Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 7** : 1–19.
ISSN-0037-8712
doi:10.5269/bspm.81469

Trends and Challenges in Neural-Augmented Epidemic Modelling: Stability, Identifiability and Interpretability

Monika, Sarla Chouhan and Deepak Dhiman*

ABSTRACT: Neural-augmented differential equation models—including Neural Ordinary Differential Equations (Neural ODEs), Physics-Informed Neural Networks (PINNs), and Universal Differential Equations—are increasingly shaping scientific machine learning and infectious disease modelling. These approaches extend classical epidemic frameworks by embedding flexible neural components into mechanistic structures, enabling adaptability to real-world complexities like time-varying transmission rates, heterogeneous contact patterns, and incomplete data. While promising for prediction, their epidemiological adoption raises challenges in stability, identifiability, and interpretability—essential for public health trust. This review traces their evolution from early compartmental systems to COVID-19 hybrid architectures, synthesizing structural identifiability theory (parameter uniqueness), Lyapunov-based stability analysis (predictable long-term behavior), and interpretability frameworks (accuracy vs. insight). Recent applications, like hybrid SEIR–neural networks for real-time forecasting, highlight practical relevance in safety-critical contexts. We identify open questions on balancing expressiveness with tractability, uncertainty quantification, and scalable tools. These directions establish neural-augmented models as reliable, interpretable, and trustworthy tools for epidemic preparedness and response.

Key Words: Neural Ordinary Differential Equations (Neural ODEs), structural identifiability, interpretability, Lyapunov stability, Physics-Informed Neural Networks (PINNs).

Contents

* Corresponding author.
2020 *Mathematics Subject Classification:* 68T05, 34A34, 68T05, 93D05, 93B07, 62M45.
Submitted January 21, 2026. Published April 11, 2026

1. Introduction

Mathematical modelling of dynamic systems has long relied on differential equations to capture the underlying mechanistic processes governing phenomena in epidemiology, biology, physics, and engineering. While traditional compartmental and mechanistic models offer interpretability and theoretical rigor, they often struggle with flexibility and generalization in the face of noisy or sparse data. To bridge this gap, recent advances in scientific machine learning have introduced neural-augmented approaches such as Neural Ordinary Differential Equations (Neural ODEs) [1] and Physics-Informed Neural Networks (PINNs) [2]. These frameworks combine the expressive power of deep learning with the structure and continuity of differential equations, resulting in models capable of learning complex dynamics directly from data.

Despite their success in predictive tasks, these hybrid models raise important questions about their reliability, particularly in domains where decisions based on model outputs have high consequences-such as epidemic forecasting, climate modelling, and drug response simulation. The integration of neural networks introduces nonlinearity, opacity, and parameter redundancy, which challenge three classical properties essential for scientific modelling:

- Stability – Is the model's behaviour well-behaved under perturbations or over time?
- Identifiability – Can we uniquely recover the model's parameters from available data?

- Interpretability – Can we understand or trust what the model is learning?

Historically, identifiability analysis has provided essential tools for assessing whether model parameters can be estimated uniquely from data. Starting from early work by Walter and Lecourtier [3] and expanded in systems biology [3,4], structural identifiability methods have laid a foundation for analysing mechanistic models. However, these techniques must be extended to address models that incorporate neural components, such as unknown nonlinearities or time-varying functions represented via neural networks.

At the same time, stability analysis-traditionally rooted in Lyapunov theory and control systems-has become increasingly relevant for training and deploying Neural ODEs. Works by Tesi et al. [5], Ip A. et al. [6], and Zhang R. et al. [7] highlight approaches for embedding stability criteria directly into the learning process, including Lyapunov-constrained optimization, Lipschitz continuity enforcement [8], and fixed-time stability guarantees for safety-critical control systems. Within the context of epidemiological modelling, where both data scarcity and high-impact decisions prevail, the interpretability and identifiability of neural-enhanced models are especially crucial. Emerging methods such as universal differential equations [9], LASSO-ODEs for model selection [10], and neural invertible flows for outbreak inference [11] exemplify attempts to build hybrid systems that retain scientific meaning while leveraging data-driven flexibility.

This review aims to provide a unified overview of the mathematical, algorithmic, and application-driven aspects of stability and identifiability in neural-augmented dynamical systems. We organize the discussion along three interrelated axes:

1.1. Foundations of Structural Identifiability and Observability

We revisit the well established theory of structural identifiability and observability, drawing on classical results and computational tools originally developed in systems biology, and adapt them more carefully to the emerging class of hybrid neural–mechanistic models. This allows for a clearer understanding of how parameters can be uniquely estimated and how system states can be reliably inferred in complex data-driven epidemiological settings.

1.2. Stability in Neural Differential Systems

We review recent methodological advances that explicitly incorporate Lyapunov-based analysis, spectral regularization constraints, and robust training objectives as strategies for ensuring stability in neural differential systems. These approaches not only safeguard against unstable learning dynamics but also enhance the reliability of predictions in long-term epidemic simulations and related dynamical processes.

1.3. Applications in Epidemic Modelling and Beyond

We highlight recent interdisciplinary efforts that apply hybrid modeling approaches to epidemic forecasting—particularly in the context of COVID-19—alongside infectious disease control and broader scientific applications. Particular attention is given to the increasing emphasis on transparency, reproducibility, and trustworthiness, which are critical for enabling these models to inform real-world policy and decision-making across diverse domains.

By bridging perspectives from applied mathematics, machine learning, and theoretical biology, this review seeks to clarify the role of stability and identifiability in constructing reliable neural-enhanced models and outline challenges that remain at the frontier of this interdisciplinary field.

