within the 90% percentile interval. In all scenarios depicted in Fig. 6, the forecast curve exhibits a broader shape relative to the curve derived from interpolated observations, tending to underestimate the data points before the peak and, conversely, overestimate the observations once the peak has passed. The peak time is consistently predicted later than the actual one, although the discrepancy remains within a maximum of four weeks. Notably, shortly after the epidemic peak (week 1), the model rapidly adapts to the changing dynamics, being able to provide accurate longer-term forecasts of the declining phase of the incidence curve, with relatively narrow prediction intervals.

Even in the case of long-term forecasting, the results presented in Appendix A, Figure A.10, are better for the 2024–2025 season than for the 2023–2024 one. The 90% prediction interval contains almost all observations for the peak scenarios and, at week 5, immediately after the peak, almost all the observations fall within the narrow 50% percentile interval, proving again the ability of the model to quickly adjust its predictions. In addition, the model's ability to predict the maximum ILI incidence increases as the epidemic peak approaches.

5. Discussion

In this work, we have introduced a novel framework for epidemic forecasting, namely SIR-INN, based on PINNs. Specifically, we endowed

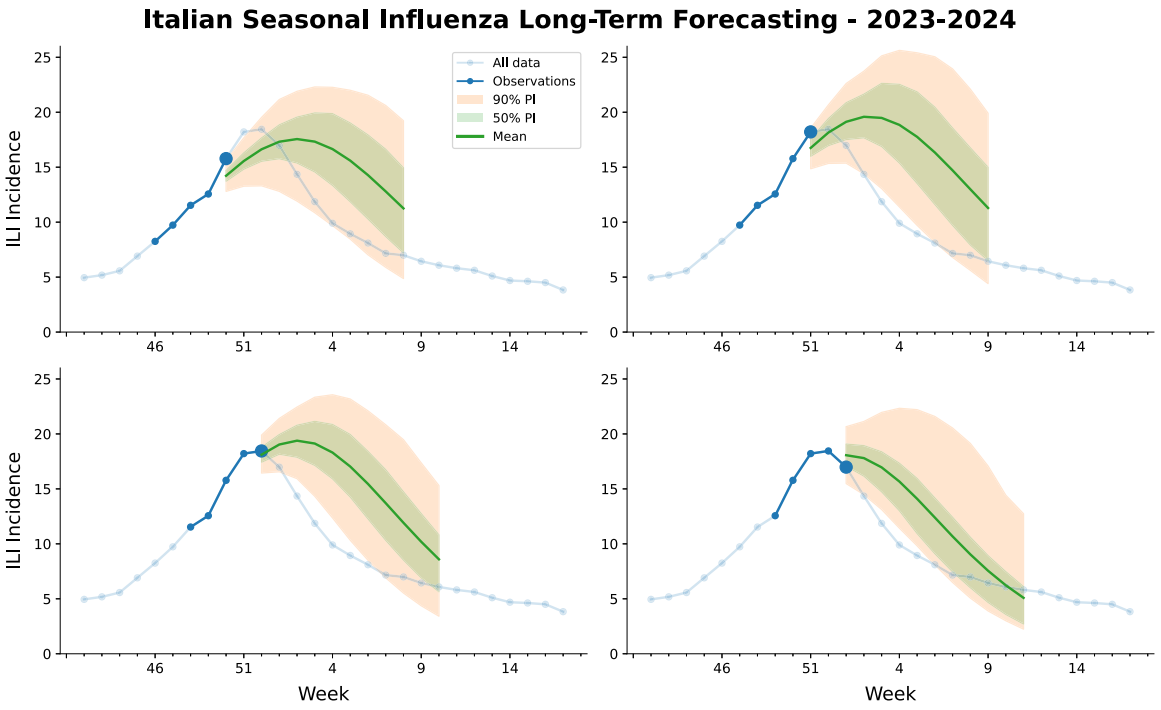

Fig. 6. Seasonal influenza forecasting ten weeks ahead via SIR-INN. The ground truth data, in blue, represent the weekly observations at national level of the 2023–2024 seasonal influenza. The highlighted blue data represent the time window of observations, with fixed length of 5, from which the SIR-based PINN infers the epidemic parameters and then forecasts on the following ten weeks ahead. The mean-model forecasts are represented with green lines and the 50% and 90% percentile intervals are shaded, respectively, in green and orange error bars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a single neural network with prior epidemiological knowledge, leveraging the well-known SIR compartmental model with constant transition rates. The neural network has been trained only once on temporal-epidemic domains close to realistic epidemic scenarios. In this way, starting from a limited set of noisy observations, our PINN is able to infer, via MCMC, the parameters that characterize a SIR-based epidemic dynamics. Following this inference step, we used our PINN to forecast the dynamics of epidemics in future time windows.

