Nepal Journal of Mathematical Sciences (NJMS)
ISSN: 2738-9928 (online), 2738-9812 (print)
Vol. 5, No. 1, 2024 (February): 35-44
DOI:https://doi.org/10.3126/njmathsci.v5i1.76447
© School of Mathematical Sciences,
Tribhuvan University, Kathmandu, Nepal

Research Article
Received Date: December 3, 2024
Accepted Date: February 12, 2025
Published Date: February 28, 2025

Mathematical Approach for Estimating the Effective Reproduction Number for COVID-19 in Nepal with Data-Driven SIR Model

Shiva Hari Subedi[1*] & Gyan Bahadur Thapa[2]

[1] *Department of Applied Sciences, Thapathali Campus, Institute of Engineering, Tribhuvan University, Nepal*
[2] *Department of Applied Science and Chemical Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal*
Corresponding Author: [*]ssubedi@tcioe.edu.np

Abstract: *The Effective Reproduction Number (R_t) is a critical parameter for understanding infectious disease dynamics and informing public health interventions. This study introduces a methodological framework for estimating R_t to analyze COVID-19 transmission dynamics in Nepal. The observational approach uses the data of the COVID-19 in Nepal within a Susceptible-Infected-Removed (SIR) model, excluding vital dynamics. The time-dependent transmission and recovery rates are calculated using the data in the model equations, and then these parameters, along with epidemic data, are fed into an Long Short Term Memory (LSTM) neural network, which learns temporal patterns from the data to improve prediction accuracy. The LSTM predicted parameters are applied into the formula to estimate R_t derived by the Next Generation Matrix method. The estimated values of R_t are validated against empirically observed data. Model performance is assessed using multiple metrics including mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and coefficient of determination (r^2) to quantify the accuracy of predictions. This methodology enables quantitative assessment of epidemic trajectories and control measures, contributing to our understanding of COVID-19 dynamics in Nepal.*

Keywords: *COVID-19, SIR model, transmission dynamics, effective reproduction number, RNN-LSTM*

1 Introduction

COVID-19, transmitted by the SARS-CoV-2 virus, rapidly became a pandemic with wide-ranging impacts, affecting health, economy, education systems, and social interactions internationally [15, 25]. National and international authorities showed efforts to control its spread by applying lockdowns, travel restrictions, mask-wearing, social distancing, large-scale testing, and the rapid development of vaccines while the researchers played a vital role by informing them with appropriate ideas for control strategies. The spread of COVID-19 in Nepal has been shaped by both global trends and local factors especially causing significant challenges during its second wave (spread of delta variant). The Susceptible-Infected-Removed (SIR) model with transmission rate (β) and recovery rate (ρ) provides a structured and analytical way to understand the spread of COVID-19, inform public health strategies, predict epidemic trajectories, estimate key parameters, and assess herd immunity. A data-driven approach makes the study more accurate, adaptable, and analytical by using real-time data to identify patterns and inform evidence-based decisions.

The effective reproduction number (R_t) is a fundamental concept in epidemiology, and is defined as the average number of secondary infections caused by a single infected individual at a particular time [5]. It is a key indicator in epidemiology to estimate the current transmission risk of a disease. From an epidemiological context, $R_t > 1$, $R_t < 1$, and $R_t = 1$ indicate a growing, declining and stable epidemic respectively [13]. The value of R_t may be influenced by immunity levels of individuals, virus variants, meteorological factors, and especially, by the interventions and control measures [10].

This study aims to perform a comparative analysis of the observed R_t values with those calculated using the predicted $\beta(t)$ and $\rho(t)$ in the Next Generation Matrix (NGM) formulation of the SIR model. The time-varying parameters $\beta(t)$ and $\rho(t)$ are estimated by feeding the data into the Recurrent Neural Network-Long Short Term Memory (RNN-LSTM) model, designed to capture temporal dependencies and learn patterns in the data over time. The outcomes of this analysis can contribute to infectious disease epidemiological studies in the context of Nepal. The rest of the paper includes a literature review, methodologies, the theoretical framework of the model, results with discussion, and some applications in Sections 2 to 6, respectively. Finally, Section 7 for conclusions summarizes the major findings of the paper, sharing some future directions.