2. Literature Review

2.1. Classical Identifiability and Stability in Compartmental Models

The foundational challenge of parameter identifiability in compartmental systems dates back to the early work by Walter and Lecourtier [3], who established criteria for handling unidentifiable models and proposed strategies for reparameterization and optimal experiment design. As epidemic models are often used to infer latent disease dynamics from sparse or noisy data, identifiability directly impacts their practical reliability. Chis et al. [12] extended this line of work by comparing numerical and analytical

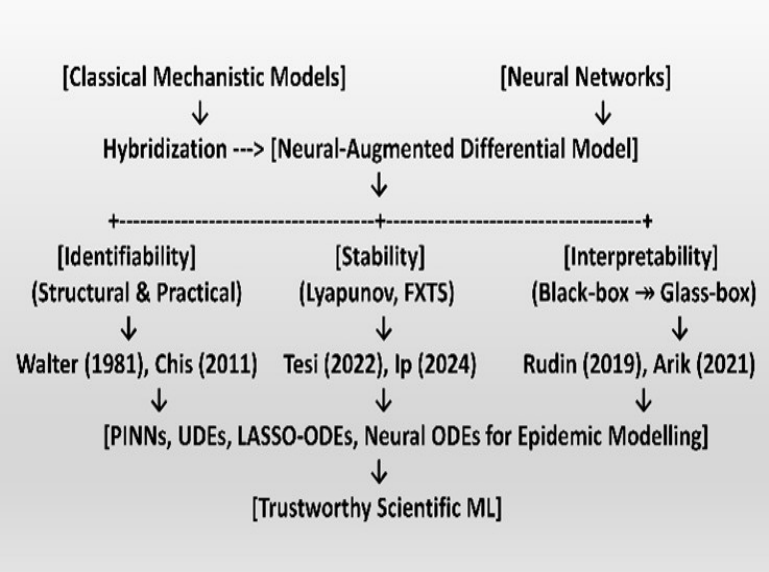

Figure 1: Conceptual diagram showing the integration of classical models and neural networks into hybrid neural-augmented differential systems. Core mathematical concerns-identifiability, stability, and interpretability-serve as pillars for building reliable and trustworthy models in scientific applications.

techniques for structural identifiability in systems biology, emphasizing the advantages of differential algebra methods and profile likelihoods.

Stability in traditional models is typically assessed via Lyapunov functions or linearization techniques. These tools provide insights into equilibrium points, reproduction numbers, and long-term epidemic trends. However, they assume that model parameters and system dynamics are explicitly defined - a limitation in scenarios involving partial knowledge or high-dimensional uncertainty.

2.2. The Rise of Neural Differential Equations

With the emergence of data-driven scientific modelling, neural ordinary differential equations (neural ODEs) introduced by Chen et al. [2] offered a new paradigm: instead of fixing the form of dynamical laws, these are learned from data. This framework retains the interpretability of ODE systems while introducing the expressive power of neural networks. However, the flexibility of neural ODEs raises new concerns about overfitting, model instability [1,5], and loss of interpretability - especially in critical domains such as epidemiology.

Physics-informed neural networks (PINNs), proposed by Raissi et al. [13], represent another hybrid approach where physical constraints are embedded into the learning process. PINNs enforce known dynamics while using neural networks to model unknown components. While their application has been promising in forward and inverse problems, especially involving PDEs, rigorous guarantees on stability and identifiability are still under development. Rackauckas et al. [9] further extended this idea via universal differential equations (UDEs), which combine mechanistic models with trainable neural components, ideal for modelling partially known biological systems. This formulation opens the door to neural-augmented epidemic models, but also amplifies the need for formal guarantees regarding identifiability and stability.

2.3. Interpretable and Stable Neural ODEs

As neural ODEs gained popularity, stability became a central concern. Haber and Ruthotto [1] proposed architectures inspired by numerical integration methods to ensure well-posed and stable learning. Their "ResNet-as-ODE" viewpoint provided a bridge between numerical ODE solvers and deep network design. Tesi et al. [5] explicitly examined the stability of neural ODEs by analysing their behaviour

under different constraints and regularizations, recommending techniques like weight normalization and spectral penalties to avoid divergence.

Kalur et al. [14] introduced stabilized neural differential equations by incorporating constraints during training that promote numerical robustness. Finlay C. et al. [8] took this further by controlling Lipschitz constants to ensure global stability, demonstrating improved robustness against adversarial inputs and noisy gradients.

Fixed-time stable neural ODEs (FxTS-Net) proposed by Zhang R. et al. [7] provided finite-time convergence guarantees, which are particularly important in epidemiology where timely interventions are crucial. These developments highlight that neural-augmented models must be trained with more than just accuracy in mind - stability must be a core design principle.

2.4. Identifiability in Neural-Augmented Models

The complexity of neural networks, particularly their nonlinearity and overparameterization, introduces severe identifiability challenges. Traditional techniques like differential algebra or observability rank conditions are not directly applicable. Liyanage et al. [15] provided a tutorial using StructuralIdentifiability.jl, demonstrating how classical identifiability analysis can still be applied to semi-mechanistic epidemic models with neural components - provided the neural structure is parameterized appropriately.

Tan and Eisenberg [10] addressed identifiability and model selection in epidemic models via LASSO-ODE, which integrates sparse regularization with mechanistic modelling. Their work provides a direction for simultaneously learning interpretable models and ensuring identifiability by penalizing unnecessary complexity.

Kiss and Simon [16] focused on identifiability in network-based epidemic models, identifying how topological features and interaction structures affect parameter inference. Their insights are particularly relevant for hybrid models where only partial interactions are mechanistically known.

Recent developments by Ip A. et al. [6] on learning Lyapunov-stable systems with neural networks further integrated control-theoretic tools into the learning pipeline, suggesting new ways to enforce stability via constraints on the neural function space, potentially improving identifiability as well.

2.5. Applications in Epidemic Forecasting and Interpretability

Neural-augmented epidemic models have shown significant promise in forecasting real-world outbreaks, particularly in contexts where traditional models struggle with nonlinearities or limited data. Friedman et al. [17] evaluated COVID-19 forecasting models across multiple nations, revealing stark differences in reliability and motivating the urgent need for explainable, trustworthy models. Arik et al. [18] developed interpretable sequence learning models for epidemic time series, emphasizing transparency in forecasting—a direction closely aligned with Rudin's [19] influential call for interpretable approaches in high-stakes decision-making, especially in public health.

Radev et al. [11] proposed Outbreak Flow, which combined invertible neural networks with Bayesian inference to track outbreak dynamics and provide improved uncertainty estimates. While powerful and flexible, such approaches remain largely black-box in nature and lack formal identifiability or stability guarantees. This critical gap underscores the necessity of frameworks that integrate both predictive accuracy and mathematical rigor—an overarching theme that this review seeks to address.

2.6. Summary and Gap Identification

While neural-augmented models offer enhanced flexibility and prediction power, they pose fundamental challenges regarding identifiability and stability - especially when used in high-stakes settings like epidemic forecasting. Despite numerous advancements in architectures (e.g., stabilized neural ODEs, UDEs), training procedures (e.g., Lipschitz control), and interpretability tools, there is no unified framework [9] for assessing whether a neural-augmented epidemic model is theoretically sound and practically reliable.