We used national data from the Italian seasonal influenza surveillance, provided by the Italian National Institute of Health (ISS) (Istituto Superiore di Sanità, 2020), as done by previous studies on influenza forecasting (Zhang et al., 2015; Perrotta et al., 2017; Brownstein et al., 2017). Our results, obtained from a direct application of the framework to national data from the influenza seasons 2023–2024 and 2024–2025, confirm the good forecasting abilities of our hybrid methodology.

Despite differences in weekly observations and predictions of associated parameters, our results indicate that both seasonal influenza outbreaks showed similar epidemiological characteristics in terms of transmissibility and average infectious period. Such epidemiological features are shown to be underpinned by a SIR model, with disease transmission rate and removal rate values piecewise constant close to 0.3 and an average basic reproduction number R_0 of approximately 1.135. Furthermore, the temporal evolution of the effective reproduction number R_t confirmed consistent epidemic dynamics across both seasons, with R_t exceeding the threshold of 1 during the growth phase and declining below it at epidemiologically plausible turning points, coherently with the observed ILI incidence curves.

We then performed four-week-ahead forecasting simulations based on inferred parameters, employing the same validation metrics adopted by Fiandrino et al. (2025). From the qualitative results, we can see that for the 2023–2024 influenza season, and especially for the 2024–2025 season, the mean forecasting solutions during the early period and the decreasing stage are often fairly accurate. This is particularly relevant considering that providing accurate forecasts during the early phase of an outbreak is challenging, primarily due to the limited number of available observations, which are often noisy. Even more challenging is the estimation of the incidence peak. Although the model does not achieve high accuracy in forecasting the season peak, it appears to anticipate its occurrence reasonably well. In general, Figs. 4 and 5 suggest that our model is more effective in capturing the temporal pattern of the influenza season 2024–2025. This impression is also confirmed by the results in Table 1, which reports the MAE of the median forecast, WIS, and 50% and 90% coverage for both seasons. In general, our SIR-INN framework appears to outperform the baseline used as a reference model in Fiandrino et al. (2025), particularly in terms of MAE and WIS for both the seasons, and also in terms of coverages for the 2024–2025 season. This makes our model a possible competitive alternative to the baseline for seasonal influenza forecasting. These findings are also supported by the results of the seasonal influenza forecast for 2023–2024 presented in Table 2, where all models participating in the Influcast four weeks ahead forecast challenge (InfluCast, 2025) were compared to ours. These results show that the performances of our model are comparable to those of current state-of-the-art models. In particular, the SIR-INN performs better in terms of the relative WIS, thereby demonstrating the accuracy of the prediction with respect not only to the median outcome but also to the prediction intervals. In summary, these results suggest that our framework could meaningfully enhance the forecasting of successive seasonal influenza epidemics, increasing the ensemble's predictive performance.

In addition, in Section 4.2.3, we analyze the performance of our model in long-term forecasting, specifically ten weeks in advance. In Fig. 6 and Figure A.10, we can see that although the influenza season is in its most challenging phase — peak estimation — and only a few observations are available, our model is able to partially capture the epidemic trend. This is particularly relevant considering that long-term forecasting often fails when using fully data-driven and neural network-based approaches (Rodríguez et al., 2024; Ye et al., 2025).

To further contextualize these results, we conducted an additional comparison with a classical mechanistic SIR model and a hybrid PINN-ODE pipeline, reported in detail in Appendix F. The comparison shows that the proposed SIR-INN framework achieves competitive performance across all metrics, and in particular outperforms both the fully mechanistic and hybrid mechanistic versions in terms of MAE and WIS, while maintaining a significantly lower computational cost. These findings suggest that the relative advantages of fully mechanistic, fully neural, and hybrid approaches may depend on the specific forecasting context, including the epidemic phase, the forecast horizon, and the evaluation metric.