2 Literature Review

The foundational SIR model introduced by Kermack et al.[14] established the groundwork for subsequent developments. Diekmann et al. [9] provided a comprehensive framework for defining and computing the basic reproduction number (R_0) in infectious disease models for heterogeneous populations, introducing the NGM approach. Since R_0 is more useful in theoretical models and can predict the nature of the outbreaks in the early stages but is less useful for tracking dynamical changes in real-world epidemics [8], many researchers [3, 12, 17] have emphasized the importance of the time-varying reproduction number for real-time monitoring of disease dynamics. Several epidemiological studies have focused on predicting R_t using compartmental modeling with universal differential equation approaches, statistical modeling techniques, and machine learning methods. Flaxman et al. [10] used a Bayesian framework to calculate R_t for COVID-19 in various countries. Mizumoto et al. [19] investigated the transmission potential of COVID-19 aboard the Diamond Princess cruise ship using mathematical modeling with time-series data. They found that the mean R_t reached up to 11 in the confined setting, but decreased significantly after quarantine measures were implemented. Arvanitis et al. [2] used a machine learning approach to predict R_t of COVID-19 in Greece and demonstrated a high accuracy in R_t prediction, which can support short-term policy and decision making. Cori et al. [6] developed a non-parametric approach for estimating R_t based on the time series of reported cases and the serial interval, utilizing an exponential growth model. Gostic et al. [11] evaluated methods for estimating the effective reproduction number (R_t) in near real-time and recommended the approach by Cori et al. They reported that accurate R_t estimation requires proper data preprocessing, effective delay handling, and precise specification of the generation interval. Ariel et al. [4] considered a single-outbreak SIR model and its corresponding estimation procedure for R_t using sensitivity analysis and statistical theory, demonstrating its application on synthetic data. They found that while the methods performed well in ideal conditions, problems were revealed through model fit diagnostics, highlighting the need for uncertainty estimation and residual analysis in SIR-type models. Cowling et al. [7] extended estimation methods for R_t during the H1N1 pandemic in Hong Kong, integrating reporting delays with continuous inflows. Their finding indicated that R_t declined from $1.4-1.5$ at the beginning of the local epidemic to around $1.1-1.2$ later, suggesting that changes were influenced by factors like school vacations or seasonality. Song et al. [24] proposed a method for estimating R_t using deep neural networks embedded in differential equations. Using data from Ontario's first wave of COVID-19 cases, they found that their deep learning approach performed better with fewer data sources. Nishiura et al. [20] analyzed time-dependent nature of R_t using renewal theory and concluded that the instantaneous reproduction number (R_t) changes abruptly over time, whereas the cohort reproduction number varies smoothly. Li et al. [16] reviewed different researches in infectious disease epidemiology to provide basic guidance into the selection and estimation methods for R_0 and R_t. They determined that the generation interval-based approach offers essential insights into the transmission dynamics of the diseases and supports the estimation of reproduction numbers. Within the scope of COVID-19 in Nepal, Pantha et al. [22] examined province-wise data from March 29, 2020, to September 27, 2020, analyzing key indicators such as epidemic patterns, growth rates, and reproduction numbers. They found that the maximum values of R_t range from 1.20 to 2.86 with the minimum value approximately 0.16. Pokhrel et al. [23] used data-driven SEIR model to analyze dynamics of COVID-19 in Nepal and estimated the median R_t during March 24 $-$ June 01, 2020, to be 1.48 (with a minimum of 0.58 and a maximum of 3.71). Adhikari et al. [1] analyzed COVID-19 Delta variant data in Nepal to study its transmission nature and determined value of R_t up to 4.2 at the beginning of second wave (June 2021 to December 2021) and approximately 2.0 at the peak time of epidemic.

3 Methodologies

The methodologies employed in this study are both descriptive and predictive in nature, utilizing a computational modeling approach that combines mathematical modeling, data analysis, machine learning techniques, and statistical indicators. The main approaches applied to conduct the study are summarized in the Table 1.

Table 1: Methodologies Applied in the Study

SN	Components	Methods/Techniques	Remarks
1	Model	SIR model without vital dynamics	Assuming homogeneous mixing and constant population
2	Data Collection	Data retrieval from WHO, Our World in Data	Retrieving daily data of S, I, R, R_t in Nepal from 2020/5/8 to 2022/12/31
3	Data Preprocessing	Normalization, Nonlinear smoothing, Outliers removal	To reduce divergent scales, noise, outliers and overfitting
4	Prediction of β and ρ	RNN-LSTM Neural network	Input: $[S,I,R,\beta,\rho]$; Target: $[\beta,\rho]$
5	Calculation of R_t	Next generation matrix	To compute R_t
6	Stability Analysis	Jacobian eigenvalue analysis	To analyze stability at equilibrium points
7	Performance Metrics	MAE, MSE, RMSE, r^2	To evaluate the accuracy of the methods
8	Programming	MATLAB	To simulate and analyze the model

4 Theoretical Framework of SIR Model Dynamics

This section covers the theoretical concepts of the model, epidemic terminologies, the derivation of the effective reproduction number, and an analysis of stability.