Moreover, identifiability techniques remain fragmented: classical methods do not scale well to complex neural structures, while modern sparsity-based approaches (e.g., LASSO-ODE) need stronger theoretical underpinnings. Therefore, a comprehensive synthesis of existing techniques - rooted in mathematical

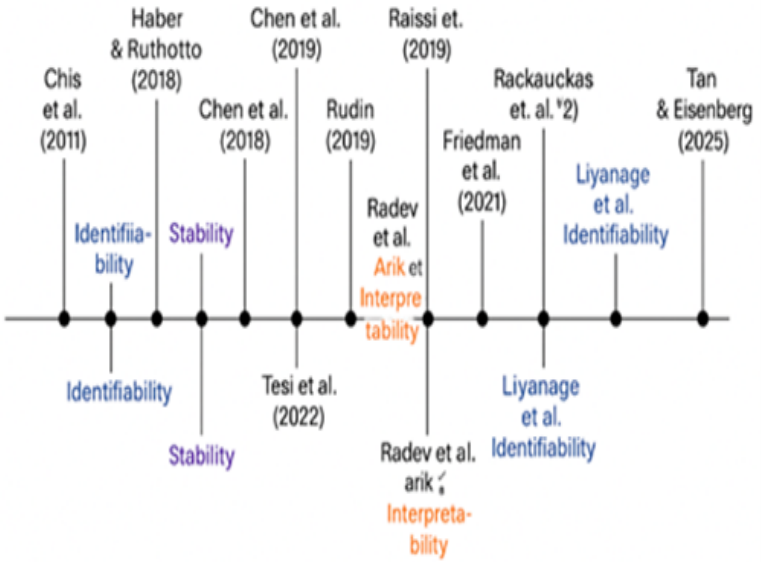

Figure 2: Timeline of Key Publications and Themes

analysis - is essential to guide the next generation of reliable, explainable, and scientifically valid hybrid epidemic models.

3. Theoretical Foundations of Stability in Neural-Augmented Models

3.1. Classical Notions of Stability in Dynamical Systems

In the study of ordinary differential equations (ODEs), stability characterizes how a system responds to perturbations near equilibrium. A fixed-point x^* of a system $\dot{x} = f(x)$ is said to be:

- Lyapunov stable if trajectories that start near x^* remain close for all future time.
- Asymptotically stable if they converge to x^*over time.
- Exponential/fixed-time stable if convergence occurs at a rate independent of initial conditions.

Lyapunov's direct method is the dominant tool for verifying these properties. A scalar function $V(x)$, known as a Lyapunov function, is constructed such that:

- $V(x) \geq 0$ for $x \neq x^*$, and $V(x^*) = 0$,
- $\dot{V}(x) = \frac{dV}{dt} = \nabla V \cdot f(x) < 0$

Such a function ensures that the system is asymptotically stable. For fixed-time stability - crucial for fast epidemic control - more stringent conditions like $\dot{V}(x) \leq -aV^{\lambda} - bV^{\delta}$ (with $a, b > 0, 0 < \lambda < 1 < \delta$) are needed, as discussed in FxTS literature (e.g. [7]).

3.2. Stability in Neural ODEs

Neural ordinary differential equations [1] define dynamics via:

$$\frac{dx}{dt} = f_e(x, t),$$

where f_e is a neural network. However, unlike classical models, the learned function f_e may not satisfy any known stability guarantees. This raises important questions:

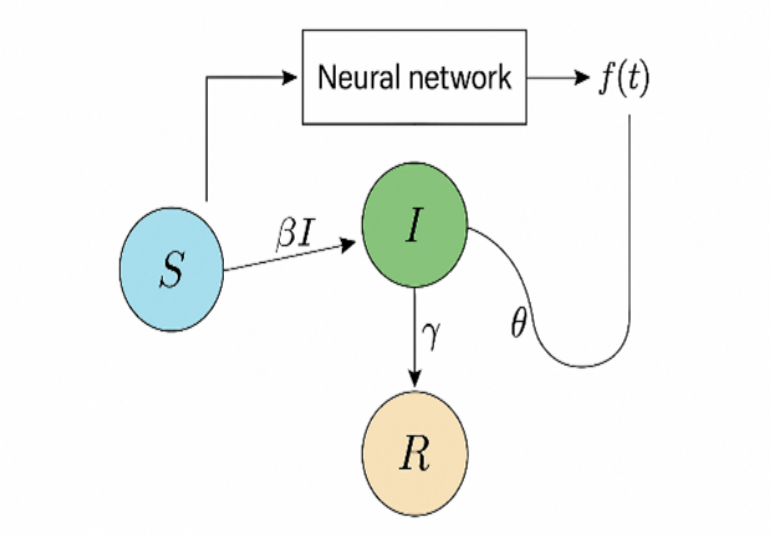

Figure 3: Neural-Augmented SIR Model

Will trajectories diverge for small perturbations in initial conditions?

Can the model generalize beyond the training data?

Will numerical solvers behave stably?

Haber and Ruthotto [1] pioneered work on designing stable deep architectures using insights from ODE discretization. By structuring networks using implicit or symplectic integrators (e.g., backward Euler or Hamiltonian flows), they ensured better long-term stability and reduced vanishing/exploding gradients during training.

Tesi et al. [5] analysed neural ODEs from a systems theory perspective, studying conditions under which solutions remain bounded and converge. Their work revealed that network depth, activation choice, and Lipschitz continuity of f_e all influence stability.

Kalur et al. [14] and Finlay C. et al. [8] proposed algorithmic solutions:

- Lipschitz regularization: Penalizing large Jacobians during training to prevent stiff dynamics.
- Implicit architectures: Using backward integration to simulate stable trajectories.
- Projection methods: Restricting the learned dynamics to lie within invariant stable sets.

These methods can be readily adapted for epidemic modelling, where the neural network represents unknown transmission or recovery functions embedded within a compartmental model.

3.3. Fixed-Time Stable Neural ODEs

Zhang R. et al. [7] proposed FxTS-Net, a framework to ensure fixed-time convergence regardless of initial conditions - particularly relevant in scenarios like outbreak mitigation. They introduced specific neural network structures that satisfy fixed-time Lyapunov criteria by design, using composite loss functions that enforce decreasing Lyapunov energy over time.

