APPENDIX: ONLINE RESOURCES

A RELATED SURVEYS

Over the past two decades, the development of infectious disease modeling has been thoroughly reviewed by several researchers. For instance, Grassly and Fraser [237] analyzed the connection between mathematical hypotheses and modeling design in the context of infectious disease transmission, summarizing key contributions to the field. In recent years, some surveys have focused on models for specific types of infectious diseases, such as malaria [246], dengue [229], influenza [231, 249], and COVID-19 [227, 232]. Thereinto, Mandal et al. [246] and Andraud et al. [229] provided insights into deterministic mathematical approaches (also known as mechanism-based models) for modeling malaria and dengue transmission, respectively, linking the development of these models to the propagation of theories on disease spread. These reviews have covered several areas, such as population-level compartmental models (also known as mass-action compartmental models), structured metapopulation models, and agent-based models. However, they did not give substantial attention to machine learning models, which have seen rapid development recently. Different from the above three surveys, Nsoesie et al. [249] and Chretien et al. [231] examined a broader range of models for influenza forecasting, including mechanism-based, statistical, and data-driven models. The most current surveys by Adiga et al.[227] and Clement et al. [232] concentrated on COVID-19 models, with Adiga et al. [227] categorizing them into statistical, mechanism-based, and hybrid models, while Clement et al. [232] offered a finer classification that included machine learning and deep learning approaches. Despite the comprehensive categories in [227, 232], they primarily focused on mechanism-based and statistical models, providing limited coverage of recent epidemiologically inspired machine learning models. Rodriguez et al. [253] summarized the extensive data sources available for epidemic forecasting and introduced various models, including data-driven, deep learning, and hybrid models that combine epidemiological principles with statistical methods.

B METHODOLOGY FOR LITERATURE SEARCH

In this section, we describe the process we employed to select the relevant publications for our survey.

B.1 Search Scope and Methods

Machine learning for infectious disease risk prediction is an interdisciplinary research topic, crossing the fields of public health and machine learning. To establish a foundation in this area, we conducted a comprehensive literature review using the PubMed database, with a search cutoff date of July 20, 2024. Our systematic search strategy employed a combination of keywords: ("infectious disease risk" OR endemic OR epidemic OR pandemic) AND (prediction OR forecasting OR nowcasting) AND ("machine learning" OR "artificial intelligence" OR algorithm OR modeling). This approach aimed to capture relevant studies at the intersection of machine learning and infectious disease risk prediction. Given the substantial volume of literature—approximately 18, 878 papers containing our search terms in their full texts—we refined our initial selection to include only those papers that contained these keywords within their titles and abstracts, ensuring a focused and relevant collection for review. In this step, we identified 2, 240 publications by searching in the PubMed database.

Given that (1) publications included in the PubMed dataset mainly focus on journals and topics related to public health, which may not capture the full breadth of machine learning research, and (2) our keyword-based search strategy, while comprehensive, cannot guarantee the inclusion of all publications relevant to machine learning for infectious disease risk prediction, we supplemented our search with a manual review. This involved examining personal bibliographies, tracking outputs

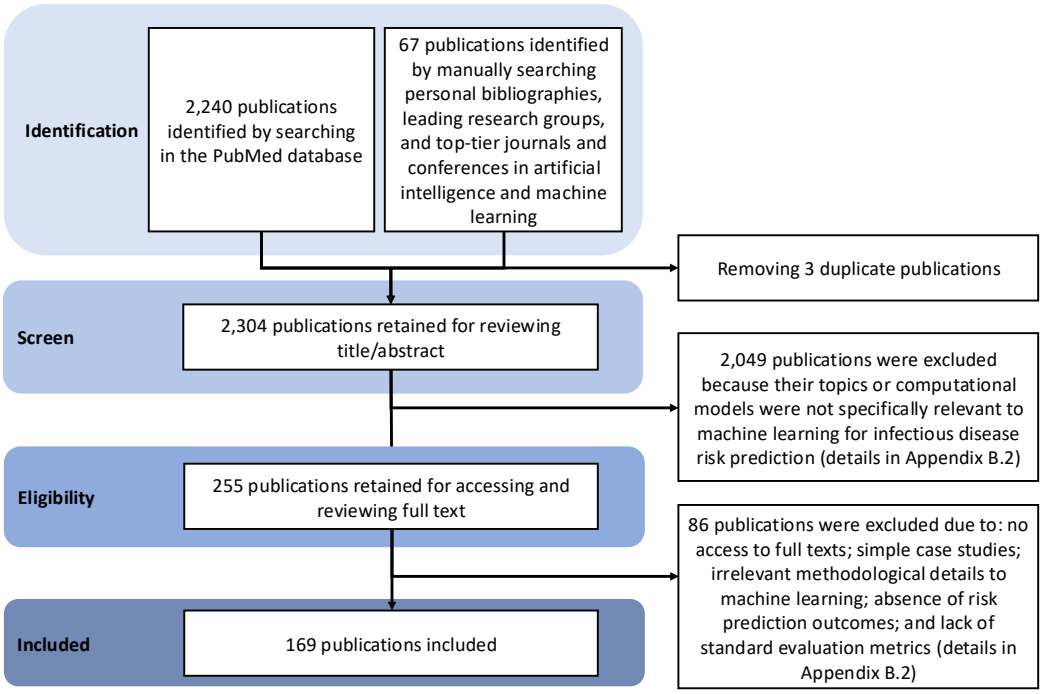

Fig. 3. Literature search flow.

from leading research groups, and searching publications from top-tier journals and conferences in artificial intelligence and machine learning. This rigorous approach was utilized with the intention to include a more comprehensive range of significant contributions to the field. After removing the duplicates between the identified publications from PubMed and those from manual searching, we retained 2, 304 publications for title and abstract review.

B.2 Selection Criteria and Results

Our first screening phase involved a thorough review of the titles and abstracts to decide whether they aligned with our inclusion and exclusion criteria. This step determined their suitability for a more in-depth full-text review. To maintain the relevance and specificity of our survey, which is focused on machine learning for infectious disease risk prediction at the human population level, we designed a set of exclusion criteria to exclude irrelevant publications. The criteria are categorized into two primary categories:

(1) **Irrelevant Topics**:
- Studies on infectious diseases not affecting human populations.
- Works focused on disease diagnosis or drug design.
- Research on health risks at the individual level resulting from infectious diseases, such as:
  - Clinical risks like ICU demand, hospitalization and severity predictions for COVID-19, and patient mortality.
  - Psychological impacts, including online behavior related to seeking psychological help and mental health disorders.

  - Opioid overdose fatalities.
  - Social consequences for individuals infected with COVID-19.

(2) **Irrelevant Computational Models**:
  - Predictions or simulations related to the effects of interventions.
  - Comparative studies between well-known existing methodologies.
  - Models not associated with machine learning algorithms, such as:
    - Approaches based on dynamic systems and control theory.
    - Purely statistical or mathematical equation-based models, and optimization methods exclusively.

According to the above criteria, we excluded 2,049 publications. This left us with a subset of 255 publications for full-text review. Subsequently, we implemented further exclusion criteria, as shown below, to ensure each study's eligibility for our survey:

- Inaccessibility to the full article text.
- Case studies merely testing pre-existing methods or comparative analyses.
- Studies where the methodological details were unrelated to machine learning models.
- Research that did not contain outcomes related to disease risk prediction.
- Studies that did not employ standard evaluation metrics to assess the performance of their proposed methods.

Upon completion of this rigorous selection process, we included a total of 169 publications, in our survey for comprehensive classification and discussion. The entire literature search workflow is illustrated in Figure 3.

C PRELIMINARY INTRODUCTION TO PURE EPIDEMIOLOGICAL AND STATISTICAL MODELS

C.1 Introduction to Epidemiological Models

In the 20th century, many epidemiological models were developed to mathematically depict the process of infectious disease transmission based on the understanding and knowledge of disease characteristics and transmission modes. In these models, a studied population is usually divided into several compartments representing different disease statuses, and a set of rules is designed to describe the transition between these statuses. These models can be subdivided based on the granularity level of modeling (from coarse to fine) into three types [227, 244]: (1) compartmental models at the population level; (2) compartmental models at the meta-population level; and (3) agent-based models at the individual level.