4.1 Model Equations

The basic SIR model with constant population N (excluding new births and natural deaths) is described by the following transmission diagram (Fig. 1) and the system of differential equations (1)-(3). Here, $S = S(t), I = I(t)$ and $R = R(t)$ denote the number of individuals in susceptible, infected and removed compartments, respectively, at any time t.

Figure 1: Flow diagram of the SIR model.

$$\frac{dS}{dt} = -\beta\frac{SI}{N} \tag{1}$$

$$\frac{dI}{dt} = \beta\frac{SI}{N} - \rho I \tag{2}$$

$$\frac{dR}{dt} = \rho I \tag{3}$$

System of ODEs for the SIR model

At initial stage, $S(0) = S_0$, $I(0) = I_0$, $R(0) = R_0$, satisfying $S_0 + I_0 + R_0 = N_0$ (the initial population). By summing and integrating the system of ODEs, we obtain $S(t) + I(t) + R(t) = C$ (a constant). Applying the initial conditions yields $S(t) + I(t) + R(t) = N_0 = N$, since N is assumed to be constant throughout the model.

4.2 Key Terminologies in the SIR Model

i. **State Variables (S, I, R):** The state variables in the SIR model are susceptible individuals (S), who are at risk of becoming infected; infected individuals (I), who are currently infected with the disease and are infectious, and the removed (recovered+dead) individuals (R), who have either recovered or died from the disease.

ii. **Transmission Rate (β) :** In the compartmental model β represents the product of contact rate and the probability of transmission per contact, and it is often referred to as the transmission rate.

iii. **Recovery Rate (ρ):** It is the rate at which infected individuals recover from the disease and transition into the R-compartment.

iv. **Infection Rate ($\beta SI/N$):** The infection rate $\beta SI/N$ in epidemic models represents the rate at which new infections are occurring in the population, specifically the rate at which susceptible individuals become infected.

v. **Basic Reproduction Number (R_0):** It represents the average number of secondary infections generated by a single infected individual in a fully susceptible population. R_0 is calculated at the early stage of an epidemic when the population is assumed to be entirely susceptible, and no interventions have been implemented.

vi. **Effective Reproduction Number (R_t):** It represents the average number of secondary infections caused by a single infected individual at time t during the course of the epidemic. R_t is calculated at any given time, so it reflects the current dynamics of the epidemic, considering factors such as population immunity (from previous infection or vaccination), public health interventions, and behavioral changes. It provides real-time overview into whether the disease is spreading ($R_t > 1$), stable ($R_t = 1$), or declining ($R_t < 1$).

vii. **Infectious Period ($1/\rho$):** The infectious Period ($1/\rho$) refers to the average duration an individual remains infectious and capable of transmitting the disease to others. After completing this period, the infected individuals enter the R compartment, either recovering or dying.

viii. **Diseases Free Equilibrium (DFE):** It is the state of the population in which there are no infectious individuals, and the disease cannot spread. It serves as a critical point in the analysis of epidemic models, particularly when evaluating the probability for an epidemic to occur or to be controlled.

ix. **Endemic Equilibrium (EE):** It is the stable state of a disease in a population where the disease stays at a constant level over time, without causing an outbreak or completely dying out.

x. **Epidemic Threshold (ET):** It is the condition under which a disease becomes endemic in the population, meaning it can persist at a constant level over time rather than dying out. Key factor analyzing endemic threshold is R_0 (for a non-immune and non-intervened population) or R_t (for an immune and intervened population).

4.3 Formulation of R_t using NGM

From the model equations (1)-(3), the ODE for the infected sub-population can be rearranged into two terms as:

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \rho I = P - Q$$

where $P = \beta \frac{SI}{N}$ (new infection term) represents the inflow to infected compartment and $Q = \rho I$ (transition term) represents the outflow from the same compartment. P captures the transmission of infection and Q captures the progression of infection (in this case, recovery) whereas together they determine how the infection spreads through the population.