Such formulations offer critical benefits for epidemic applications:

- Rapid stabilization of predictions in uncertain settings.
- Guarantees on finite-time convergence of infection/recovery trajectories.
- Greater trust in model outputs during early outbreak phases.

3.4. Control-Theoretic Perspectives

Ip A. et al. [6] contributed a control-theoretic framework for learning stable systems using neural networks. Their approach builds a constrained optimization framework where the neural network must satisfy known Lyapunov conditions, enforced through penalty terms or constrained layers.

This approach aligns well with hybrid epidemic modelling, where the neural component governs uncertain or poorly understood processes like human mobility, seasonality, or policy effects. Enforcing stability through Lyapunov-informed training enhances interpretability and robustness, especially in out-of-distribution forecasting.

3.5. Summary and Open Challenges

Despite significant progress in stabilizing neural ODEs, challenges remain: • How can we construct explicit Lyapunov functions for hybrid systems?

• Can we embed stability guarantees into training without sacrificing expressiveness?

• How do these guarantees "We propose a mathematical workflow for analysing a neural-augmented epidemic model:" scale to stochastic settings or partial observability?

The above work offers a foundation for integrating rigorous stability analysis into epidemic forecasting models. In particular, fixed-time and Lyapunov-based methods provide promising avenues for embedding reliability into data-driven epidemic simulations.

4. Identifiability in Neural-Augmented Compartmental Models

4.1. Classical Concepts of Identifiability

Identifiability refers to the ability to uniquely estimate model parameters from observed data. In the context of compartmental epidemic models, structural identifiability addresses whether it is theoretically possible to determine parameters (like transmission or recovery rates) given perfect, noise-free observations.

Historically, Walter and Lecourtier [3] identified the challenges posed by unidentifiable compartmental models and recommended techniques such as model reparameterization, input design, or reduction of model complexity. Later, Chis et al. [12] provided a comparative overview of analytical methods (e.g., differential algebra, generating series) and numerical approaches (e.g., profile likelihoods), emphasizing their limitations in nonlinear, overparameterized systems.

Villaverde [4] extended these ideas to nonlinear biological systems, illustrating that observability and identifiability are deeply interlinked, especially when dealing with unmeasured compartments or latent states - a common feature in epidemic models.

These foundations underscore that identifiability is not merely a technical detail - it is central to model credibility. Without identifiability, fitted parameters might match observed data well but lead to drastically different predictions in extrapolation.

4.2. Challenges Introduced by Neural Components

Embedding neural networks in compartmental models introduces functional nonparametric components, which are typically not directly identifiable unless constrained.

Consider the hybrid model:

$$\begin{aligned} \frac{ds}{dt} &= f_\theta(S, I, t) \\ \frac{dI}{dt} &= = f_\theta(S, I, t) \\ \frac{dR}{dT} &= = \gamma I \end{aligned}$$

where f_θ is a neural approximation of the transmission rate. The function f_θ is not uniquely identifiable from data unless:

It is parametrically constrained (e.g., assumed to be monotonic or Lipschitz),

Sufficient observations of the compartments (S, I, R) over time are available,

Initial conditions and measurement noise are well-characterized.

Rackauckas et al. [9] proposed Universal Differential Equations (UDEs), enabling neural functions to augment known mechanistic structures. However, they noted the importance of regularization and prior knowledge to guide learning and mitigate identifiability issues.

4.3. Recent Advances in Identifiability for Neural Models

(i) Bayesian Inference and Invertible Networks

Radev et al. [11] introduced Outbreak Flow, a Bayesian inference framework using invertible neural networks. By modelling the posterior distribution of transmission dynamics, they provided uncertainty estimates alongside point forecasts. While not directly addressing identifiability, their use of normalizing flows offers a promising direction for modelling uncertainty in hybrid models.

(ii) StructuralIdentifiability.jl

Liyanage et al. [15] presented a Julia package specifically designed to test structural identifiability of epidemic models. While primarily developed for symbolic ODE models, this tool offers a computational pathway to extend identifiability analysis to neural-augmented systems, especially when neural networks are approximated via truncated polynomial or rational expansions.

(iii) Regularization and Sparse Selection

Tan and Eisenberg [10] proposed LASSO-ODE, a method that applies L1-regularization to neural-enhanced mechanistic models, enabling simultaneous structure selection and identifiability. This strategy effectively prunes unnecessary neural weights, leading to more parsimonious and interpretable models. Their framework showed success in:

- Identifying minimal models from noisy outbreak data,
- Estimating time-varying parameters like reproduction number $R_t(t)$,
- Avoiding overfitting, especially in short time series common in emerging outbreaks.

(iv) Network-based Identifiability

Kiss and Simon [16] addressed identifiability in network-based epidemic models - particularly important as modern models incorporate spatial or contact-graph structures. Their insights apply to hybrid settings, where the contact function or mobility patterns may be learned via neural networks but must remain interpretable and identifiable.

4.4. Interplay Between Stability and Identifiability

An emerging theme in recent literature is the interdependence of stability and identifiability:

Stability ensures that parameter estimates are robust to perturbations in initial conditions and data.

Identifiability ensures that model dynamics are attributable to unique parameters or functional components.

If a model is unstable, small changes in inputs or parameters can lead to wildly different outputs, making identifiability analysis meaningless. Conversely, if multiple neural functions can explain the same data, stability guarantees may apply to a non-unique set of dynamics - leading to potential misuse of forecasts.

Ip A. et al. [6] and Zhang R. et al. [7] suggest that enforcing fixed-time convergence or Lyapunov stability during training can indirectly enhance identifiability by constraining the model's functional flexibility. This represents a promising integrated direction for future research.

4.5. Summary and Research Gaps

Despite advances in tools and theory, major gaps remain:

- No general framework exists for testing structural identifiability of neural-augmented models analytically.
- Hybrid models often lack interpretability, especially when neural components are deep or unregularized.
- Time-varying identifiability (i.e., practical identifiability over sliding time windows) remains largely unexplored.
- Few studies combine Lyapunov-based stability enforcement with identifiability-aware training objectives.