Together, these findings align with what is increasingly emphasized in the forecast epidemiology literature. Adopting a hybrid approach that integrates the mechanistic understanding of disease transmission with the learning and generalization capabilities of neural networks appears to be a promising direction to improve forecast performance (Rodríguez et al., 2024; Ye et al., 2025; Qian et al., 2025). The developed framework allows us to benefit from both approaches, thus maximizing forecasting capabilities in both short-term and long-term predictions. On the one hand, to preserve the underlying dynamic structure of influenza transmission, we use the simplicity of the SIR model, supported by theoretical studies, with fixed epidemiological parameters. On the other hand, the neural network component allows us to exploit data-driven patterns, improving the flexibility and generalization ability of the model (Raissi et al., 2019; Rodríguez et al., 2024). Furthermore, we take additional advantage of the hybrid nature of the framework by implementing an MCMC method, separately from the neural network, to automatically capture the temporal variability of the SIR parameters through independent local estimation at each forecasting round. While this rolling re-estimation allows the model to rapidly adapt to the changing dynamics, it also implies that the inferred parameters should be interpreted as locally consistent approximations rather than global epidemiological constants characterizing the full season. Our results suggest that even complex epidemic dynamics may reflect a simple, shared mathematical structure that deserves consideration in purely Machine Learning (ML)-based forecasting models (Ye et al., 2025).

The resulting framework, despite the combination of diverse and potentially complex hybrid components, remarkably preserves simplicity, explainability, efficiency, and cost-effectiveness. Our SIR-INN consists of a single neural network trained only once on synthetic data, in contrast to the majority of machine learning (and hybrid) approaches for epidemic forecasting (Kharazmi et al., 2021; Rodríguez et al., 2023; Millevoi et al., 2024; Qian et al., 2025). This makes the method highly flexible and generalizable. The neural network does not require retraining each time a new observation is available or when a new influenza season occurs. The SIR-INN framework also allows us to obtain parameter distributions independently of the neural network training, thereby eliminating the need for retraining it while allowing for the quantification of forecast uncertainty. Performing Uncertainty Quantification (UQ) is essential in epidemiological forecasting, as it contributes to model robustness and supports more reliable decision-making in public health. Given the inherent variability and noise of epidemiological data, capturing forecast uncertainty is necessary to avoid overconfidence and misinterpretation of the model output (Funk et al., 2019; Bracher et al., 2021; Sherratt et al., 2023). However, in neural network-based approaches where the output is a single trajectory, incorporating this uncertainty remains challenging and is therefore still rarely implemented in the current literature (Linka et al., 2022; Shi et al., 2025). Our work can be seen as an initial and promising attempt toward incorporating uncertainty into forecasting within this class of hybrid approaches. However, further work is still needed in this direction. The model recurrently exhibits overconfidence, resulting in overly tight percentile intervals and consequently underperforming in terms of coverage. This behavior is less pronounced around the peak phase, where wider predictive intervals reflect the inherent difficulty of peak estimation, ensuring greater reliability. Nevertheless, this represents a relevant shortcoming that should be taken into account when interpreting the probabilistic forecasts.

In summary, our proposed framework, namely SIR-INN, provides an efficient and accurate surrogate for learning epidemiological models, enabling less costly forecasting while maintaining both generalizability and accuracy, even in long-term predictions. Given its simple design, the model offers valuable insight into more efficient and interpretable implementations of hybrid models for forecasting other dynamical systems with uncertainty quantification. A natural and immediate extension of this work is its application to other SIR-based epidemic models. In fact, it would only require retraining the neural network once after selecting parameter ranges that are better aligned with the specific disease under study. Furthermore, an interesting direction for generalization is the prediction of epidemics coupled with human behavior changes. In such cases, the individual approaches — the mechanistic modeling and the neural network component — remain limited by their partial perspectives on the system, failing to benefit from one another during the forecasting process. In conclusion, our work represents an initial step toward an increased integration of AI tools with mechanistic epidemic models to improve forecasting approaches and contributes to the establishment of new standards in hybrid modeling.

CRediT authorship contribution statement

Martina Rama: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Gabriele Santin:** Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Giulia Cencetti:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Michele Tizzoni:** Writing – review & editing, Validation, Supervision, Conceptualization. **Bruno Lepri:** Writing – review & editing, Validation, Supervision, Conceptualization.