C.1.1 Population Level Models. Compartmental models at the population level usually include a set of differential equations (DEs) that depict the dynamics of state variables and thereby quantitatively represent disease risk. These models comprise a representative and classical group of epidemic models that are used to mathematically depict disease transmission. Various DEs have been developed for a wide range of infectious diseases, such as influenza, malaria, dengue, Aids, and COVID-19. These models assume that a disease transmission environment is homogeneous, i.e., individuals are mixed evenly within the environment and those with the same status have the same probability of moving from their current status to another status. In the following, we introduce various compartmental models at the population level that have been constructed for diseases that propagate in different ways, e.g., respiratory diseases and vector-borne diseases. The **s**usceptible–**i**nfected–**r**ecovered (SIR) model is a classic compartmental model [242]. It has a simple structure with three statuses (susceptible, infected, and recovered) and two parameters (effective contact rate

β and recovery rate γ), and is widely used to model the dynamics of infectious diseases [255], especially respiratory diseases, such as influenza, ILI, and COVID-19. Following the development of the SIR model, many other compartmental models with more sophisticated structures were designed to describe more complex scenarios, such as **s**usceptible–**e**xposed–**i**nfected–**r**ecovered (SEIR) models, which consider the latent period of a disease [238, 243] and use the parameter β to represent the probability of an individual entering the incubation period after being in contact with an infectious individual, and the parameter α to represent the probability of an individual leaving the latent period; and the **s**usceptible–**e**xposed–**i**nfected–**r**ecovered–**d**eath (SEIRD) model [247, 251], which considers deaths due to disease. Many variations of compartmental models have been developed for vector-borne diseases, such as malaria and dengue, to depict disease transmission between vectors (e.g., mosquitoes) and humans [229, 246]. For instance, the Ross model [254] is the most fundamental model to describe vector-borne diseases, while the Macdonald model [245] is based on the Ross model but also considers the latent status of vectors. In addition to considering the latent status of vectors, the Anderson and May model [228] considers the latent status of humans, and **s**usceptible–**l**atent–**i**nfected–**r**ecovered (SLIR) models [258, 259] consider the recovered status of humans in terms of acquired immunity.

C.1.2 Meta-population Level Models. Sometimes, the homogeneous-mixing assumption in population level models does not accurately reflect the actual situation of disease transmission because individuals in a host group may have different characteristics, such as different susceptibilities to disease and abilities to recover from infection. These characteristics significantly influence disease spread throughout a population and also determine how epidemic interventions should be enacted. Therefore, in addition to models based on the assumption that disease spreads occurs in a homogeneous environment (i.e., that individuals have the same probability of coming into contact with each other and of moving from one status to another), many models—i.e., compartmental models at the meta-population level—have been developed that are not based on this assumption; instead, they are based (to some extent) on a heterogeneous assumption. Studies have divided populations into subgroups and designed model structures according to different population properties, such as age structures [230, 236], geographical distributions [233, 250, 252], and human behavioral patterns [241].

C.1.3 Individual Level Models. As mentioned, epidemiological models at the meta-population level consider the heterogeneity of subgroups of a whole population. However, their characterization of the heterogeneity of population traits is still limited because of the low resolution of subgroup partitions. Contact between hosts and hosts, or between hosts and vectors, is the natural way in which infectious diseases are transmitted in the real world. Thus, **a**gent-**b**ased **m**odels (ABMs) are usually built on a network constructed at the individual level and simulate interactions between individuals, such that they model disease transmission in more realistically than epidemiological models at the meta-population level. EpiSims, proposed by Eubank et al., is an agent-based simulation tool for modeling disease spread caused by human mobility [235]. EpiSims simulates the physical contact patterns of humans by constructing a bipartite social contact network that consists of two types of vertices: individual vertices and location vertices. Compared with the results of compartmental models at the population and meta-population levels, the simulations generated by agent-based models are closer to the real-world situation because the characteristics of these models' networks are similar to those of real networks. Similarly, Hoertel et al. developed a stochastic agent-based microsimulation model for modeling the COVID-19 epidemic in France [240]. The two above-described studies show that fine-grain agent-based models enable the flexible setting of interventions and can help to reveal potentially effective intervention strategies.

C.2 Introduction to Time-series Statistical Models

Some classic statistical models for time-series prediction are based on linear model structures, such as **a**utore**g**ressive (AR) models, which linearly combine past observations within a time window p (i.e., $x_{t-1}, x_{t-2}, \cdots, x_{t-p}$) with the disturbance term E_t (the equation of AR model is shown in Eq. (S1) [42]); **m**oving **a**verage (MA) models, which complement AR models by linearly combining disturbance terms within a time window q (i.e., $E_t, E_{t-1}, \cdots, E_{t-q}$) (the equation of MA model is shown in Eq. (S2) [42]); and their combination, known as autoregressive moving average (ARMA) models (the model equation is shown in Eq. (S3) [42]), which characterize the features of single time-series dynamics to forecast future trends [42, 106].

$$y \text{ or } x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + E_t \tag{S1}$$

$$y \text{ or } x_t = E_t + \beta_1 \cdot E_{t-1} + \cdots + \beta_q \cdot E_{t-q} \tag{S2}$$

$$y \text{ or } x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + E_t + \beta_1 E_{t-1} + \cdots + \beta_q E_{t-q} \tag{S3}$$

However, the above-mentioned classic statistical models can only be applied to time series that are stationary, and so cannot be applied in many situations, as time series are often non-stationary due to the effects of seasonal factors, persistent interventions, or other determining factors. Therefore, variations of the above-mentioned models have been proposed to cope with these situations. For example, **a**utore**g**ressive **i**ntegrated **m**oving **a**verage (ARIMA) models [42] remove obvious trends (such as upward or downward trends)—which are caused by determining factors—by using d-order differencing processes. This affords stationary time series, to which ARMA models can be applied. **S**easonal **ARIMA** (SARIMA) models [42] remove the effects of seasonality by performing lagged differencing processes with a period s. Another variation of the AR model is the **a**uto**r**egressive e**x**ogenous (ARX) model. The ARX model also uses disease dynamics data to make predictions but takes other risk-related factors into account; these factors are denoted extra or exogenous variables and formulated as a weighted sum item that is in addition to the original weighted sum item.

D COMPLEMENTED LITERATURE SUMMARY

D.1 Deep Learning for Modeling Intrinsic and Complex Transmission Patterns

D.1.1 LSTM. The LSTM model is a type of **r**ecurrent **n**eural **n**etwork (RNN) which is designed for capturing the long-term dependency from time-series data [239]. Venna et al. proposed an LSTM-based deep learning model that consists of multiple LSTM cells sequentially stacked over time [177]. In the sequential structure, every LSTM cell takes two inputs (except the first cell, which only takes the dynamic data as input): the dynamic data at a single time point and the output of the previous cell, to generate a prediction for the next time step. That study also examined the effects of climate variables by applying the symbolic time-series approach and the effects of regions with geographical proximity by applying weighted summation to adjust the output of an LSTM to generate the final prediction. In addition, **m**ultivariate **LSTM** (M-LSTM) model developed by Nikparvar et al. [114] and Qu et al. [132] aimed to deal with multi-variate time series including other disease-related factors, such as human mobility and human interventions, to improve the prediction of spatiotemporal disease risks. Vadyala et al. proposed a K-means-LSTM model that uses the K-means algorithm to group time series data with similar patterns in different regions and employed the LSTM to capture the temporal dependency and make predictions [176].

D.1.2 GNNs. The **s**pectral **tem**poral **g**raph **n**eural **n**etwork (StemGNN) [25] uses a **g**raph **c**onvolutional **n**etwork (GCN) structure to model temporal dependency and inter-series correlations to predict newly confirmed COVID-19 cases. Specifically, instead of modeling the time series in the time domain, it utilizes the **g**raph **F**ourier **t**ransform (GFT) to model inter-series correlations within

the spectral domain and the **d**iscrete **F**ourier **t**ransform (DFT) to model intra-series temporal correlations within the frequency domain, and then feeds the representation of correlations into a GNN.

D.1.3 CNNs + RNNs. Xie et al. proposed inter- and intra-**s**eries **e**mbeddings **f**usion **net**work (SEFNet), which includes two parallel modules to capture the cross-regional and inner-regional dependency by the multi-scale unified convolution component with the attention mechanism and LSTM network separately [200]. Utku et al. also proposed a hybrid deep learning model that incorporates the CNN and GRU modules for capturing spatial and temporal patterns, respectively [175].