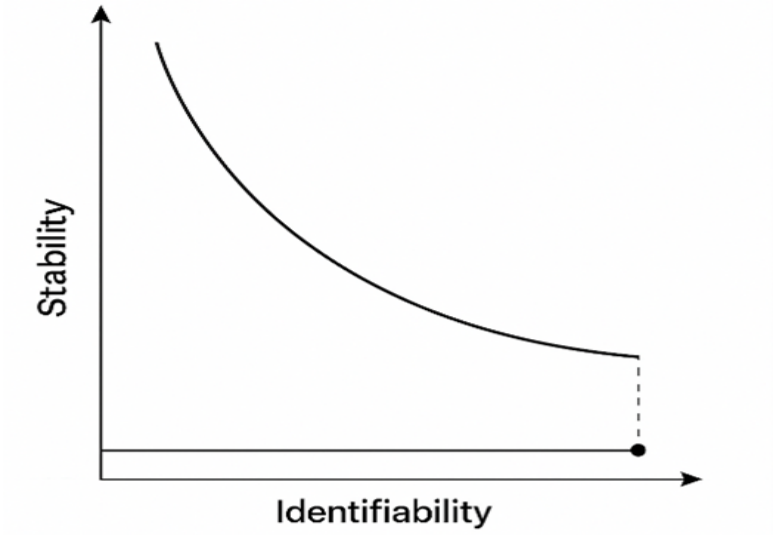

Figure 4: Stability vs. Identifiability Trade-offcx

5. Interpretability and Practical Relevance in Epidemic Forecasting

5.1. The Need for Interpretability in High-Stakes Modelling

Epidemic models often influence critical public health decisions - including lockdowns, vaccination drives, and hospital resource planning. Hence, transparency and interpretability of models become as important as their accuracy. Unlike purely statistical models, compartmental models (like SIR, SEIR) inherently provide interpretable parameters: transmission rates (β), recovery rates (γ), and reproductive numbers (R_θ). However, when neural networks are introduced, this interpretability can be compromised.

Cynthia Rudin [19] strongly argued against black-box models in high-stakes scenarios, calling instead for the adoption of inherently interpretable models. This aligns with epidemic modelling, where both stakeholders and policymakers demand explanations - not just predictions.

5.2. Neural Networks and the Black-Box Problem

Neural networks, by nature, learn internal representations that are difficult to interpret. When applied to epidemic models, they may:

Learn time-varying transmission functions without explicit form,

Obscure the relationship between interventions and changes in infection dynamics,

Introduce redundancy that hampers understanding of driving mechanisms.

As a result, model acceptance in policy domains becomes difficult. Moreover, the non-uniqueness of neural functions further reduces their trustworthiness, especially when multiple equally fitting models can provide conflicting forecasts.

5.3. Strategies for Enhancing Interpretability

Several recent papers have proposed techniques to make hybrid neural-epidemic models more interpretable. For instance, monotonic neural networks can enforce non-decreasing relationships in learned transmission-rate functions, preserving known qualitative behavior of epidemic dynamics. Similarly, neural additive models decompose predictions into structured components, making it easier to attribute changes in infection trajectories to specific covariates or interventions.

(i) Interpretable Neural Components

Arik et al. [18] proposed using attention-based mechanisms and interpretable neural sequence models to forecast COVID-19 trends. By highlighting which time steps or covariates influenced predictions, they provided insights into both model logic and underlying dynamics.

(ii) Sparse Neural Representations

Tan and Eisenberg [10] in their LASSO-ODE framework achieved model sparsity using L1 regularization. This method allows neural networks to learn minimal functional forms, essentially performing feature selection and aiding interpretation.

(iii) Hybrid Symbolic-Neural Models

In Rackauckas et al. [9] Universal Differential Equations (UDEs), part of the model is mechanistic, and part is learned. If the learned part is constrained (e.g., via polynomial bases, monotonic activation functions), interpretability improves without sacrificing expressiveness.

(iv) Symbolic Approximations of Neural Terms

Recent tools (e.g., Structural Identifiability.jl) enable researchers to approximate neural components with symbolic surrogates (e.g., rational functions) for downstream identifiability and interpretability analysis.

5.4. Balancing Expressiveness, Accuracy, and Interpretability

There exists a three-way trade-off:

Expressiveness: neural components can model complex, nonlinear, and non-stationary phenomena.

Accuracy: purely data-driven models can outperform rigid mechanistic models on short-term prediction.

Interpretability: analytical forms and identifiable parameters are necessary for insight, communication, and policy relevance. A balanced hybrid model must:

- Restrict neural components with prior knowledge,
- Use regularization to prevent overfitting,
- Combine domain-specific compartments with data-driven corrections.

5.5. Future Directions

Promising research avenues include:

- **Physics-informed interpretability:** Embed constraints like conservation of population or known epidemic thresholds directly into the neural training process.
- **Explainable AI (XAI):** Apply post-hoc explanation methods (e.g., SHAP, LIME) to epidemic forecasting models.
- **Interpretable Neural ODEs:** Design architectures that retain mathematical properties (e.g., monotonicity, boundedness) and have interpretable derivatives.
- **Visualization of Learned Functions:** Graphical tools that plot the learned transmission or contact functions over time can provide intuitive insights.

5.6. Summary

Interpretability is not an optional add-on in epidemic modelling - it is central to model utility. While neural augmentation offers flexibility and improved fit, it must be grounded in mechanisms, sparsity, and transparency. The integration of identifiability, stability, and interpretability forms the triad necessary for next-generation models that are not only accurate but also trustworthy and actionable.

6. Methodological Framework for Mathematical Analysis of Stability and Identifiability

This section outlines the mathematical techniques used to assess the stability, identifiability, and interpretability of neural-augmented compartmental epidemic models. These models combine dynamical systems theory with machine learning, requiring hybrid methods drawn from both fields.

6.1. Structure of Neural-Augmented Epidemic Models

A general neural-augmented compartmental model (e.g., SIR with neural transmission rate) may be written as:

$$\begin{aligned} \frac{dS}{dt} &= -f_\theta(t,S,I,R).S.I, \\ \frac{dI}{dt} &= f_\theta(t,S,I,R).S.I - \gamma.I, \\ \frac{dR}{dt} &= \gamma.I, \end{aligned}$$

where: f_θ is a neural network (parameterized by weights θ), γ is a constant recovery rate. This model integrates learnable nonlinear functions into classical ODE structures.

6.2. Stability Analysis Tools

Stability ensures that solutions do not exhibit uncontrolled behavior and that the system returns to equilibrium after perturbation.