Code availability

The code to reproduce the results of this paper is publicly available at https://github.com/martina-rama/SIR-INN.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Bruno Lepri acknowledges the support of the PNRR ICSC National Research Centre for High Performance Computing, Big Data, and Quantum Computing (CN00000013), under the NRRP MUR program funded by NextGenerationEU.

Appendix. Supplementary Data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epidem.2026.100919.

Data availability

Link to data sources and code repository are present in the manuscript.

References

Berkhahn, S., Ehrhardt, M., 2022. A physics-informed neural network to model COVID-19 infection and hospitalization scenarios. Adv. Contin. Discret. Model. 2022 (1), 61.

Bertaglia, G., Lu, C., Pareschi, L., Zhu, X., 2022. Asymptotic-preserving neural networks for multiscale hyperbolic models of epidemic spread. Math. Models Methods Appl. Sci. 32 (10), 1949–1985.

Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I.C.-H., Hickmann, K.S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., et al., 2016. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. BMC Infect. Dis. 16, 1–10.

Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., Finelli, L., 2014. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: A systematic review of the literature. BMC Infect. Dis. 14 (1), 1–20.

Bosse, N.I., Gruson, H., Cori, A., van Leeuwen, E., Funk, S., Abbott, S., 2022. Evaluating forecasts with scoringutils in r. arXiv preprint arXiv:2205.07090.

Bracher, J., Ray, E.L., Gneiting, T., Reich, N.G., 2021. Evaluating epidemic forecasts in an interval format. PLoS Comput. Biol. 17 (2), e1008618.

Brownstein, J.S., Chu, S., Marathe, A., Marathe, M.V., Nguyen, A.T., Paolotti, D., Perra, N., Perrotta, D., Santillana, M., Swarup, S., et al., 2017. Combining participatory influenza surveillance with modeling and forecasting: Three alternative approaches. JMIR Public Health Surveill. 3 (4), e7344.

Cai, S., Mao, Z., Wang, Z., Yin, M., Karniadakis, G.E., 2021. Physics-informed neural networks (PINNs) for fluid mechanics: A review. Acta Mech. Sin. 37 (12), 1727–1738.

Carrat, F., Vergu, E., Ferguson, N.M., Lemaitre, M., Cauchemez, S., Leach, S., Valleron, A.-J., 2008. Time lines of infection and disease in human influenza: A review of volunteer challenge studies. Am. J. Epidemiol. 167 (7), 775–785.

Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F., 2022. Scientific machine learning through physics–informed neural networks: Where we are and what's next. J. Sci. Comput. 92 (3), 88.

Del Valle, S.Y., McMahon, B.H., Asher, J., Hatchett, R., Lega, J.C., Brown, H.E., Leany, M.E., Pantazis, Y., Roberts, D.J., Moore, S., et al., 2018. Summary results of the 2014–2015 DARPA Chikungunya challenge. BMC Infect. Dis. 18, 1–14.

Desai, A.N., Kraemer, M.U., Bhatia, S., Cori, A., Nouvellet, P., Herringer, M., Cohn, E.L., Carrion, M., Brownstein, J.S., Madoff, L.C., et al., 2019. Real-time epidemic forecasting: challenges and opportunities. Health Secur. 17 (4), 268–275.

Evans, L.C., 2022. Partial Differential Equations. Vol. 19, American Mathematical Society.

Fiandrino, S., Bizzotto, A., Guzzetta, G., Merler, S., Baldo, F., Valdano, E., Urdiales, A.M., Bella, A., Celino, F., Zino, L., et al., 2025. Collaborative forecasting of influenza-like illness in Italy: The influcast experience. Epidemics 100819.

Folland, G.B., 1995. Introduction to Partial Differential Equations. Vol. 102, Princeton University Press.

Fox, S.J., Kim, M., Meyers, L.A., Reich, N.G., Ray, E.L., 2024. Optimizing disease outbreak forecast ensembles. Emerg. Infect. Dis. 30 (9), 1967.

