DYNAMIC ENSEMBLE MODELS FOR CLIMATE-DRIVEN EPIDEMIC PREDICTION

Jinpyo Hong[1*], *Rachel E. Baker*[2,3,†]

[1] School of Engineering, Brown University, RI, USA

[2] Department of Epidemiology, School of Public Health, Brown University, RI, USA

[3] Institute at Brown for Environment and Society, Brown University, RI, USA

1. Abstract

Epidemics driven by climate variability pose critical challenges to public health systems, as factors such as rising temperatures, disrupted seasonal cycles, and extreme weather events substantially impact disease transmission dynamics. This study introduces a novel framework that combines deep learning with graph-based modeling to predict Respiratory Syncytial Virus (RSV) incidence. By integrating climate, epidemiological, and socioeconomic data across diverse U.S. states, our approach tackles the challenges of generalizing models across geographically varied regions. Leveraging a dynamic ensemble technique enhanced by transfer learning, the framework incorporates temporal, spatial, and climatic dependencies to achieve robust, region-specific time-series forecasts. The model not only captures localized trends but also adapts to heterogeneous regional patterns, offering a scalable and accurate solution for epidemic forecasting. These findings highlight the potential of data-driven methodologies to advance climate-resilient public health strategies, enabling real-time monitoring, targeted interventions, and resource optimization.

2. Introduction

1.1 Backgrounds

- ✓ Climate change is reshaping infectious disease dynamics, but traditional models like SIR struggle to adapt to rapidly shifting environmental conditions.
- ✓ Respiratory Syncytial Virus (RSV) causes over 33 million infections, 3 million hospitalizations, and 120,000 deaths annually
- ✓ RSV transmission driven not only by biology, but also by climatic variability, population density, and healthcare access.

1.2 Contributions

- ✓ We introduce a climate-aware ensemble ML framework that combines weekly RSV incidence with environmental indicators, which improves forecasting accuracy and is designed to generalize across diverse socioeconomic and environmental settings.

3. Methodology

2.1 Overview of Proposed RSV Forecasting Framework

- Hybrid deep learning model integrating temporal, spatial, and transfer learning
- Combines **LSTM, CNN, and GNN** with a **transfer learning layer**
- Designed for **scalability, robustness,** and **cross-regional generalizability**

2.2 Dual-Branch Temporal and Spatial Modeling

- **LSTM Branch**
  - ✓ Captures long-term dependencies in sequential data
  - ✓ Learns evolving trends across seasonal, climatic, and socioeconomic variables
- **CNN Branch**
  - ✓ Detects abrupt spikes or anomalies in time series
  - ✓ Complements the broader context learned by LSTM
- Outputs fused via fully connected layers to balance **trend vs. spike sensitivity**

2.3 Spatial Modeling with GNN

- Models spatial spread across regions using **graph-based structure**
  - ✓ Nodes represent regions; edges encode geographic & mobility proximity
  - ✓ Adjacency matrix incorporates physical + population connectivity
  - ✓ Iterative message passing reveals spatial correlations
- Outputs are region-specific embeddings used to enhance forecasts