The Jacobian of P and Q with respect to I (say F and V respectively) can be derived as follows:

$$J[P(I)] = \left[\frac{\partial P}{\partial I}\right] = \left[\frac{\beta S}{N}\right] = F$$

$$J[Q(I)] = \left[\frac{\partial Q}{\partial I}\right] = \left[\rho\right] = V$$

The next generation matrix is

$$NGM = FV^{-1} = \left[\frac{\beta S}{N}\right]\left[\frac{1}{\rho}\right] = \left[\frac{\beta S}{N}\frac{1}{\rho}\right]$$

Since the NGM is 1×1 matrix, it has only one eigenvalue $\lambda = \frac{\beta S}{\rho N}$. The effective reproduction number R_t is the spectral radius (dominant eigenvalue) of NGM, so,

$$R_t = \frac{\beta S}{\rho N}$$

This gives the average number of secondary infections caused by a single infected individual in the remaining susceptible population at time t. When $t = 0$ (at the beginning of the epidemic), $S \approxeq N$ so that $R_0 \approxeq \frac{\beta}{\rho}$ which is the basic reproduction number.

4.4 Equilibrium Points and Stability Analysis

Equilibrium points of a system of ODEs for SIR model exist when the system is in steady state, *i.e.* when

$$\frac{dS}{dt} = 0, \quad \frac{dI}{dt} = 0, \quad \frac{dR}{dt} = 0$$

.
So in this state,

$$-\beta \frac{SI}{N} = 0,\ \beta \frac{SI}{N} - \rho I = 0,\ \rho I = 0$$

Disease Free Equilibrium (S^*, I^*, R^*) occurs when $I^* = 0$, so that $S^* + I^* + R^* = N$ gives $S^* + R^* = N$. Assuming the classical condition for the DFE, where the entire population remains susceptible in the absence of infection [9], we can set $R^* = 0$ to get $(S^*, I^*, R^*) = (N, 0, 0)$.

Endemic equilibrium (S^{**}, I^{**}, R^{**}) occurs when $I^{**} > 0$ and since $I^{**} \neq 0$, the second equation $\beta \frac{S^{**} I^{**}}{N} - \rho I^{**} = 0$ gives $\beta \frac{S^{**}}{N} = \rho$ or $S^{**} = \frac{\rho N}{\beta}$. Therefore, EE is $(S^{**}, I^{**}, R^{**}) = (\frac{\rho N}{\beta}, I_E, R_E)$, where I_E and R_E denote the number of infected and removed individuals respectively at the EE point.

The stability of the system of ODEs can be analyzed by determining the eigenvalues of their Jacobian matrix at the equilibrium points.
The Jacobian matrix for system of ODEs (1)-(3) is given by

$$J(S,I,R) = \begin{pmatrix} \frac{\partial}{\partial S}(\frac{dS}{dt}) & \frac{\partial}{\partial I}(\frac{dS}{dt}) & \frac{\partial}{\partial R}(\frac{dS}{dt}) \\ \frac{\partial}{\partial S}(\frac{dI}{dt}) & \frac{\partial}{\partial I}(\frac{dI}{dt}) & \frac{\partial}{\partial R}(\frac{dI}{dt}) \\ \frac{\partial}{\partial S}(\frac{dR}{dt}) & \frac{\partial}{\partial I}(\frac{dR}{dt}) & \frac{\partial}{\partial R}(\frac{dR}{dt}) \end{pmatrix} = \begin{pmatrix} -\frac{\beta I}{N} & -\frac{\beta S}{N} & 0 \\ \frac{\beta I}{N} & \frac{\beta S}{N} - \rho & 0 \\ 0 & \rho & 0 \end{pmatrix}$$

At DFE point $(S^*, I^*, R^*) = (N, 0, 0)$:

$$J(S^*, I^*, R^*) = \begin{pmatrix} 0 & -\beta & 0 \\ 0 & \beta - \rho & 0 \\ 0 & \rho & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\beta & 0 \\ 0 & \beta - \rho & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad (performing\ Row_3 = Row_1 + Row_2 + Row_3)$$

.
Therefore, eigenvalues of $J(S^*, I^*, R^*)$ are $\lambda = 0, \beta - \rho, 0$. The zero eigenvalue does not provide necessary information about stability. The eigenvalue $\beta - \rho$ gives the progression rate of the infections, which determines whether the epidemic will grow, decay, or become stable near equilibrium.