Funk, S., Camacho, A., Kucharski, A.J., Lowe, R., Eggo, R.M., Edmunds, W.J., 2019. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. PLoS Comput. Biol. 15 (2), e1006785.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Heffernan, J.M., Smith, R.J., Wahl, L.M., 2005. Perspectives on the basic reproductive ratio. J. R. Soc. Interface 2 (4), 281–293.

Holcomb, K.M., Mathis, S., Staples, J.E., Fischer, M., Barker, C.M., Beard, C.B., Nett, R.J., Keyel, A.C., Marcantonio, M., Childs, M.L., et al., 2023. Evaluation of an open forecasting challenge to assess skill of West Nile virus neuroinvasive disease prediction. Parasites Vectors 16 (1), 11.

InfluCast, 2025. InfluCast - web platform. URL https://influcast.org/. (Accessed 26 April 2025).

Istituto Superiore di Sanità, 2020. Sistema di sorveglianza integrata RespiVirNet, istituto superiore di sanità. https://www.epicentro.iss.it/influenza/respivirnet. (Accessed 26 April 2025).

Johansson, M.A., Apfeldorf, K.M., Dobson, S., Devita, J., Buczak, A.L., Baugher, B., Moniz, L.J., Bagley, T., Babin, S.M., Guven, E., et al., 2019. An open challenge to advance probabilistic forecasting for dengue epidemics. Proc. Natl. Acad. Sci. 116 (48), 24268–24274.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nat. Rev. Phys. 3 (6), 422–440.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Ser. A, Contain. Pap. A Math. Phys. Character 115 (772), 700–721.

Kharazmi, E., Cai, M., Zheng, X., Zhang, Z., Lin, G., Karniadakis, G.E., 2021. Identifiability and predictability of integer-and fractional-order epidemiological models using physics-informed neural networks. Nat. Comput. Sci. 1 (11), 744–753.

Lagergren, J.H., Nardini, J.T., Baker, R.E., Simpson, M.J., Flores, K.B., 2020. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. PLoS Comput. Biol. 16 (12), e1008462.

Lauer, S.A., Brown, A.C., Reich, N.G., 2021. Infectious disease forecasting for public health. Popul. Biol. Vector-Borne Dis. 45–68.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Linka, K., Schäfer, A., Meng, X., Zou, Z., Karniadakis, G.E., Kuhl, E., 2022. Bayesian physics informed neural networks for real-world nonlinear dynamical systems. Comput. Methods Appl. Mech. Engrg. 402, 115346.

Lopez, V.K., Cramer, E.Y., Pagano, R., Drake, J.M., O'Dea, E.B., Adee, M., Ayer, T., Chhatwal, J., Dalgic, O.O., Ladd, M.A., et al., 2024. Challenges of COVID-19 case forecasting in the US, 2020–2021. PLoS Comput. Biol. 20 (5), e1011200.

Madden, W.G., Jin, W., Lopman, B., Zufle, A., Dalziel, B., E. Metcalf, C.J., Grenfell, B.T., Lau, M.S., 2024. Deep neural networks for endemic measles dynamics: Comparative analysis and integration with mechanistic models. PLoS Comput. Biol. 20 (11), e1012616.

Mao, Z., Jagtap, A.D., Karniadakis, G.E., 2020. Physics-informed neural networks for high-speed flows. Comput. Methods Appl. Mech. Engrg. 360, 112789.

Miles, P.R., 2019. Pymcmcstat: A python package for Bayesian inference using delayed rejection adaptive Metropolis. J. Open Source Softw. 4 (38), 1417. http://dx.doi.org/10.21105/joss.01417.

Millevoi, C., Pasetto, D., Ferronato, M., 2024. A physics-informed neural network approach for compartmental epidemiological models. PLoS Comput. Biol. 20 (9), e1012387.

Neal, R.M., 1993. Probabilistic inference using Markov chain Monte Carlo methods.

Ning, X., Guan, J., Li, X.-A., Wei, Y., Chen, F., 2023a. Physics-informed neural networks integrating compartmental model for analyzing COVID-19 transmission dynamics. Viruses 15 (8), 1749.

Ning, X., Jia, L., Wei, Y., Li, X.-A., Chen, F., 2023b. Epi-DNNs: Epidemiological priors informed deep neural networks for modeling COVID-19 dynamics. Comput. Biol. Med. 158, 106693.