D.1.4 GNNs + Others. In the **cro**ss-**l**ocation **a**ttention-based **GNN** (ColaGNN) [41], the RNN modules extract temporal dependencies from historical disease trends, and location-aware attention is used to infer the spatial influence between different regions from these learned hidden features. Furthermore, a dilated convolution module is employed to learn attributes for each location from historical trends, thereby capturing multi-scale local temporal dependencies. Based on the learned spatial dependencies (serving as the network structure) and temporal dependencies (serving as node attributes), a graph message-passing mechanism is used to integrate the spatiotemporal information, which is then used to generate ILI predictions. The **hier**archical **s**patial-**t**emporal framework (HierST) [221] includes a temporal module that combines two time-series architectures—the **l**ong- and **s**hort-**t**erm time series **net**work (LSTNet) and the **n**eural **b**asis **e**xpansion **a**nalysis for **t**ime **s**eries (N-BEATS)—to model temporal dependency; and a spatial module that contains the gated EdgeGNN, which adaptively adjusts the connections of edges, and the NodeGNN, which learns the representation of node features. The novelty of this approach is also reflected by the introduction of prior knowledge of common sense to constrain the model inference. Specifically, given that the predictions for different administrative levels (i.e., country, state, and county) should be close to each other, they designed a consistency optimization objective that includes items representing the difference between predictions at different spatial scales in addition to the difference between ground truth and predictions. The **epi**demic forecasting model based on **f**unctional **n**eural **p**rocess (EPIFNP) proposed by Kamarthi et al. [72] also includes temporal and spatial modules, which are implemented by a probabilistic neural sequence encoder and a stochastic correlation graph, respectively. Instead of generating point estimates of forecast value, the EPIFNP model generates the probability distribution of prediction via a probabilistic generative process model to evaluate the uncertainty of prediction. The **s**patio-**t**emporal **a**ttention-based **n**eural **n**etwork (STANN) proposed by Lv et al. captured the dynamic spatiotemporal disease risk patterns using temporal attention modules and **g**raph **at**tention networks (GAT) [108]. The **s**patial-**t**emporal **s**ynchronous **g**raph **t**ransformer network (STSGT) proposed by Banerjee et al. also integrated the attention mechanism with GNN module [17]. The EpiGNN proposed by Xie et al. consisted of different deep learning modules to capture the temporal and spatial patterns, and its modeling is also mainly based on graph-structured dependencies [199]. They designed **l**ocal **t**ransmission **r**isk (LTR) encoding and **g**lobal **t**ransmission **r**isk (GTR) encoding modules to capture the spatial effects from regions surrounding and far away. Besides, they introduced the **r**egion-**a**ware **g**raph **l**earner (RAGL) module to fuse different kinds of information learned, i.e., temporal encoding, local transmission encoding, and global transmission encoding. The STGCN proposed by Sasikumar et al. stacks multiple **s**patio-**t**emporal **a**ttention **conv**olutional (STAConv) modules. The STAConv module is designed for capturing the spatial and temporal dependencies by two temporal gated convolutional layers and one spatial graph convolutional layer, respectively, and includes a multi-head attention mechanism to merge these dependencies for prediction [152]. The **g**raph **a**ttention-based **s**apatial **t**emporal (GAST) model proposed by Zhu et al. has a similar architecture with the ColaGNN model

[41] but replaced the graph passing component with a GAT module [224]. Liu et al. incorporated the GRU with the GNN to develop the GRGNN module, which extracts the features from spectral, frequency, and time domains based on the combination of GFT, DFT, and GRU modules [99].

D.1.5 Other Machine Learning Techniques. In the context of disease risk prediction, Liu et al. introduced a model that leverages matrix profile—an unsupervised learning technique—to segment and cluster time series data based on similarity [98]. This clustering informs an attention-based LSTM model within an encoder-decoder setup to predict disease dynamics. Lin et al. presented a **s**elf-supervised **m**essage-**p**assing **n**eural **n**etwork (SMPNN) that employs message passing in graph structures to embed features from spatial information and web search data in a self-supervised manner, facilitating the prediction of local disease incidence with these cross-location patterns [94]. Wu et al. developed an online multi-task regression algorithm that captures the spatial and temporal dependencies by the multi-task chain structure and the lag time series, respectively, and detects the concept drift by Hoeffding tree and **ad**aptive **win**dowing (ADWIN) drift detector [198]. DeepCOVIDNet, developed by Ramchandani et al., combines an embedding module to integrate diverse risk data from multiple regions with a DeepFM module—a neural network based on factorization machines—to address dependencies among embeddings [135]. Akhtar et al. adopted the **n**onlinear **a**uto**r**egressive models with e**x**ogenous inputs (NARX) neural network, incorporating **m**ulti**l**ayer **p**erceptron (MLP) with the ARX to predict disease risk over time and space based on heterogeneous risk-related data [7]. A recent work by Papagiannopoulou et al. introduces the **re**gional **i**nfluenza-**l**ike-**i**llness **f**orecasting (ReILIF) method, which uses the LSTM to encode the temporal dependency and **i**ntermediate **f**usion **n**etworks (IFNs) to integrate multi-modal data [123]. The fusion mechanism proposed in this study is able to integrate exogenous data to produce informative representations. Moreover, its flexibility allows for further generalizations, facilitating the development of advanced risk prediction models with heterogeneous risk-related data.

D.2 Epidemiological Parameter Inference From Data

D.2.1 Data Assimilation. In a set of data assimilation models used for epidemic prediction, the **K**alman **f**ilter (KF) [257] and its variants (such as **e**nsemble **a**djustment **K**alman **f**ilter (EAKF) [11]), as well as **p**article **f**ilter (PF) [13] methods, have been used to estimate model statuses. Yang et al. developed a forecast system, which is based on KF/PF and a simple SIR model, to predict irregular non-seasonal influenza epidemics in Hong Kong from January 1998 to December 2013 [206]. Pei et al. developed a model-data assimilation framework based on a metapopulation compartmental model to accurately predict the spatial spread of influenza [126]. In this metapopulation compartmental model, which is based on a humidity-driven SIRS model [159] that they had used in their previous studies [156, 157][260], they divided a population into different groups in terms of geographical locations (i.e., different states), and incorporated two types of human mobility (i.e., fixed commuting flows and irregular movement of visitors). Zeng et al. developed a forecasting approach by integrating a meta-population mosquito-borne SIR model with an EAKF, to predict dengue fever dynamics across twelve cities in Guangdong Province, China [209]. Sebbagh et al. proposed to use the **e**xtended **K**alman **f**ilter (EKF) to estimate the parameters in SIRD model [154]. Dukic et al. extended the SEIR model to the state-space framework to enable better characterization of temporal changes in the dynamics and estimated its parameters by a particle filtering algorithm [47].

D.2.2 GLEaM Simulation-based Methods. In [212], firstly, initial infection estimates are derived from Twitter microblogging data and traditional surveillance sources. Secondly, epidemiological parameters are calibrated through Monte Carlo simulations within a four-dimensional parameter space, using selected sampling points, while the GLEaM model is used to generate disease spread

simulations. Lastly, the most accurate models are identified using a multi-model inference strategy that minimizes information loss, as quantified by the Akaike information criterion.

D.2.3 MSE Loss-based Models. Sun et al. proposed the **d**ynamic-**s**usceptible-**e**xposed-**i**nfective-**q**uarantined (D-SEIQ) model [168], estimating key epidemiological parameters such as the basic reproduction number and incubation period using the MSE loss functions for long-term prediction of cumulative COVID-19 case numbers. Wang et al. formulated a similar loss based on MSE to estimate the parameters of epidemiological models. Based on this loss function, they formalized the learning procedure as the AutoODE algorithm, which infers the model parameters of mechanism-based models by an automatic differentiation method [189]. In addition, based on a case study on the forecasting of COVID-19 dynamics, they proposed the spatiotemporal SuEIR model, which is an extension of the SuEIR model [226] that better models spatiotemporal patterns of COVID-19 spread. Liao et al. introduced the **t**ime **w**indow-based **SIR** model (TW-SIR), which effectively captures variations in two critical epidemiological parameters—the basic reproduction number and the exponential growth rate—of COVID-19 dynamics [93]. These parameters, which were reflected by the infection and recovery/death rates, were estimated using a polynomial regression algorithm. Zhao et al. proposed an algorithm, combining the logistical model and SEIR model to estimate and adjust the epidemiological parameters to make more accurate predictions by fitting the logistic function [217]. Wang et al. proposed three **g**eneralized **b**oosting **m**achine learning (GBM) models to analyze and estimate the transmission rate from the public health policies and mobility data based on a **s**usceptible, **e**xposed, symptomatic **i**nfected, **a**symptomatic **i**nfected, Removed (SEIAR) model [191]. Camargo et al. developed an incremental learning approach, which is based on a dynamic ensemble method trained by bagging scheme, to build predictive models for SEIRD variables in the context of COVID-19 [24]. Mallick et al. focused on a meta-population SEIR model that divides the population into sixteen subgroups according to the geographical locations of sixteen states in India, with model parameters being adeptly adjusted using the **l**east **a**bsolute **s**hrinkage and **s**election **o**perator (LASSO) algorithm [109]. Krivorotko et al. proposed an algorithm for automatic calibration. The algorithm optimizes the parameters by minimizing the difference between the observed data (ground truth) and the values generated by an ABM that describes the COVID-19 transmission dynamics [86].