Lyapunov Stability Theory For a dynamical system $\dot{x} = f(x)$, a Lyapunov function $V(x)$ is a scalar function satisfying:

and $V(x) > 0$ for all $x \neq x^*$, $\dot{V}(x) = \nabla[V(x)]^T f(x) \leq 0$ in a neighborhood of x^*. Recent works such as [6] and [14] extend this to Neural ODEs using Lyapunov neural networks that learn functions $V(x)$ satisfying these properties.

Example 6.1 *For the infected class I(t), a candidate Lyapunov function could be* $V(I) = (\frac{I^2}{2})$, *and* $V(I) = \dot{I}.I$. *Conditions on the infection dynamics* $\dot{I}$ *are then derived to ensure* $V(I) \leq 0$, *implying stability of the infection equilibrium.*

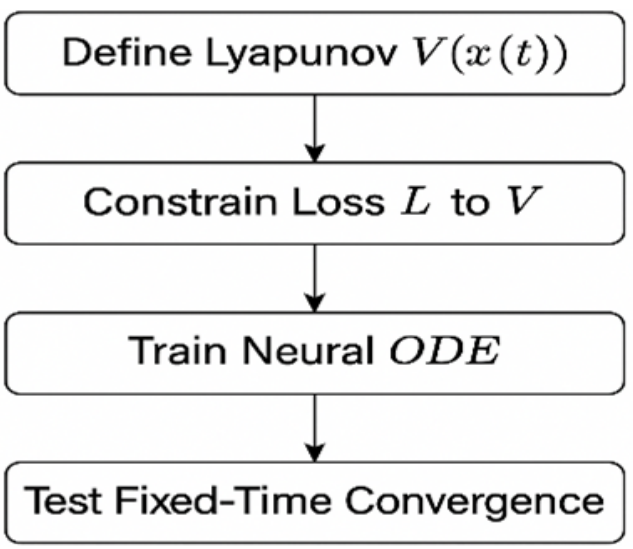

Figure 5: Lyapunov-Based Training Pipeline

Fixed-Time Stability (FxTS) Zhang R. et al. [7] proposed FxTS-Net, ensuring convergence to equilibrium within a fixed time (independent of initial conditions). This is particularly important for rapid containment strategies in epidemic modelling.

Lipschitz Constraints Bounding the Lipschitz constant of f_θ is a method to guarantee stability and prevent model blow-up. Finlay C. et al. [8] applied these techniques to robustify neural ODEs by controlling their gradient norms.

6.3. Identifiability Analysis Techniques

Identifiability ensures that the parameters and neural functions embedded in the model can be uniquely inferred from data.

Structural Identifiability via Differential Algebra Methods pioneered by Walter & Lecourtier [3] and Chis et al. [12] use symbolic manipulation of ODEs to determine whether parameters (including neural sub-functions) are structurally identifiable.

- If the neural component f_θ is parameterized using basis functions (e.g., polynomials), symbolic methods can still apply.
- Villaverde [4] extended these methods to nonlinear and neural-augmented systems.

Practical Identifiability via Sensitivity Analysis Even structurally identifiable parameters may be practically unidentifiable if they are insensitive to data noise. Tools include:

- Fisher Information Matrix (FIM),
- Profile Likelihoods,
- Monte Carlo simulations.

Software Tools

- StructuralIdentifiability.jl [15] offers automated analysis of identifiability in Julia.
- LASSO-ODE [10] enforces identifiability through sparse regression in the ODE framework, ideal for model selection.

6.4. Interpretability - Preserving Constraints

Interpretability can be mathematically enforced by:

- Penalizing model complexity (e.g., L1 norms),
- Restricting function classes (e.g., monotonic neural nets),
- Symbolically constraining model components.

For example, replacing arbitrary neural networks with spline functions or polynomial basis expansions allows identifiability tools to remain valid and improves model explainability.

6.5. Hybrid Framework Proposal

We propose a mathematical workflow to analyze a neural-augmented epidemic model:

- **Define a neural-compartmental model:** Select known compartments and embed unknown dynamics as neural networks.
- **Apply Lyapunov-based stability tests:** Use neural Lyapunov functions or fixed-time criteria.
- **Transform neural components:** Re-express f_θ in symbolic or sparse forms if needed.
- **Check identifiability:** Apply symbolic or numerical identifiability tools.
- **Incorporate interpretability constraints:** Use sparse, low-order, or physics-informed neural nets.
- **Validate using real data:** Estimate parameters, evaluate forecast quality, and confirm identifiability.

Summary The proposed methodological framework integrates:

- Control theory (for stability),
- Differential algebra and sparse learning (for identifiability),
- Structural and symbolic constraints (for interpretability).

Such an approach is not only mathematically rigorous but also essential for building epidemic models that are trustworthy, reliable, and scientifically insightful.

7. Research Challenges and Open Problems

Despite promising advances, the development of neural-augmented epidemic models that are simultaneously stable, identifiable, and interpretable remains a deeply challenging task. This section outlines key difficulties in current research and identifies open problems that demand further exploration, particularly from a mathematical perspective.

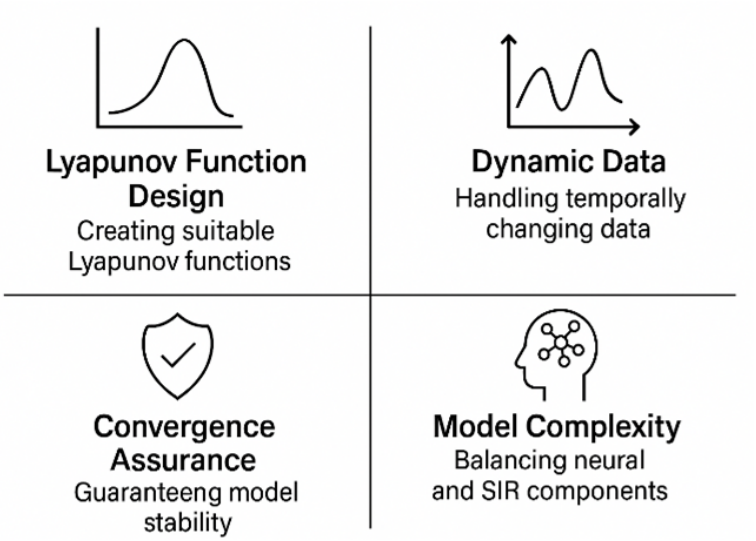

Figure 6: Challenges in Hybrid Epidemic Modelling

7.1. Theoretical Challenges in Stability Analysis

Nonlinearities and Learning Instabilities

- Neural networks introduce high degrees of nonlinearity and non-smoothness into compartmental models.
- Classical stability tools (e.g., Lyapunov functions) often fail to generalize due to the lack of analytical expressions for learned dynamics.