Nishiura, H., Chowell, G., 2009. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In: Mathematical and Statistical Estimation Approaches in Epidemiology. Springer, pp. 103–121.

Nixon, K., Jindal, S., Parker, F., Reich, N.G., Ghobadi, K., Lee, E.C., Truelove, S., Gardner, L., 2022. An evaluation of prospective COVID-19 modelling studies in the USA: From data to science translation. Lancet Digit. Health 4 (10), e738–e747.

O'Dea, E.B., Drake, J.M., 2022. A semi-parametric, state-space compartmental model with time-dependent parameters for forecasting COVID-19 cases, hospitalizations and deaths. J. R. Soc. Interface 19 (187), 20210702.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035, URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Perrotta, D., Tizzoni, M., Paolotti, D., 2017. Using participatory web-based surveillance data to improve seasonal influenza forecasting in Italy. In: Proceedings of the 26th International Conference on World Wide Web. pp. 303–310.

Qian, Y., Zhang, K., Marty, E., Basu, A., O'dea, E.B., Wang, X., Fox, S.J., Rohani, P., Drake, J.M., Li, H., 2025. Physics-informed deep learning for infectious disease forecasting. J. R. Soc. Interface 22 (232).

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.

Ray, E.L., Wang, Y., Wolfinger, R.D., Reich, N.G., 2025. Flusion: Integrating multiple data sources for accurate influenza predictions. Epidemics 50, 100810.

Reich, N.G., McGowan, C.J., Yamana, T.K., Tushar, A., Ray, E.L., Osthus, D., Kandula, S., Brooks, L.C., Crawford-Crudell, W., Gibson, G.C., et al., 2019. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the US. PLoS Comput. Biol. 15 (11), e1007486.

RespiVirNet, 2020, 2020. RespiVirNet data extracts. URL https://github.com/fbranda/influnet. (Accessed 26 April 2025).

Richey, M., 2010. The evolution of Markov chain Monte Carlo methods. Am. Math. Mon. 117 (5), 383–413.

Rodríguez, A., Cui, J., Ramakrishnan, N., Adhikari, B., Prakash, B.A., 2023. EINNs: Epidemiologically-informed neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37, pp. 14453–14460.

Rodríguez, A., Kamarthi, H., Agarwal, P., Ho, J., Patel, M., Sapre, S., Prakash, B.A., 2024. Machine learning for data-centric epidemic forecasting. Nat. Mach. Intell. 6 (10), 1122–1131.

Shaier, S., Raissi, M., Seshaiyer, P., 2021. Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease informed neural networks. arXiv preprint arXiv:2110.05445.

Sherratt, K., Gruson, H., Johnson, H., Niehus, R., Prasse, B., Sandmann, F., Deuschel, J., Wolffram, D., Abbott, S., Ullrich, A., et al., 2023. Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. Elife 12, e81916.

Shi, Y., Wei, P., Feng, K., Feng, D.-C., Beer, M., 2025. A survey on machine learning approaches for uncertainty quantification of engineering systems. Mach. Learn. Comput. Sci. Eng. 1 (1), 1–39.

Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., Vespignani, A., et al., 2018. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics 22, 13–21.

Waskom, M.L., 2021. Seaborn: Statistical data visualization. J. Open Source Softw. 6 (60), 3021. http://dx.doi.org/10.21105/joss.03021.

Wolffram, D., Abbott, S., An der Heiden, M., Funk, S., Günther, F., Hailer, D., Heyder, S., Hotz, T., van de Kassteele, J., Küchenhoff, H., et al., 2023. Collaborative nowcasting of COVID-19 hospitalization incidences in Germany. PLoS Comput. Biol. 19 (8), e1011394.

Ye, Y., Pandey, A., Bawden, C., Sumsuzzman, D.M., Rajput, R., Shoukat, A., Singer, B.H., Moghadas, S.M., Galvani, A.P., 2025. Integrating artificial intelligence with mechanistic epidemiological modeling: A scoping review of opportunities and challenges. Nat. Commun. 16 (1), 1–18.

Zhang, Q., Gioannini, C., Paolotti, D., Perra, N., Perrotta, D., Quaggiotto, M., Tizzoni, M., Vespignani, A., 2015. Social data mining and seasonal influenza forecasts: the FluOutlook platform. In: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part III 15. Springer, pp. 237–240.