Zhang et al. applied an LSTM model to predict the epidemiological parameters of a modified **s**usceptible-**i**nfected-**q**uarantined-**r**ecovered (SIQR) model [211]. Khan et al. incorporated the impacts of eight COVID-19 interventions into a **s**uspected-**i**nfected-**q**uarantined-**r**emoved-**d**eceased (SIQRD) model, introducing the modified SIQRD (m-SIQRD) model [78]. They utilized an attention-based machine learning approach with a CNN-GRU architecture to calibrate the model's parameters and employed **g**enerative **a**dversarial **n**etworks (GANs) for data augmentation to enhance predictions. Feng et al. leveraged both LSTM and GRU models to estimate parameters in the SEIR model [50]. Nguyen et al. proposed the BeCaked model, which uses the combination of the LSTM model and **v**ariational **a**uto**e**ncoder (VAE) to predict the parameters of the SIRD model [113]. Ding et al. introduced the **b**ack-**p**rojection **i**nfected–**s**usceptible–**i**nfected-based **LSTM** (BPISI-LSTM) model, showcasing the application of LSTM networks in epidemiological modeling [44]. Jung et al. used FNNs to parameterize the parameters of a new SIR model with time-dependent parameters [69]. Liu et al. proposed an SIRV model, where 'V' denotes the population that has completed the whole vaccination process. Then, they adopted five methods, including NAR, LSTM, ARIMA, Gaussian function, and the polynomial function, to estimate the daily varying transmission rate of this model [100].

D.3 Epidemiological Mechanism-guided Models

D.3.1 Non-linear Models. Zhu et al. developed a transmission model for dengue that integrates prior knowledge from various aspects of disease spread, including the infectious bites from mosquitos to humans, human mobility, and the potential infectivity of mosquitoes, and they employed epidemiological indicators (i.e., VCAP and EIR) to better characteristic the dengue transmission [223]. Shi et al. constructed a nonlinear stochastic model that also utilizes VCAP and EIR for its formulation [162]. In addition, when considering the effects of cross-regional transmission, they used a periodic function to depict the periodic transmission patterns. Thus, their nonlinear stochastic model consists of the terms of local infections and imported infections. **n**ext-**g**eneration **m**atrix (NGM) [234, 256], another concept drawn from epidemiology, has also been employed in disease risk prediction, enhancing our understanding of infection dynamics. This matrix, similar to the VCAP and EIR, originates from epidemiological models and characterizes the complex, non-linear interactions between various epidemiological parameters. It is designed to track the evolution of these parameters over successive time steps, thereby facilitating the forecasting of how a disease might spread or change over time. Building on this concept, Liu et al. developed a multivariate regression model that leverages the NGM of a vector-human compartmental framework at the meta-population level. This model aims to assess and predict malaria risk by integrating heterogeneous risk-related factors, including climate variability, socio-economic conditions, land cover distribution, mosquito populations, and demographic data [97]. Based on the domain knowledge of malaria transmission, some epidemiological parameters, such as the latent period of parasites in vectors, daily biting rate, and mosquito death rate, can be calculated by the disease-related factors (e.g., temperature and rainfall). Meanwhile, other parameters that are unknown and subject to change over time and space—such as mosquito density and transmission efficiency—are dynamically inferred as the model is trained, allowing for a more accurate prediction of malaria risk.

D.3.2 Network Inference. Wang et al. introduced an approach called power-law **d**egree and **d**ata **p**riori jointly **r**egularized non-negative network **i**nference (D^2PRI), which is built upon a SIR model at the meta-population level [182]. This approach treats interactions between infected individuals across various locations as a transmission process over a disease propagation network and formulates the parameter inference of edge weights in the network and the disease transmission rate in the SIR model as an integrated network inference problem. Moreover, the model includes regularization terms that are informed by prior knowledge of network structures—such as the power-law distribution of node degrees and data derived from human mobility patterns—to guide the parameter estimation. Similarly, Prasse et al. also formulated the prediction of COVID-19 outbreaks across multiple cities as a network inference problem. Their **n**etwork-**i**nference-based **p**rediction **a**lgorithm (NIPA) is grounded in a meta-population level SIR model and aims to estimate infection rates within the contact network and the recovery rates for different cities [128].

D.4 Epidemiological Regularization and Constraints for Optimization

D.4.1 Bayesian Inference. Osthus et al. introduced a **d**ynamic **B**ayesian (DB) influenza forecasting approach that models discrepancies between simulations from epidemiological models and real-world observations [120]. This approach considers that prediction uncertainties are not solely attributable to observational errors and, therefore, models a wILI as the sum of three items: the logit of the infections that are described by the SIR model, a common discrepancy item for all influenza seasons, and a specific discrepancy item for each influenza season. Watson et al. proposed to integrate the epidemiological model, Bayesian velocity model, and random forest into a single model to predict the COVID-19 dynamics [195].

D.4.2 RNNs. Wang et al. [184, 185] proposed an epidemic prediction framework, named **d**eep learning based **e**pidemic **f**orecasting with **s**ynthetic **i**nformation (DEFSI), to conduct short-term and high-resolution ILI incidence prediction. The novelty of DEFSI is that it generates fine-scale ILI incidence data from an agent-based simulator (EpiFast) of an SEIR model, whose transmission parameters are estimated from the surveillance data. The obtained synthetic data are used to train a two-branch LSTM model to capture the within-season and between-season temporal dependencies of the incidence trends, and the model outputs are merged to generate final predictions. Bousquet et al. used the LSTM model to estimate the time-varying contact rate and deceased rate of the SIRD model [20]. Wang et al. proposed the Neural-SEIR framework, which also utilizes the LSTM model to capture the temporal patterns and predict key epidemiological parameters of the SEIR model [181]. LSTM-CA model proposed by Wang et al. used the cellular automaton (CA) to model the SEIR in a grid/spatial manner and used LSTM to capture the temporal patterns of disease transmission and predict the dynamics for each cell to calculate the infectious rate, recovered rate, and death rate [186].

D.4.3 Mixed Deep Learning Modules. Cao et al. proposed the **me**ta**po**pulation epidemic **g**raph **n**eural **n**etworks (MepoGNN), which combines the meta-population (rather than the population-level) SIR model and GNNs to predict multiple regions' disease dynamics [26]. There are two modules in the proposed model: the spatial-temporal module that uses the **t**emporal **c**onvolutional **n**etwork (TCN) and GCN layers to encode the spatial and temporal hidden patterns and estimate the time-varying epidemiological parameters, and the metapopulation SIR module to calculate the disease dynamics with the epidemiological parameters and differential equations. By designing the meta-population SIR model, they consider the disease transmission between different regions explicitly. Similarly, the epidemiology-aware deep learning framework proposed by Liu et al. also incorporates meta-population SIR models as the constraints of the objective function [96]. Different from [26], which uses the ODEs directly, Liu et al. formulated the **n**ext-**g**eneration **m**atrix (NGM) of the meta-population SIR models to predict the new infections with the epidemiological parameters and designed the epidemiological loss. In the deep learning part, they used the spatial module to learn the dependency between regions and the temporal module to infer the changing epidemiological parameters.

D.4.4 EINNs. Torku et al. proposed the **s**tochastic **e**pidemiology-**i**nformed **n**eural **n**etwork (SEINN) model to capture the hidden patterns of disease transmission by estimating the parameters and states of SIR models with neural networks [173]. Oluwasakin and Khaliq integrated the PINN with logistic models to propose the **l**ogistics-**i**nformed **n**eural **n**etwork (LINN) algorithm. This method is specifically tailored to predict the infection dynamics of the COVID-19 virus variant, Omicron. It can learn the time-varying transmission rate from the daily and cumulative number of cases [117].