- $\beta - \rho > 0 \implies \frac{\beta}{\rho} > 1 \implies R_0 > 1 \implies$ the epidemic is unstable, and the disease will spread exponentially.
- $\beta - \rho < 0 \implies \frac{\beta}{\rho} < 1 \implies R_0 < 1 \implies$ the disease is unstable and will eventually die out.
- $\beta - \rho = 0 \implies \frac{\beta}{\rho} = 1 \implies R_0 = 1 \implies$ the disease is in a stable endemic state.

At EE point $(S^{**}, I^{**}, R^{**}) = (S_E, I_E, R_E)$, where $S_E = \rho N/\beta$:

$$J(S^{**}, I^{**}, R^{**}) = \begin{pmatrix} -\frac{\beta I_E}{N} & -\frac{\beta S_E}{N} & 0 \\ \frac{\beta I_E}{N} & \frac{\beta S_E}{N} - \rho & 0 \\ 0 & \rho & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\rho & 0 \\ \frac{\beta I_E}{N} & \frac{\beta S_E}{N} - \rho & 0 \\ 0 & \rho & 0 \end{pmatrix}, \qquad (performing\ Row_1 = Row_1 + Row_2)$$

i.e,

$$J(S^{**}, I^{**}, R^{**}) = \begin{pmatrix} 0 & 0 & 0 \\ \frac{\beta I_E}{N} & \frac{\beta S_E}{N} - \rho & 0 \\ 0 & \rho & 0 \end{pmatrix}, \qquad (performing\ Row_1 = Row_1 + Row_3)$$

Therefore, eigenvalues of $J(S^{**}, I^{**}, R^{**})$ are $\lambda = 0,\ \beta S_E/N - \rho,\ 0$. The eigenvalue $\beta S_E/N - \rho$ determines whether the disease is growing, decaying, or becoming stable at endemic equilibrium.

- $\beta S_E/N - \rho > 0 \implies \frac{\beta S_E}{\rho N} > 1 \implies R_t > 1 \implies$ the epidemic is unstable, and the disease is spreading exponentially (epidemic growth).
- $\beta S_E/N - \rho < 0 \implies \frac{\beta S_E}{\rho N} < 1 \implies R_t < 1 \implies$ the disease is unstable but towards disease-free stable state and is coming to an end.
- $\beta S_E/N - \rho = 0 \implies \frac{\beta S_E}{\rho N} = 1 \implies R_t = 1 \implies$ the system is in a marginally stable endemic state, where the disease persists at a constant level.

5 Results and Discussion

This section presents and discusses the outcomes of the RNN-LSTM method and NGM formula to predict epidemic dynamic parameters (β, ρ) and R_t, by displaying comparative plots of observed and predicted values (Figure: 2 & 3) and by tabling the performance metrics of the procedures (Table: 2 & 3).

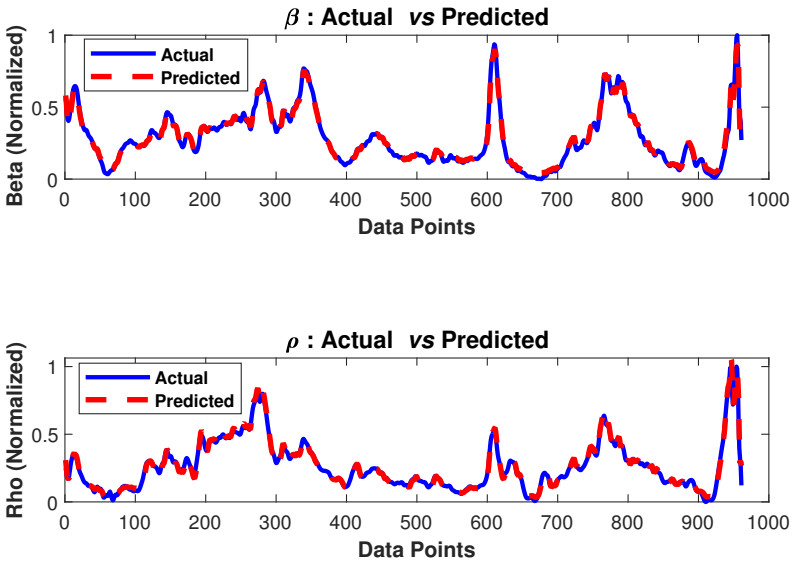

Figure 2: RNN-LSTM outcomes (for β and ρ)