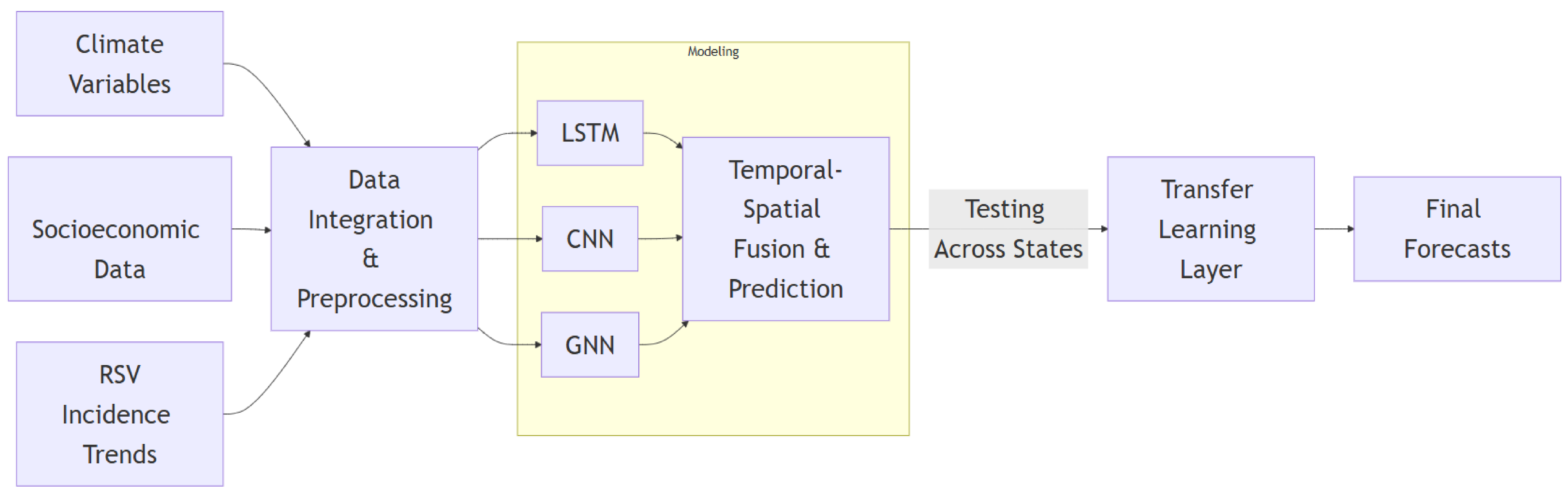

Fig 1. Overall architecture of proposed model framework

2.4 Cross-State Transfer Learning Layer

- Adapts knowledge from data-rich source states to data-scarce target states
  - ✓ Domain adaptation layer minimises distribution gaps
  - ✓ Refines shared representation of climate, demographic, and disease patterns

2.5 Public Health Impact

- Accurate and generalizable forecasts across varied environments
- Supports early intervention, efficient resource allocation, and policy planning
- Enables preparedness for climate-sensitive infectious diseases like RSV

3. Experiments

3.1 Experiment Setup

- **Dataset**
  - ✓ County level RSV surveillance data from multiple U.S. states (weekly incidence counts)
  - ✓ Climate variables: temperature, precipitation, snow depth, wind speed (NOAA)
  - ✓ Socioeconomic features: population size, birth rate (state-level)
- **Training Strategy**
  - ✓ Train a dual-branch temporal model: LSTM captures long-term seasonal trends; CNN detects short-term anomalies
  - ✓ Incorporate GNN to learn spatial dependencies based on regional proximity and mobility
  - ✓ Apply transfer learning to adapt the model from data-rich states to data-scarce ones
  - ✓ Use 4-week rolling windows as input sequences to capture recent trends and outbreak signals

3.2 Evaluation Results

- Objective performance across representative states (Table 1)
  - ✓ Model achieves high R^2 values (≥ 0.90), demonstrating strong predictive accuracy
  - ✓ Consistently low Mean Squared Error, confirming robustness across geographic regions
  - ✓ Fusion of LSTM (for long-term dynamics) and CNN (for short-term changes) provided a balanced response to RSV's temporal variability
- Cross-state transferability study
  - ✓ Model trained on South Dakota generalizes well to Colorado, capturing distinct biannual cycles (Figure 2)
  - ✓ Forecasts maintain performance despite region mismatch, indicating effective transfer learning, which highlights model's ability to adapt to diverse epidemiological patterns

Region	Massachusetts	Minnesota	North Carolina	Kansas
Mean Squared Error	141.85	156.87	158.86	140.75
R^2	0.918	0.908	0.959	0.923