Open Problem: Can we develop automated Lyapunov construction methods or symbolic approximations tailored for learned systems?

Guaranteeing Stability under Training

- Many models exhibit unstable behaviour during or after training due to overfitting or adversarial dynamics.
- Fixed-time stability (FxTS) is still new in the context of epidemic modelling and lacks widespread implementation tools.

Open Problem: How can we embed stability criteria into training objectives (e.g., loss functions) for epidemic neural ODEs?

7.2. Limitations in Identifiability

Structural vs. Practical Identifiability

- A model may be structurally identifiable in theory but practically unidentifiable due to poor data quality or model sloppiness.
- Neural networks often absorb multiple mechanisms into one function, making it hard to isolate specific effects.

Open Problem: How do we redesign or regularize neural architectures to improve practical identifiability in epidemic contexts?

Lack of Symbolic Representations

- Differential algebra techniques require symbolic model equations.
- But neural components (e.g., deep nets) are often black-box functions, incompatible with symbolic computation.

Open Problem: Can we use symbolic surrogates or sparse regression approximations (like LASSO-ODE) to enable identifiability analysis of neural ODE models?

7.3. Trade-off Between Expressiveness and Interpretability

- High-capacity networks improve forecasting accuracy but reduce scientific interpretability.
- Interpretable models are often under-parameterized and lack flexibility in capturing complex dynamics.

Open Problem: What are the optimal trade-off strategies between accuracy, stability, identifiability, and interpretability? **Potential directions:**

- Hybrid models where only a small part of the model is learned.
- Constraining networks with biological priors.
- Symbolic regressors that approximate the NN output with understandable functions.

7.4. Data-Driven Challenges

Data Scarcity and Noise

- Real epidemic datasets are noisy, sparse, and irregular.
- This weakens the identifiability of parameters and increases uncertainty in neural dynamics.

Open Problem: How can we quantify uncertainty in both parameters and forecasts for hybrid epidemic models?

Time-Varying Dynamics

- Epidemic parameters change with public behaviour, policies, and environmental factors.
- Static models, even with neural components, often fail to track real-time change effectively.

Open Problem: Can we design adaptive or recurrent neural components that maintain identifiability while capturing evolving trends?

Challenge Type	Key Issues	Possible Solutions
Stability	Lyapunov design, Lipschitz constraints	Learnable Lyapunov nets, fixed-time theory
Identifiability	Black-box neural nets, unidentifiable parameters	Symbolic surrogates, sparse regression
Interpretability	Trade-off with expressiveness	Sparse and constrained NN architectures
Data quality	Sparse, noisy, shifting	Data assimilation, Bayesian NN inference
Tool limitations	Lack of neural-aware analysis tools	Extend existing toolkits, e.g., StructuralIdentifiability.jl
Policy implication	Generalization and trust	Emphasize symbolic models, interpretability metrics

7.5. Tooling and Software Gaps

- Existing tools for structural identifiability (e.g., STRIKE-GOLDD, DAISY, StructuralIdentifiability.jl) are designed for traditional ODEs, not black-box neural systems.
- No widely adopted tools currently exist for stability-constrained training of neural ODEs.

Open Problem: How can we extend or develop software frameworks for identifiability and stability analysis in hybrid neural–mechanistic models?

7.6. Generalization and Policy Impact

- Even if a model fits past data and is stable, it may fail to generalize or offer misleading guidance when deployed.
- Interpretable and identifiable models are more likely to be trusted by policymakers, but current neural methods often fall short.

Open Problem: How do we validate model trustworthiness in policy-critical scenarios such as pandemic planning, vaccination rollout, or behavioral forecasting?

7.7. Summary of Challenges

8. Future Directions

To make neural-augmented epidemic models practically usable for science and policy, we must advance the mathematical and computational tools that guarantee stability, identifiability, and interpretability. This section proposes key future research directions to address the challenges identified earlier.

8.1. Symbolic and Sparse Surrogates for Neural Components

Neural networks, though powerful, are often difficult to interpret or analyze analytically. A promising direction is to:

- Replace black-box components with sparse symbolic approximations (e.g., via LASSO-ODE, SINDy, or symbolic regression).
- Use these approximations for identifiability analysis and scientific interpretation.

Future Work: Develop frameworks where neural networks are first trained, then approximated by symbolic expressions that retain interpretability and permit analytical identifiability checks.

8.2. Integrated Stability-Constrained Training Frameworks

There is a growing need for models that are not only accurate but provably stable, especially in safety-critical epidemiological forecasting.

- Stability-aware training using Lyapunov-constrained losses, fixed-time stability theory, or Lipschitz control (e.g., FxTS-Net, Lipschitz-bounded networks).
- Embedding Lyapunov candidate generation into the training pipeline.

Future Work: Create open-source libraries for stability-constrained neural ODE training, integrating symbolic Lyapunov verifiers and adaptive step-size solvers.

8.3. Practical Identifiability in Hybrid Models

While structural identifiability has a strong mathematical foundation, practical identifiability under noise and partial observability remains underexplored.

- Extend existing tools (e.g., StructuralIdentifiability.jl) to account for neural components by combining sensitivity analysis, Bayesian inference, and Monte Carlo methods.
- Integrate data assimilation techniques to dynamically recalibrate model parameters and NN weights in real-time.

Future Work: Benchmark identifiability and uncertainty quantification methods on real-world datasets with hybrid models.

8.4. Unified Toolchains for Scientific Machine Learning

Current pipelines for neural-augmented models are fragmented: training neural networks, verifying stability, and testing identifiability are done separately.

- Develop integrated toolchains that allow symbolic epidemic models, neural components, and identifiability/stability verifiers to co-exist in a single workflow.
- **Examples:** Extending DifferentialEquations.jl, PyTorch/Julia bridges, or hybrid platforms integrating STRIKE-GOLDD with neural training libraries.

Future Work: Create interactive notebooks or packages for educators, epidemiologists, and control theorists to collaboratively design safe and interpretable models.

8.5. Explainability for Trustworthy Forecasting

Improving explainability of models is essential for adoption in policy-making.

- Use attention mechanisms, saliency maps, or interpretable neural architectures for forecasting.
- Embed causal assumptions into model design to provide stronger scientific explanations.