D.5 Pure Machine Learning Models for Temporal Autocorrelation Analysis

D.5.1 Classical Machine Learning. Nsoesie et al. developed the **D**irichlet **p**rocess (DP) to classify and predict the influenza curves [115]. The DP model can cluster the current time series of influenza activity with similar simulated and historical time series while identifying the different patterns compared to the current time series; then it can utilize this information to make predictions. Saqib et al. introduced a hybrid model that combines polynomial regression with Bayesian ridge techniques [150]. This model utilizes an n-degree polynomial to model the relationship between historical data and future predictions, capturing the underlying trends with a flexible curve. Concurrently, it refines the estimation of parameters through Bayesian ridge regression, which treats predictions not as fixed values but as probability distributions. Another representative technology, the GP, has also been effectively utilized to model the dependencies within time series data [77, 225]. Zimmer and

Yaesoubi developed a GP-based framework explicitly tailored for forecasting seasonal epidemics [225]. Their approach differs from that of Senanayake et al. [155], who designed kernels to capture both spatial and temporal dependencies. Instead, Zimmer and Yaesoubi focused on designing kernels for characterizing the temporal dependencies between within-seasonal and between-seasonal time series. In a separate work, Ketu et al. introduced a **m**ulti-**t**ask **g**aussian **p**rocess (MTGP) regression model that operates with dual input-output configurations: the historical series (past observations) and the reference series (a combination of past observations and future points to be predicted). This model enhances the correlation between these series by employing an improved kernel matrix, thereby refining the forecasting process [77]. Wang et al. combined the Logistic growth forecasting model with the Prophet model, which is a machine learning model developed by Facebook[7], to predict the epidemic trend of COVID-19 [188]. Empirical Bayesian is also used to model the temporal characteristics. In contrast assigning a pre-defined prior distribution to the parameters of a model (i.e., a distribution that is irrelevant to the observational data), empirical Bayesian models usually estimate prior distributions from historical observations. A typical example is the semiparametric empirical Bayes framework for epidemic modeling proposed by Brooks et al. [22]. This framework first estimates the prior, i.e., the shape of the ILI curve, the noise, the peak height, the peak week, and the pacing, using a set of uniform distributions over the historical observations. It then generates the underlying ILI curve of the current ILI season by linearly adjusting the piecewise quadratic curves of historical seasons using the current year's CDC baseline weekly ILI level. In addition, the classic Kalman filter also has been applied to the forecasting of COVID-19 confirmed cases by Singh et al. [165].

D.5.2 RNNs. Gupta et al. implemented a straightforward deep learning architecture with two LSTM layers to predict the numbers of confirmed cases and deaths [62]. Kolozsvari et al. used an encoder–decoder structure with LSTM units to predict the new cases of the first wave of COVID-19 in several countries [83].

D.5.3 Other Deep Learning Models. Dash et al. introduced the **N**eural**P**rophet (NP) model [39], which features the **deep** **a**uto-**r**egressor neural **net**work (Deep-AR-Net), which is implemented by simple NN with multiple neurons, to improve upon the prediction capabilities of the original Prophet model developed by Facebook. Prophet is an additive model that accounts for trends, seasonality, and holiday effects in its forecasts. The Deep-AR-Net, integrated within the NeuralProphet framework, comprises two neural network modules: an **a**uto-**r**egressor (AR) and a **l**agged-**r**egressor (LR). These modules are specifically designed to deal with complex and nonlinear dependencies and irregular patterns that are often present in infectious disease data, thereby enhancing the model's predictive performance. Dong et al. proposed the dendritic neural regression (DNR) method, which is a type of artificial neural network model inspired by the dendritic structure of biological neurons, to predict COVID-19 dynamics [45]. The difference between it and traditional neural networks is that it uses a multiplication operator to capture the nonlinear relationships between input feature signals in the dendrite layer.

D.6 Correlated Signals of Disease Dynamics

D.6.1 Web-based Activity. Yang et al. utilized GET data to devise the **a**uto**r**egression with **Go**ogle search data (ARGO) method for influenza epidemic estimation [205]. They obtained the search terms that are strongly correlated with the ILI from Google Correlate[8] (which stopped providing

[7]https://github.com/facebook/prophet. Accessed February 13, 2025

[8]www.google.com/trends/correlate/. Cannot be accessed on March 25, 2023

data after March 28, 2015) and search trends from Google Trends[9], and then used these data as input for an ARX model. Kandula et al. also used Web-based search activity data from GET to nowcast ILI dynamics at subregional geographic scales (i.e., state-level) [73]. They formalized strongly correlated query terms for a specific region in a period as explanatory variables and used them to train an ARIMA model to predict the response variable (i.e., ILI observations). Then, they treated the forecast of the ARIMA model as an additional explanatory variable and used it with the original explanatory variables to train a random forest model for making final predictions of the response variable. Gonzalez et al. merged the outputs of **f**eed **f**orward **n**eural **n**etwork (FFNN), **s**um **o**f **s**ines (SoS), and smoothed endemic channels by a linear equation to predict **a**cute **r**espiratory **i**nfection (ARI) based on key terms in the Google search engine as well as the historical ARI data [60]. Schneider et al. applied a Lasso regression model based on Google search queries to predict the ILI rates in the Netherlands [153]. Similar search query data from Baidu[10], one of the largest search engines in China, was used by Yuan et al. to monitor influenza case counts in China [208]. They also used a multiple regression model to predict case counts based on the previous case count and a composite index of searches. Dai et al. explored the use of the deep learning architecture, i.e., an attention-based LSTM model, on the Baidu Index series data to predict the influenza risk [38].

D.6.2 Multi-type Data.

Classical machine learning models. Wang et al. utilized the signals from symptom search trends, Google mobility data, and COVID-19 vaccination coverage as the input of a dynamic supervised machine learning algorithm based on log-linear regression, aiming at predicting the COVID-19 cases in the UK [190]. Diao et al. proposed a GLM based on Poisson and negative binomial regression models to predict the malaria risk by considering the effects of the climatic variables and interventions, together with the historical malaria incidence [43]. Ciaccio et al. categorized a set of heterogeneous features into two groups—static and dynamic variables—based on their inherent characteristics [35]. Initially, they trained **p**artial **l**east **s**quares **r**egression models (PLSR) with these static variables, such as countries' general health and healthcare situation, and dynamic variables, such as climate and non-pharmaceutical interventions, to determine their individual contributions to country-level mortality. Subsequently, they applied a rolling-window Elastic Net regression model, utilizing the variables identified as most significant, to predict mortality for the upcoming 60 days.

Deep learning models. Kiang et al. utilized the meteorological and environmental data, e.g., rainfall, temperature, relative humidity, and vegetation index, as well as disease dynamic data, i.e., the malaria cases, to predict the malaria risk [79]. They used a simple neural network structure, MLP, to model the correlations between the disease-related factors and the disease dynamics. To evaluate the fine-grained dengue risk, Liu et al. collected the street-view images as environmental features of urban and processed them by a pre-trained CNN [95]. These environmental feature vectors were then fused with other features (e.g., temperature, rainfall, and past case count), together serving as the input data, to make dengue risk forecasting by MLP and SVM. Eltoukhy et al. developed a NARX model to deal with multiple types of data, including the dynamics data, such as case and death data, as well as the external factors, such as population, median age index, healthcare expenditure, and air quality [49]. Said et al. collected a range of indicators of demographic, socioeconomic, and health situations and used the K-Means clustering method to group countries with similar profiles based on these indicators [148]. For each cluster of countries sharing comparable characteristics, they constructed the multivariate time series data, including the temporal data of lockdown strategies and the

[9]https://trends.google.com/trends/. Accessed February 13, 2025
[10]https://index.baidu.com/. Accessed February 13, 2025

historical COVID-19 dynamics, and adopted the Bi-LSTM network to predict the daily cumulative COVID-19 cases. Rashed et al utilized meteorological data and the estimation of human mobility available from Google and NTT DoCoMo, Inc. in Japan, and developed a deep learning algorithm that consists of multi-path LSTM layers to predict COVID-19 dynamics [136]. Amendolara et al. used climate data (temperature and precipitation), local wind speed, demographic data (population size), and intervention data (vaccination and vaccination efficacy), as well as the historical ILI data, to predict the ILI dynamics with the Bi-LSTM network [9]. Yang et al. combined the attention mechanism and an LSTM model to propose the **m**ulti**a**ttention-**L**STM deep-learning (MAL) model to fuse the heterogeneous data of historical ILI risk, virological surveillance, climate, demography, and search engines activity to predict the ILI risk [204]. De Oliveira et al. developed a predictive model based on stacked auto-encoders for forecasting COVID-19 trends [40]. This model integrates multivariate time series data, which includes not only the recorded COVID-19 cases and deaths but also environmental factors such as temperature, humidity, and air quality index.