Table 1. Preliminary Model Results

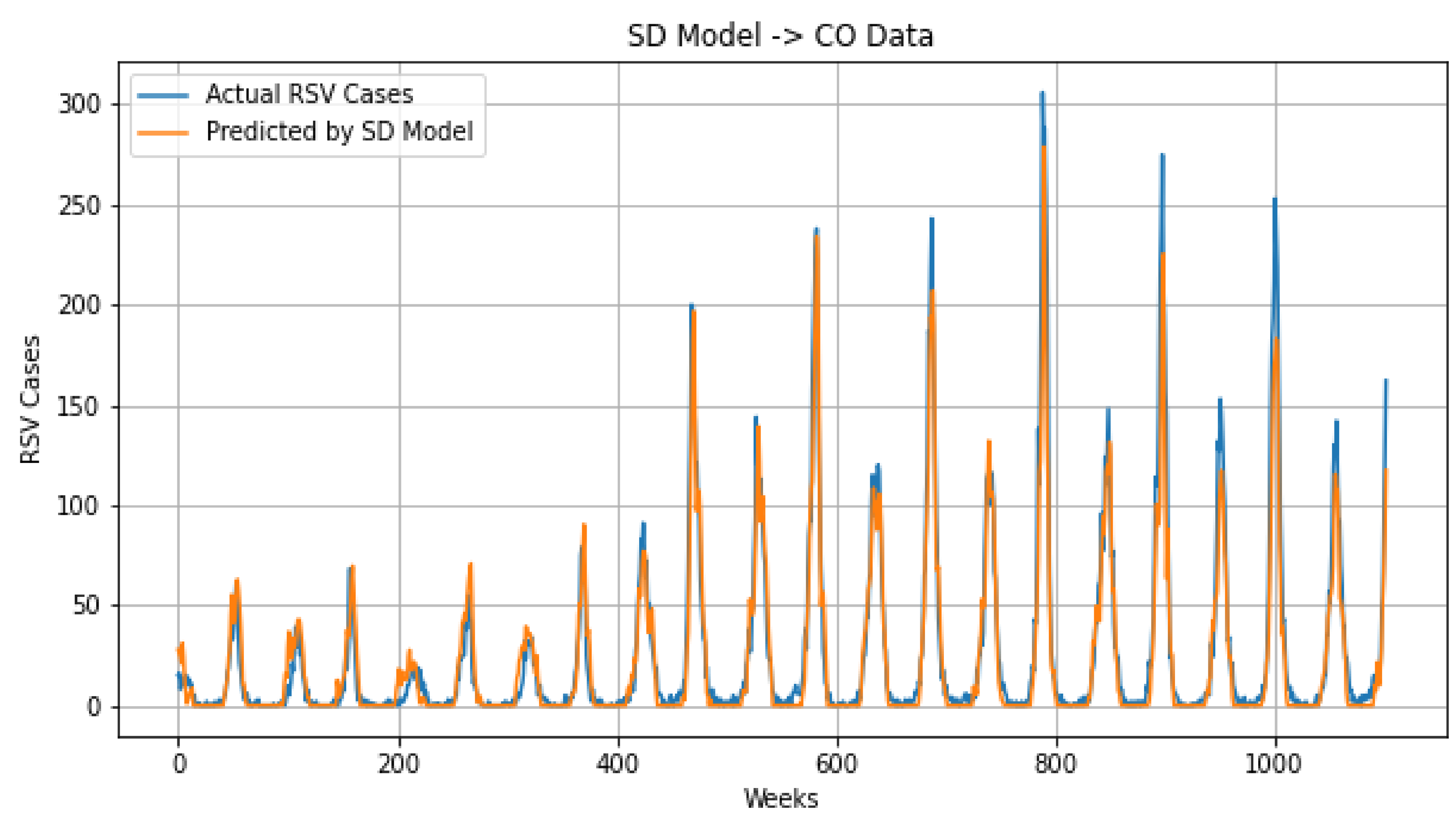

Fig 2. Distribution of prosody embeddings in predictions and ground truth

Jinpyo_hong@brown.edu