Future Work: Research quantitative metrics for interpretability tailored to epidemic models, and correlate them with public trust and policymaker acceptance.

8.6. Policy-Driven Model Validation

Ultimately, the utility of any model is judged by how well it informs public health decisions.

- Validate hybrid models against intervention scenarios, not just curve fitting.
- Simulate policy counterfactuals (e.g., lockdown, vaccination rollout) using hybrid models with uncertainty bounds.

Future Work: Partner with public health authorities to pilot hybrid models for scenario planning and risk communication.

8.7. Educational and Interdisciplinary Training

For the field to grow, new researchers must be trained in both rigorous mathematics and modern machine learning.

- Develop course modules, review papers, and graphical tools to explain concepts like structural identifiability, neural ODEs, and fixed-time stability.

Future Work: Promote interdisciplinary collaboration between mathematicians, epidemiologists, and AI researchers to co-develop robust hybrid modelling frameworks.

9. Conclusion and Summary of Contributions

This review advances the field by synthesizing stability, identifiability, and interpretability challenges in neural-augmented epidemic models, proposing a unified methodological framework (Section 6) that integrates Lyapunov-based stability tests, symbolic identifiability tools like StructuralIdentifiability.jl, and sparse regularization via LASSO-ODE. Key insights include the interdependence of fixed-time stability (e.g., FxTS-Net) and practical identifiability under data scarcity, as well as strategies for balancing neural expressiveness with policy-relevant transparency through constrained architectures. Unlike prior works, this paper bridges mathematical theory and epidemiological applications by outlining a six-step workflow—from neural-compartmental model definition to real-data validation—that ensures hybrid models are theoretically sound and computationally feasible. These contributions clarify how neural ODEs, PINNs, and UDEs can evolve from powerful predictors to trustworthy tools for outbreak forecasting and intervention planning. The discussion highlights remaining gaps, such as scalable symbolic surrogates for black-box neural components and stability-constrained training pipelines, paving the way for interdisciplinary toolchains that enhance model reliability in high-stakes public health scenarios.

Key Contributions of This Review:

- It provides a structured synthesis of stability, identifiability, and interpretability issues specific to neural-augmented epidemic models, connecting scattered results into a coherent picture.
- It proposes a practical six-step mathematical workflow that combines Lyapunov-based analysis, identifiability tools, and interpretability constraints for designing hybrid neural–mechanistic models.
- It highlights how recent architectures and methods—such as stabilized neural ODEs, UDEs, and sparse regression approaches—can be adapted to epidemic forecasting under data scarcity and high stakes.
- It identifies open problems and software gaps, outlining a research agenda toward scalable, mathematically grounded, and policy-ready neural-augmented epidemic modelling frameworks.

References

1. E. Haber and L. Ruthotto, *Stable architectures for deep neural networks*, Inverse Problems **34** (2018), Art. 014004.
2. R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, *Neural ordinary differential equations*, Adv. Neural Inf. Process. Syst., 2018.
3. E. Walter and Y. Lecourtier, *Unidentifiable compartmental models: What to do?*, Math. Biosci. **56** (1981), 1–25.
4. A. F. Villaverde, *Observability and structural identifiability of nonlinear biological systems*, Complexity **2019** (2019), Art. 8497093.
5. P. Tesi, G. Bonassi, and N. van de Wouw, *On the stability of neural ODE models*, Neural Networks **148** (2022), 260–271.
6. A. Ip, M. Chowdhury, and M. Arcak, *Learning Lyapunov-stable systems with neural networks*, IEEE Trans. Neural Netw. Learn. Syst. **35** (2024), 711–723.
7. R. Zhang, Y. Wang, and Z. Liu, *FxTS-Net: Fixed-time stable neural ODEs for safety-critical control*, IEEE Control Syst. Lett. **9** (2025), 388–393.
8. C. Finlay and A. Oberman, *Training robust neural networks using Lipschitz bounds*, arXiv:2005.02929, 2020.
9. C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, et al., *Universal differential equations for scientific machine learning*, arXiv:2001.04385, 2020.

10. J. Tan and M. C. Eisenberg, *LASSO-ODE: A framework for mechanistic model identifiability and selection in disease transmission modeling*, arXiv:2505.17252, 2025.

11. S. T. Radev, F. Graw, L. Chen, and A. Green, *OutbreakFlow: Model-based Bayesian inference of disease outbreak dynamics with invertible neural networks*, Adv. Neural Inf. Process. Syst., 2021.

12. O. T. Chis, J. R. Banga, and E. Balsa-Canto, *Structural identifiability of systems biology models: A critical comparison of methods*, PLoS ONE **6** (2011), Art. e27755.

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

14. A. Kalur, A. D. Jagtap, G. E. Karniadakis, and A. Chakrabarty, *Stabilized neural differential equations: Theory and algorithms*, SIAM J. Sci. Comput. **46** (2024), A90–A115.

15. Y. R. Liyanage, O. Saucedo, N. Tuncer, and G. Chowell, *A tutorial on structural identifiability of epidemic models using StructuralIdentifiability.jl*, arXiv:2505.10517, 2025.

16. I. Z. Kiss and P. L. Simon, *On parameter identifiability in network-based epidemic models*, Bull. Math. Biol. **85** (2023), Art. 18.

17. J. Friedman, P. Liu, E. Gakidou, and IHME COVID-19 Forecasting Team, *Predictive performance of international COVID-19 mortality forecasting models*, Nat. Commun. **12** (2021), Art. 2604.

18. S. Ö. Arik, C. Li, and T. Pfister, *Interpretable sequence learning for COVID-19 forecasting*, Nat. Mach. Intell. **3** (2021), 977–988.

19. C. Rudin, *Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead*, Nat. Mach. Intell. **1** (2019), 206–215.

Monika,
Department of Mathematics,
Shri Govindram Seksaria Institute of Technology and Science,
Indore, Madhya Pradesh India.
E-mail address: professormonikasingh@gmail.com

and

Sarla Chouhan,
Department of Mathematics,
Shri Govindram Seksaria Institute of Technology and Science,
Indore, Madhya Pradesh India.
E-mail address: chouhan.sarla81@gmail.com

and

Deepak Dhiman,
Centre for Distance and Online Education,
Mangalayatan University,
Aligarh, Uttar Pradesh, India.
E-mail address: deepak.dhiman09@gmail.com