D.7 Uncertainty Quantification

Ghoshal and Tucker proposed **B**ayesian **c**onvolutional **n**eural **n**etworks (BCNNs) utilizing Monte-Carlo dropweights [57]. This approach has been applied to assess the confidence in model predictions for chest X-ray images, a crucial step in the diagnosis of COVID-19. Shamsi et al. introduced a deep learning framework that incorporates uncertainty-aware transfer learning, designed for the classification of chest X-ray and **C**omputed **T**omography (CT) images [160].

D.8 Model and Prediction Robustness

Cheng et al. used the stacking technique to ensemble four machine learning models—ARIMA, RF, **s**upport **v**ector **r**egression (SVR), and extreme gradient boosting—to predict the weekly ILI dynamics [32]. Yakovyna et al. introduced a supervised-unsupervised ensemble model based on stacking, which clustered the input data with many heterogeneous features and selected important features using the Boruta, decision tree, and RF models before applying a stacked ensemble of various algorithms [202].

Zhang et al. proposed the **r**andom-**f**orest-**bagging** **b**road **l**earning **s**ystem (RF-Bagging-BLS) approach, which employs RF to weight and select key features, and utilizes bootstrap strategies to randomly sample the selected features to train multiple BLS prediction models [210]. Cui et al. suggested a two-layer nested heterogeneous ensemble learning method, which first trains the individual models with different types by using the sampling data from bootstrapping and then ensembles them for predicting COVID-19 mortality [36]. Galasso et al. used an RF regression model to predict the COVID-19 cases, which aggregates the output of multiple tree-based estimators trained based on a subset of the randomly sampled input features [52].

E FORMULATIONS IN INFECTIOUS DISEASE RISK PREDICTION

E.1 Problem Statement

In the task of predicting infectious disease risks, given observations of disease dynamics-related data $\boldsymbol{x}$, a common goal of machine learning approaches is to train a model f to accurately predict the future disease dynamics y in one location or multiple locations using historical data:

$$y = f(\boldsymbol{x}). \tag{S4}$$

Usually, $\boldsymbol{x}$ denotes the model input, and it could be the historical data of the indicator of disease risks, such as the disease case number or disease prevalence. In some circumstances, it could also include other risk-related data, such as climate data, mobility data, and population data, to enhance the prediction performance. y is the model output, i.e., the future risks to be predicted, which is

usually the indicator of infectious disease risks. In general, in the temporal dimension, the input and output could cover one time step or multiple time steps; in the spatial dimension, the input and output could cover one location or multiple locations. In the following content, we use $\boldsymbol{x}$ to denote the model input, y to denote the model output, and we do not make assumptions on their dimensions. We use f to denote the general format of model's prediction functions; the specific formulations of them depend on the used model structures in different works.

E.2 Formulations of GLMs

The general formulation of GLMs is given as follows [42]:

$$p(y|\boldsymbol{x}, \boldsymbol{w}, \sigma^2) = \exp\left[\frac{y\boldsymbol{w}^T\boldsymbol{x} - A(\boldsymbol{w}^T\boldsymbol{x})}{\sigma^2} + \log h(y, \sigma^2)\right], \tag{S5}$$

where y denotes the response variable to be predicted, $\boldsymbol{x}$ denotes the input feature vector used to predict y, $\boldsymbol{w}$ denotes the weighting parameter vector on $\boldsymbol{x}$, σ^2 represents the dispersion term, $A(\cdot)$ represents the log normalizer, $h(\cdot,\cdot)$ represents the base measure, and $\exp[\cdot]$ denotes a given probability distribution from the exponential family. Furthermore, there is a mean function g^{-1} that maps $\boldsymbol{w}^T\boldsymbol{x}$ to the mean value of the response variable: $\mu = g^{-1}(\boldsymbol{w}^T\boldsymbol{x})$ [248]. Based on the distribution selected for modeling the data, the mean function varies accordingly, such as $\mu = \exp(\boldsymbol{w}^T\boldsymbol{x})$ for Poisson distribution, $\mu = \boldsymbol{w}^T\boldsymbol{x}$ for Gaussian distribution, and $\mu = \frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{x}}}$ for Bernoulli distribution.

E.3 Formulations of GP

The GP models assume that the random variable $f(\boldsymbol{x}_i)$ in continuous domains (e.g., time or space) follow a Gaussian distribution with the mean $\mu = m(\boldsymbol{x}_i)$ and the variance σ_i, and that the joint distribution of a finite set of these variables $\boldsymbol{f} = [f(\boldsymbol{x}_1), \cdots, f(\boldsymbol{x}_M)]$ will follow the multivariate Gaussian distribution with the mean $\boldsymbol{\mu} = [m(\boldsymbol{x}_1), \cdots, m(\boldsymbol{x}_M)]$ and the covariance $\Sigma_{ij} = \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j)(i, j = 1, \cdots, M)$, where the $\mathcal{K}$ is the kernel function and M is the number of observations [19][248]. In GP models, given the training set $\mathbf{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M\}$, we have $\boldsymbol{f}_X \sim \mathcal{N}(\boldsymbol{\mu}_X, \mathbf{K}_{X,X})$, where $\mathbf{K}_{X,X}$ is the $M \times M$ covariance matrix of the data set $\mathbf{X}$. Given a test set $\mathbf{X}_*$, the joint distribution $p(\boldsymbol{f}_X, \boldsymbol{f}_{X_*}|\mathbf{X}, \mathbf{X}_*)$ is represented as follows:

$$\begin{pmatrix} \boldsymbol{f}_X \\ \boldsymbol{f}_{X_*} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_{X_*} \end{pmatrix}, \begin{pmatrix} \mathbf{K}_{X,X} & \mathbf{K}_{X,X_*} \\ \mathbf{K}_{X,X_*}^\top & \mathbf{K}_{X_*,X_*} \end{pmatrix}\right). \tag{S6}$$

The covariance of these variables is calculated by choosing the appropriate kernel function $\mathcal{K}(\cdot,\cdot)$ and is used to describe the characteristics of processes.

E.4 Loss Function of Deep Learning Optimization

In our survey, for the simplicity of notations, we use $\hat{y}_d = f_d(\boldsymbol{x}, \boldsymbol{\theta}_d)$ to represent the prediction of y_d from the input feature $\boldsymbol{x}$ via a non-linear function of the deep learning model f_d. Here the f_d could be specified by any DNN model structure, and the $\boldsymbol{\theta}_d$ denotes the corresponding model parameters. Generally, $\boldsymbol{\theta}_d$ is optimized (over the parameter space $\boldsymbol{\Theta}_d$) using the following loss function:

$$\arg\min_{\boldsymbol{\theta}_d \in \boldsymbol{\Theta}_d} \mathcal{L}_d(\boldsymbol{\theta}_d) \tag{S7}$$

$$\mathcal{L}_d(\boldsymbol{\theta}_d) = \ell(\hat{y}_d, y_d) = \ell(f_d(\boldsymbol{x}, \boldsymbol{\theta}_d), y_d), \tag{S8}$$

where $\mathcal{L}_d$ is the predictive loss, quantified by the difference between the model prediction $\hat{y}_d$ and the ground truth label y_d. In the disease risk prediction task, various distance metrics could be used to measure the difference, such as the ℓ_1-norm loss (mean absolute error) or ℓ_2-norm loss (mean squared error).

E.5 Formulations of Data Assimlation

The SIRS–EAKF framework can estimate the posterior of probabilistic distributions of system states (i.e., susceptible populations S_t and infected populations I_t) and epidemiological parameters (e.g., the mean infectious period D, the average duration of immunity L, and the maximum and minimum of daily basic reproductive number R_{0max} and R_{0max}) in the used SIRS model. In [156], Shama et al. represented model states and epidemiological parameters by a set of variables $Z_t = (S_t, I_t, R_{0max}, R_{0min}, L, D)$. Then the posterior of Z_t can be represented as follows:

$$p(Z_t|y_t, y_{t-1}, \cdots) \propto p(y_t|Z_t)p(Z_t|y_{t-1}, \cdots), \tag{S9}$$

where the first term on the right-hand side is the likelihood of observational disease risk given states and parameters, while the second term is the prior distribution of the states and parameters. For KF, these two terms are assumed to be Gaussian distributions; in contrast, for PF, these two terms are not under these assumptions.