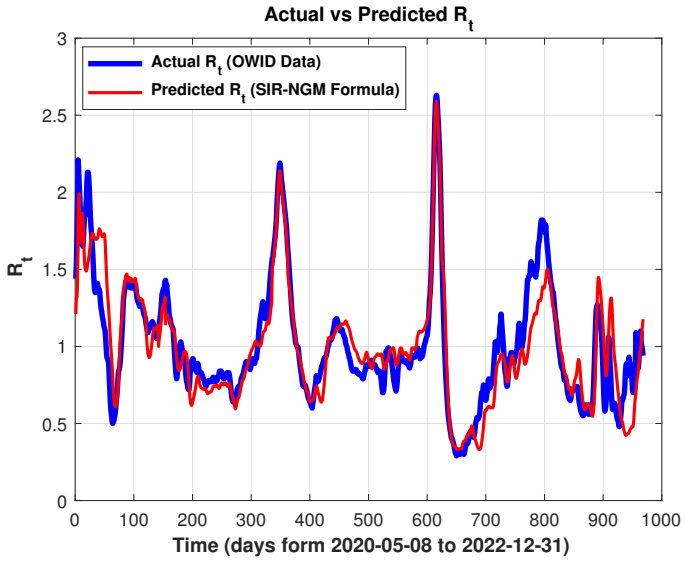

Figure 3: Actual (observed in OWID-data) and predicted (calculated by SIR-NGM-formula) R_t

Table 2: Performance metrics of RNN-LSTM predictions

Metric	**Beta**	**Rho**
MSE	0.00083214	0.00049888
RMSE	0.028847	0.022336
MAE	0.020822	0.017354
r^2	0.9774	0.97781

Table 3: Performance Metrics for Actual *vs.* Predicted R_t

Metric	**Value**
MAE	0.1249
MSE	0.0156
RMSE	0.1249
r^2	0.8984

In this study, the daily data for the compartments (S, I, R) and the time-varying reproduction number (R_t) were collected from reputable sources: the World Health Organization (WHO) [25] and Our World in Data (OWID) [21]. The preprocessed data for S, I, R and R_t were fed into a hybrid RNN-LSTM model to capture temporal dependencies in the time-series data. This model effectively predicted with greater accuracy, utilizing its ability to model complex, dynamic relationships over time. The NGM formulation was used to calculate R_t from the predicted values of β and ρ as a function of these time-varying parameters. This calculation helps for understanding the epidemic's potential to spread and evolve at each point in time.

The outcomes of the RNN-LSTM model are compared with the observed data values (Figure: 2), demonstrating that the model is an effective tool for predicting time-varying epidemic parameters. The performance metrics (Table: 2) further support the model's effectiveness. The observed and predicted values of β and ρ do not represent the actual epidemic values because the Rectified Linear Unit (ReLU) layer in the neural network normalizes the target values into the range of $[0,1]$. However, this does not affect the calculation, as the ratio of beta and rho is sufficient for estimating R_t. Additionally, the graphs of β and ρ provide an information about their comparative impacts on the disease and expose the trends of the parameters over time. The similar trends of the parameters are due to the assumption that $S+I+R=N$ (constant) which implies $dS/dt = -dR/dt - dI/dt$ and due to the comparatively small number of infected individuals as well as their changes in the context of Nepal.

According to data from WHO [25], OWID [21], and information provided by the Ministry of Health and Population (MOHP) Nepal [18], the first wave of COVID-19 in Nepal began in May 2020 and lasted until early 2021, peaking in October. The second wave (Delta variant) occurred from April 2021 to October 2021. The third wave (Omicron variant) started in December 2021 and lasted until February 2022, peaking in January. OWID's growth rate-based estimation of R_t for Nepal closely matches the R_t calculated in this study using the SIR-NGM formula with LSTM-predicted parameters (Figure: 3). The fluctuation in the predicted R_t aligns with the COVID-19 waves

in Nepal. The low error and high r^2 values (Table: 3) further validate the findings. During the first wave, our study estimates R_t to range from 0.6 to 2.0, which is consistent with the study conducted by Pokhrel et al. [23]. During the peak of the second wave, our study has found that R_t exceeds 2, aligning with the findings of Adhikari et al.[1]. This suggests that the model, which utilizes the NGM formula and RNN-LSTM predictions, is effective in simulating the epidemic dynamics and estimating the reproduction number.

There are some limitations to this study. First, the study is based on the classical SIR model, assuming a constant and homogeneous population without considering age-based variations, which may not fully capture the actual epidemic dynamics of the whole country. Second, our main focus is on comparing R_t estimates to measure the effectiveness of the RNN-LSTM method within the classical SIR model, rather than analyzing the overall trend of R_t throughout the epidemic. So, the study does not consider the factors such as government interventions, vaccination coverage, population migration, hospital capacity, meteorological conditions which could affect the accuracy and applicability of our findings. The data collection process, methodology, and timing in Nepal significantly influence the outcomes of this study. Variations in data accuracy, inconsistencies in reporting, delays in updates, and differences in testing capacity can impact the estimation of R_t. Furthermore, under-reporting, changes in testing policies, and regional disparities in data availability may introduce biases, affecting the reliability of our findings.

6 Applications

The estimated R_t has diverse applications in epidemiological research including

i. Performing real-time monitoring of disease spread.

ii. Designing targeted intervention strategies.

iii. Predicting the course of epidemics under different scenarios.

iv. Assessing the impact of public health measures like vaccination.

v. Estimating time varying transmission rate β if infectious period $1/\rho$ is known.

vi. Evaluating variants of concern.

vii. Assessing herd immunity threshold.

7 Conclusions

This study demonstrates the effectiveness of the hybrid RNN-LSTM model in predicting time-varying epidemic parameters (β and ρ) and estimating the effective reproduction number (R_t) using the SIR-NGM formulation. The predicted R_t closely aligns with observed values, capturing major fluctuations and reflecting COVID-19 waves in Nepal. The novel aspect of this study lies in the application of the neural network technique within a data-driven SIR model to achieve a refined parameter estimation approach in the context of Nepal. A higher accuracy in performance metrics provides strong scientific justification for the proposed approach, highlighting its effectiveness in capturing epidemic dynamics and improving predictive reliability. Since the model can be generalized for other countries, similar infectious diseases, different datasets, and varying conditions, it is adaptable to diverse epidemiological settings. Incorporating larger datasets and additional influencing factors further enhances the accuracy and robustness of predictions. The divergence between actual and predicted R_t in certain dates could be due to model limitations, data quality issues, and time lag in prediction. So, the limitations discussed in the previous section highlight the areas for improvement. In Nepal, data-driven studies face challenges like incomplete datasets, inconsistent collection, lack of standardization, manual entry errors, limited access, reporting delays, missing metadata, and resource constraints. Despite these constraints, the estimated R_t offers valuable insights for real-time disease monitoring, intervention strategies, and public health decision-making.

Future work will focus on refining the model by augmenting additional compartments with interventions, meteorological, and demographic factors which could provide a more comprehensive understanding of epidemic dynamics. This can be modified to perform qualitative analysis of R_t to predict the current state and future trajectory of an epidemic, helping to guide policy and public health decisions. To address the impact of poor data quality on the study, data preprocessing, cross-validation with multiple sources, uncertainty quantification, bias correction, sensitivity analysis, and transparent reporting can be employed.

References

[1] Adhikari, K., Gautam, R., Pokharel, A., Dhimal, M., Uprety, K. N., & Vaidya, N. K. **(2022)**. Insight into Delta variant dominated second wave of COVID-19 in Nepal. *Epidemics*, **41:** 100642.

[2] Arvanitis, A., Furxhi, I., Tasioulis, T., & Karatzas, K. **(2021)**. Prediction of the effective reproduction number of COVID-19 in Greece: a machine learning approach using Google mobility data. *J. Decis. Anal. Int. Comp.*, **1(1):** 1–21.

[3] Blumberg, S., & Lloyd-Smith, J. O. **(2013)**. Inference of R_0 and transmission heterogeneity from the size distribution of stuttering chains. *PLoS Comput. Biol.*, **9(5):** e1002993.

[4] Cintrón-Arias, A., Castillo-Chávez, C., Bettencourt, L. M. A., Lloyd, A. L., & Banks, H. T. **(2009)**. The estimation of the effective reproductive number from disease outbreak data. *Math. Biosci. Eng.*, **6(2):** 261–282.

[5] Contreras, S., Villavicencio, H. A., Medina-Ortiz, D., Saavedra, C. P., & Olivera-Nappa, A. **(2020)**. Real-time estimation of R_t for supporting public-health policies against COVID-19. *Front. Public Health*, **8:** 556689.

[6] Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. **(2013)**. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.*, **178(9):** 1505–1512.

[7] Cowling, B. J., Lau, M. S. Y., Ho, L.-M., Chuang, S.-K., Tsang, T., Liu, S.-H., Leung, P.-Y., Lo, S.-V., & Lau, E. H. Y. **(2010)**. The effective reproduction number of pandemic influenza: Prospective estimation. *Epidemiology*, **21(6):** 842–846.