E.6 Loss Function of Epidemiological Parameterization

The general formulation of the loss function for the difference between states simulated using epidemiological models and the ground truth of these states can be represented as follows:

$$\mathcal{L}_e(\boldsymbol{\theta}_e) = \ell(\hat{\boldsymbol{y}}_e, \boldsymbol{y}_e) = \ell(f_e(\boldsymbol{\theta}_e, \boldsymbol{s}_0), \boldsymbol{y}_e), \tag{S10}$$

where $\boldsymbol{\theta}_e$ (in the parameter space $\boldsymbol{\Theta}_e$) denotes epidemiological parameters (e.g., contact rate and recovery rate) in a given epidemiological model, $\boldsymbol{y}_e$ denotes the ground truth of the target variable (usually are model states with records, e.g., infected case number and death number), $\hat{\boldsymbol{y}}_e$ denotes predictions on the target variable, $\boldsymbol{s}_0$ denotes the initial value of model states, and f_e denote the given epidemiological model, which is generally described by an ODE. With such an ODE representation, $\hat{\boldsymbol{y}}_e$ can be calculated using model parameters and initial values. Given the above loss function, the optimal model parameters can be inferred by minimizing the loss:

$$\hat{\boldsymbol{\theta}}_e = \arg\min_{\boldsymbol{\theta}_e \in \boldsymbol{\Theta}_e} \mathcal{L}_e(\boldsymbol{\theta}_e). \tag{S11}$$

E.7 Loss Function of Epidemiological Regularization and Constraints

The general formulation of the loss function for models on the type of epidemiological regularization and constraints is shown as follows:

$$\mathcal{L} = \mathcal{L}_d(\boldsymbol{\theta}_d) + \mathcal{L}_e(\boldsymbol{\theta}_e) = \ell(\hat{y}_d, y_d) + \ell(\hat{\boldsymbol{y}}_e, \boldsymbol{y}_e), \tag{S12}$$

where $\mathcal{L}_d$ denotes the prediction loss to ensure the prediction accuracy of the deep learning model, and $\mathcal{L}_e$ denotes the epidemiological-constrained loss. The approaches to introduce $\mathcal{L}_e$ vary in different works.

E.8 Equations for Model Performance Evaluation

The following indicators, the RMSE (Eq. S13), MAE (Eq. S14), MAPE (Eq. S15), and RMSPE (Eq. S16), calculate the deviation of predicted values Y^* from ground truth Y, where Y_t^* and Y_t denote the predicted value and the ground truth at time step t, respectively.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - Y_t^*\right)^2} \tag{S13}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|Y_t - Y_t^*| \tag{S14}$$

$$MAPE = \left(\max_t \frac{|Y_t - Y_t^*|}{Y_t} \right) * 100 \tag{S15}$$

$$RMSPE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\frac{Y_t - Y_t^*}{Y_t} \right)^2} \tag{S16}$$

For CORR (Eq. S17), $\bar{Y^*}$ and $\bar{Y}$ denote the mean value of the predicted trend and the real trend, respectively, during the time slot from 1 to T.

$$CORR = \frac{\sum_{t=1}^{T} \left(Y_t - \bar{Y} \right) \left(Y_t^* - \bar{Y^*} \right) \Big)}{\sqrt{\sum_{t=1}^{T} \left(Y_t - \bar{Y} \right)^2} \sqrt{\sum_{t=1}^{T} \left(Y_t^* - \bar{Y^*} \right)^2}} \tag{S17}$$

The CS is the integral of $\|k_M(c) - c\|$ over c from $0, \cdots, 1$ (Eq.S18). Thereinto, $k_M(c)$ denotes prediction interval coverage which calculating the percentage of observed values falling into the c (i.e., 50% or 95%) confidence interval of predicted distributions of M.

$$CS(M) = \int_0^1 \|k_M(c) - c\| \approx 0.01 \sum_{c \in \{0, 0.01, \cdots, 1\}} \|k_M(c) - c\| \tag{S18}$$

F ADDITIONAL TABLES

In this section, we present extended tables that supplement Table 1 and Table 2 from the main text. Specifically, Table 3 is the extended version of Table 1, which summarizes each work related to capturing the temporal autocorrelation and spatial dependency of infectious disease risk, with a particular emphasis on their machine learning components. Due to the substantial expansion of Table 2, we have divided it into two separate tables: Table 4 and Table 5. These tables summarize each work in terms of their targeted diseases, epidemiological components, and machine learning components, specifically focusing on the epidemiological parameterization models and epidemiology-embedded learning models, respectively.

SUPPLEMENTARY REFERENCES

[S227] Aniruddha Adiga, Devdatt Dubhashi, Bryan Lewis, Madhav Marathe, Srinivasan Venkatramanan, and Anil Vullikanti. 2020. Mathematical models for covid-19 pandemic: a comparative analysis. *Journal of the Indian Institute of Science* 100, 4 (2020), 793–807.

[S228] Roy M Anderson and Robert M May. 1991. *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, London. https://books.google.com.hk/books?id=HT0--xXBguQC

[S229] Mathieu Andraud, Niel Hens, Christiaan Marais, and Philippe Beutels. 2012. Dynamic epidemiological models for dengue transmission: a systematic review of structural approaches. *PloS One* 7, 11 (2012), e49085.

[S230] Carlos Castillo-Chavez, Herbert W Hethcote, Viggo Andreasen, Simon A Levin, and Wei M Liu. 1989. Epidemiological models with age structure, proportionate mixing, and cross-immunity. *Journal of Mathematical Biology* 27, 3 (1989), 233–258.

[S231] Jean-Paul Chretien, Dylan George, Jeffrey Shaman, Rohit A Chitale, and F Ellis McKenzie. 2014. Influenza forecasting in human populations: a scoping review. *PloS One* 9, 4 (2014), e94130.

[S232] J Christopher Clement, VijayaKumar Ponnusamy, KC Sriharipriya, and R Nandakumar. 2021. A survey on mathematical, machine learning and deep learning models for COVID-19 transmission and diagnosis. *IEEE Reviews in Biomedical Engineering* 15 (2021), 325–340.

[S233] Chris Cosner, John C Beier, Robert Stephen Cantrell, D Impoinvil, Lev Kapitanski, Matthew David Potts, A Troyo, and Shigui Ruan. 2009. The effects of human movement on the persistence of vector-borne diseases. *Journal of Theoretical Biology* 258, 4 (2009), 550–560.

[S234] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. 1990. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* 28 (1990), 365–382.

Table 3. Summary of spatiotemporal dependency learning models.

Categories		References	Machine learning components
Traditional machine learning	Matrix factorization and nearest neighbor	[29]	Matrix factorization based regression using nearest neighbor embedding
	Generalized linear models	[215]	Poisson regression
	Gaussian process	[194]	Dynamic Poisson autoregressive model with exogenous inputs variables
		[155]	GP with spatial and temporal kernel
Deep learning	RNN	[170]	Hierarchically stacked RNN
		[177]	Sequentially stacked LSTM
		[114, 132]	Multi-variate LSTM
		[176]	K-means for grouping time series data with similar patterns and LSTM for temporal prediction
	GNN	[75]	GNNs with spatial and temporal connections
		[51]	Distributional regression and GNN
		[25]	GCN combining with GFT and DFT
	Mixed deep modules	[197]	CNN (spatial), RNN (temporal), and residual link
		[200]	Multi-scale convolutions with attention mechanism (spatial) and LSTM (temporal)
		[175]	CNN (spatial) and GRU (temporal)
		[41]	Dilated convolution and RNN (temporal) and GNN with attention matrix (spatial)
		[221]	LSTNet/N-Beats (temporal) and GAT (spatial)
		[72]	Functional neural process
		[108]	Temporal attention layer and GAT
		[17]	GNN module with attention mechanism
		[199]	Multi-scale convolutions (temporal) and GNN (spatial)
		[152]	Temporal gated convolutional layer, spatial graph convolutional layer, and multi-head attention mechanism
		[224]	Dilated convolution, RNN, and GAT
		[99]	GFT, DFT, and GRU
	Encoder-decoder	[3]	LSTM and IDEC to encode and cluster the temporal dependency respectively
		[74]	CAE with LSTM
Other machine learning techniques	Unsupervised learning	[98]	Matrix profile and attention-based LSTM
	Self-supervised learning	[94]	Message passing neural network
	Multi-task learning	[198]	Online multi-task regression
	Factorization machine	[135]	Embedding module for various features and deep factorization machines
	Neural autoregressive model	[7]	Dynamic neural network and ARX
	Intermediate fusion network	[123]	LSTM, IFNs, and attention mechanism

Table 4. Summary of epidemiological parameterization models.

Categories	Ref.	Targeted diseases	Epidemiological components	Machine learning components
Epidemiological parameter inference from data	[156, 157]	Influenza	Humidity-driven SIRS model	EAKF/PF
	[206]	Influenza	SIR model	KF/PF
	[126]	Influenza	Metapopulation compartmental model	EAKF/PF
	[209]	Dengue	Metapopulation network	EAKF
	[154]	COVID-19	SIRD	EKF
	[47]	Influenza	SEIR	State-space framework and PF
	[172]	A/H1N1	GLEaM	Monte Carlo maximum likelihood analysis
	[212]	Influenza	GLEaM	Monte Carlo maximum likelihood analysis
	[226]	COVID-19	SuEIR	Loss function with logarithmic-type MSE
	[168]	COVID-19	D-SEIQ	Loss function with MSE
	[189]	COVID-19	Spatiotemporal-SuEIR	AutoODE
	[93]	COVID-19	TW-SIR	Loss function with MSE
	[217]	COVID-19	SEIR	Logistic model
	[191]	COVID-19	SEIAR	GBM
	[24]	COVID-19	SEIRD	Ensemble method based on bagging scheme
	[109]	COVID-19	Networked SEIR	LASSO
	[86]	COVID-19	Covasim	Optuna optimizer
	[220]	COVID-19	ISI	Pretrained NLP module and LSTM
	[89]	COVID-19	SIR and SIRD	GCN and LSTM
	[211]	COVID-19	SIQR	LSTM
	[78]	COVID-19	m-SIQRD	Attention based parameter estimation
	[50]	COVID-19	Time-varying SEIR	GRU
	[113]	COVID-19	SIRD	LSTM and VAE
	[44]	COVID-19	ISI	LSTM
	[69]	COVID-19	SIR	FNN
	[100]	COVID-19	SIRV	NAR, LSTM and statistical methods
Epidemiological parameters modeling	[12]	COVID-19	Improved SEIR model	Generalized additive model
	[14]	COVID-19	Stochastic SIR process	Mixed effects model
	[23]	COVID-19	SIR model	Mixed-effects model and GP

[S235] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 6988 (2004), 180–184.

[S236] João A N Filipe, Eleanor M Riley, Christopher J Drakeley, Colin J Sutherland, and Azra C Ghani. 2007. Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. *PLoS Computational Biology* 3, 12 (2007), e255.

[S237] Nicholas C Grassly and Christophe Fraser. 2008. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology* 6, 6 (2008), 477–487.

[S238] Herbert W Hethcote and P Van den Driessche. 1991. Some epidemiological models with nonlinear incidence. *Journal of Mathematical Biology* 29, 3 (1991), 271–287.

[S239] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[S240] Nicolas Hoertel, Martin Blachier, Carlos Blanco, Mark Olfson, Marc Massetti, Marina Sánchez Rico, Frédéric Limosin, and Henri Leleu. 2020. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine* 26,

Table 5. Summary of epidemiology-embedded learning models.

Categories	Ref.	Targeted diseases	Epidemiological components	Machine learning components
Epidemiological mechanism-guided models	[163]	Malaria	VCAP/EIR	RNN
	[223]	Dengue	VCAP/EIR	Nonlinear stochastic model
	[162]	Malaria	VCAP/EIR	Nonlinear stochastic model
	[97]	Malaria	NGM	Multiplevariate regression with non-linear parameters
	[182]	Airborne disease	Metapopulation SIR	Non-negative network inference model with power-law distribution and data prior
	[128]	COVID-19	Metapopulation SIR	Network-inference-based prediction algorithm
Epidemiological regularization and constraints for optimization	[76]	COVID-19	SIR	Spatio-temporal tensor factorization
	[66]	COVID-19	SEIR	Social media based simulation model
	[120]	Influenza	SIR	Dynamic Bayesian
	[195]	COVID-19	SIRD	Bayesian velocity model and random forest
	[184, 185]	Influenza	SEIR	LSTM
	[20]	COVID-19	SIRD	LSTM
	[181]	COVID-19	SEIR	LSTM + exponential smoothing
	[186]	COVID-19	Grid SEIR	LSTM
	[53]	COVID-19	SIR	GAT and GRU
	[183]	COVID-19	SIRD	Dynamic attention-based GCN
	[26]	COVID-19	SIR	TCN, GCN
	[96]	Influenza	SIR	CNN, RNN, and residual link
	[141]	COVID-19 and influenza	SEIRM	PINN and bi-directional GRU
	[173]	COVID	Stochastic SIR	PINN
	[117]	COVID-19	SIR	PINN and logistic model

9 (2020), 1417–1421.

[S241] John A Jacquez, Carl P Simon, James Koopman, Lisa Sattenspiel, and Timothy Perry. 1988. Modeling and analyzing HIV transmission: the effect of contact patterns. *Mathematical Biosciences* 92, 2 (1988), 119–199.

[S242] William Ogilvy Kermack and Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical* 115, 772 (1927), 700–721.

[S243] Michael Y Li and James S Muldowney. 1995. Global stability for the SEIR model in epidemiology. *Mathematical Biosciences* 125, 2 (1995), 155–164.

[S244] Jiming Liu and Shang Xia. 2020. *Computational Epidemiology.* Springer Nature Switzerland, Cham.

[S245] George Macdonald et al. 1957. *The Epidemiology and Control of Malaria.* Oxford University Press, London, UK.

[S246] Sandip Mandal, Ram Rup Sarkar, and Somdatta Sinha. 2011. Mathematical models of malaria-a review. *Malaria Journal* 10, 1 (2011), 1–19.

[S247] Andrea Maugeri, Martina Barchitta, Sebastiano Battiato, and Antonella Agodi. 2020. Estimation of unreported novel coronavirus (SARS-CoV-2) infections from reported deaths: a susceptible–exposed–infectious–recovered–dead model. *Journal of Clinical Medicine* 9, 5 (2020), 1350.

[S248] Kevin P Murphy. 2022. *Probabilistic Machine Learning: An Introduction.* MIT press, Cambridge, MA.

[S249] Elaine O Nsoesie, John S Brownstein, Naren Ramakrishnan, and Madhav V Marathe. 2014. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses* 8, 3 (2014), 309–316.

[S250] Sen Pei and Jeffrey Shaman. 2020. Initial simulation of SARS-CoV2 spread and intervention effects in the continental US. https://doi.org/10.1101/2020.03.21.20040303 medrxiv:2020.03.21.20040303

[S251] Elena Loli Piccolomini and Fabiana Zama. 2020. Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model. arXiv:2003.09909 [q-bio.PE] https://arxiv.org/abs/2003.09909

[S252] Olivia Prosper, Nick Ruktanonchai, and Maia Martcheva. 2012. Assessing the role of spatial heterogeneity and human movement in malaria dynamics and control. *Journal of Theoretical Biology* 303 (2012), 1–14.

[S253] Alexander Rodríguez, Harshavardhan Kamarthi, Pulak Agarwal, Javen Ho, Mira Patel, Suchet Sapre, and B Aditya Prakash. 2022. Data-centric epidemic forecasting: A survey. arXiv:2207.09370 [cs.LG] https://arxiv.org/abs/2207.09370

[S254] Ronald Ross. 1908. *Report on the Prevention of Malaria in Mauritius.* Waterlow & Sons Ltd, London.

[S255] Juliana Tolles and ThaiBinh Luong. 2020. Modeling epidemics with compartmental models. *The Journal of the American Medical Association* 323, 24 (2020), 2515–2516.

[S256] Pauline Van den Driessche and James Watmough. 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences* 180, 1-2 (2002), 29–48.

[S257] Greg Welch, Gary Bishop, et al. 1995. *An Introduction to the Kalman Filter.* Technical Report. Chapel Hill, NC, USA.

[S258] Hyun M Yang. 2000. Malaria transmission model for different levels of acquired immunity and temperature-dependent parameters (vector). *Revista De Saude Publica* 34, 3 (2000), 223–231.

[S259] Hyun M Yang and Marcelo U Ferreira. 2000. Assessing the effects of global warming and local social and economic conditions on the malaria transmission. *Revista De Saude Publica* 34, 3 (2000), 214–222.

[S260] Wan Yang, Marc Lipsitch, and Jeffrey Shaman. 2015. Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences* 112, 9 (2015), 2723–2728.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009