[8] Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., & Jacobsen, K. H. **(2019)**. Complexity of the basic reproduction number (R_0). *Emerg. Infect. Dis.*, **25(1):** 1–4.

[9] Diekmann, O., Heesterbeek, J., and Metz, J. **(1990)**. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, **28(4):** 365–382.

[10] Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J.W., Monod, M., Imperial College COVID-19 Response Team, Ghani, A.C., Donnelly, C.A., Riley, S., Vollmer, M.A.C., Ferguson, L.C., & Bhatt, S. **(2020)**. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, **584(7820):** 257–261.

[11] Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R., Niehus, R., Hay, J. A., De Salazar, P. M., Hellewell, J., Meakin, S., Munday, J. D., Bosse, N. I., Sherrat, K., Thompson, R. N., White, L. F., Huisman, J. S., Scire, J., Bonhoeffer, S., Stadler, T., Wallinga, J., Funk, S., Lipsitch, M., & Cobey, S. **(2020)**. Practical considerations for measuring the effective reproductive number, R_t. *PLOS Comput. Biol.*, **16(12):** 1–21.

[12] Heesterbeek, H., Anderson, R. M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., Eames, K. T. D., Edmunds, W. J., Frost, S. D. W., Funk, S., Hollingsworth, T. D., House, T., Isham, V., Klepac, P., Lessler, J., Lloyd-Smith, J. O., Metcalf, C. J. E., Mollison, D., Pellis, L., Pulliam, J. R. C., Roberts, M. G., Viboud, C., & Isaac Newton Institute IDD Collaboration.**(2015)**. 'Modeling infectious disease dynamics in the complex landscape of global health. *Science*, **347(6227):** aaa4339.

[13] Keeling, M. J. & Rohani, P. **(2008)**. *Modeling infectious diseases in humans and animals*. Princeton University Press.

[14] Kermack, W. O., McKendrick, A. G., & Walker, G. T. **(1927)**. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Series A*, **115(772):** 700–721.

[15] Kumar, V., Alshazly, H., Idris, S. A., & Bourouis, S. **(2021)**. Evaluating the impact of COVID-19 on society, environment, economy, and education. *Sustainability*, **13(24):** 13642.

[16] Li, K., Wang, J., Xie, J., Rui, J., Abudunaibi, B., Wei, H., Liu, H., Zhang, S., Li, Q., Niu, Y., & Chen, T. **(2023)**. Advancements in defining and estimating the reproduction number in infectious disease epidemiology. *China CDC Wkly*, **5(37):** 829–834.

[17] Lipsitch, M., Cohen, T., Cooper, B., Robins, J., Ma, S., James, L., Gopalakrishna, G., Chew, S., Tan, C., Samore, M., Fisman, D., & Murray, M. **(2003)**. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, **300(5627):** 1966–1970.

[18] Ministry of Health and Population, Nepal **(2020–2022)**. MOHP Nepal: COVID-19 Updates and Situation Reports. *Available at:* `https://www.mohp.gov.np`.

[19] Mizumoto, K. & Chowell, G. **(2020)**. Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship. *Infect. Dis. Model.*, **5:** 264–270.

[20] Nishiura, H. & Chowell, G. **(2009)**. The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends: In *Mathematical and Statistical Estimation Approaches in Epidemiology*, 103–121. Springer Netherlands, Dordrecht.

[21] Our World in Data. **(2020-2022)**. COVID-19 Data Explorer. *Available at:* `https://docs.owid.io/projects/etl/api/covid/` Accessed: [2024-08-20].

[22] Pantha, B., Acharya, S., Joshi, H. R., & Vaidya, N. K. **(2021)**. Inter-provincial disparity of COVID-19 transmission and control in Nepal. *Sci. Rep.*, **11:** 13363.

[23] Pokhrel, A. K., Joshi, Y. P., & Bhattarai, S. **(2020)**. Epidemiological trend of COVID-19 in Nepal and the importance of social distancing to contain the virus. *Applied Sci. Tech. Ann.*, **1(1):** 15–25.

[24] Song, P. & Xiao, Y. **(2022)**. Estimating time-varying reproduction number by deep learning techniques. *J. Appl. Anal. Comput.*, **12(3):** 1–13.

[25] World Health Organization. **(2020-2022)**. Coronavirus disease (COVID-19) pandemic/WHO Coronavirus (COVID-19) Dashboard/WHO Nepal: COVID-19 Situation Reports. *Available at:* `https://www.who.int/nepal` Accessed: [2024-03-30].

□□