Journal of Statistical Modelling: Theory and Applications
Vol. 6, No. 2, 2025, pp. 101-115
Yazd University Press 2025

JSMTA

Research Paper

Hybrid modeling of COVID-19 progression in Iran: Active case estimation from February 2020 to March 2022

Toktam VazifehDoust Ahmadi[1], Sayyed Reza Alavian[1], Mehdi Jabbari Nooghabi[1], Nasrin Talkhi[2], Somayeh Ghiasi Hafezi[2], Parastoo Tajzadeh[3,*]
[1]Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran
[2]Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran
[3]Department of Medical Laboratory Sciences, Kashmar Faculty of Medical Sciences, Mashhad University of Medical Sciences, Iran

Received: September 29, 2025/ Revised: May 30, 2026/ Accepted: June 07, 2026

Abstract: The COVID-19 pandemic has unfolded in multiple waves across Iran, presenting complex challenges for public health management. Accurate modeling of active cases is essential for understanding transmission dynamics and guiding interventions. This study aims to investigate the progression of active COVID-19 cases across six epidemic waves in Iran using statistical, machine learning, and mathematical models to estimate key transmission rates and evaluate model performance. Models including autoregressive integrated moving average, multilayer perceptron, Holt-Winter, trigonometric seasonal, Prophet, and Bayesian structural time series were evaluated using root mean square, mean absolute, and mean absolute percentage errors. Mathematical approaches were applied to estimate transmission rates: infected to active, active to recovered, and active to death. Comparative performance was assessed across all six waves. Model performance varied by wave and variable. Multilayer perceptron and Bayesian structural time series showed superior accuracy for infected and active cases, respectively, while autoregressive integrated moving average excelled in predicting recoveries and deaths. Felberg consistently outperformed other models in estimating death cases. Transmission rates revealed that infected to active peaked during the third and fourth waves and declined post-vaccination. The fifth wave showed increased active to recovered and reduced active to death, indicating improved recovery and reduced mortality. Combining statistical, machine learning, and mathematical models offers a robust framework for analyzing COVID-19 dynamics. Transmission rate estimation provides actionable insights for epidemic control, highlighting the impact of vaccination and public health compliance on disease progression.

*Corresponding author: `Tajzadehp@mums.ac.ir`

Keywords: Active cases, COVID-19, Epidemic waves, Felberg method, IACR optimization, Iran, Machine learning, Statistical modeling, Transmission rate.
Mathematics Subject Classification (2010): 92D30.

1 Introduction

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, was first identified in Wuhan, China, and rapidly spread across the globe. This respiratory illness presents with symptoms such as fever, cough, and shortness of breath, and in severe cases can lead to pneumonia, acute respiratory distress syndrome, renal failure, and death. On January 30, 2020, the World Health Organization (WHO) declared COVID-19 a global public health emergency (World Health Organization, 2020). In Iran, the first confirmed case was reported in Qom Province on February 19, 2020, followed by rapid transmission in Gilan, Markazi, and Tehran provinces. In response, the National Corona Committee and provincial task forces were established to manage and contain the outbreak (Ministry of Health and Medical Education of Iran, 2020; World Health Organization, 2019).

Throughout the six waves of the pandemic in Iran, tracking the number of active cases-defined as individuals currently infected, whether symptomatic or asymptomatic-has been critical for public health decision-making. Active cases are calculated by subtracting recoveries and deaths from the total confirmed cases. This metric is essential for early detection and prevention of communicable diseases and serves as a key indicator of epidemic dynamics (Ghazanfari et al., 2024).

Several studies in Iran have explored the use of statistical and mathematical models to forecast COVID-19 trends. Talkhi et al. (2021) compared time series models and found that the multilayer perceptron (MLP) and Holt-Winter models yielded the lowest prediction errors for confirmed cases and fatalities (Talkhi et al., 2021). Pourghasemi et al. (2020) conducted a comprehensive study involving regression analysis, spatial modeling, risk mapping, and machine learning techniques such as random forest to assess COVID-19 mortality and transmission patterns (Pourghasemi et al., 2020). Zeroual et al. (2020) evaluated six deep learning models-including long short-term memory (LSTM), bidirectional-LSTM (Bi-LSTM), Convolutional-LSTM (Conv-LSTM), Bi-Conv-LSTM, gated recurrent unit (GRU), and Bi-GRU-for predicting new cases and deaths using datasets from Iran and Australia (Zeroual et al., 2020).

Active Cases indicate the total number of individuals infected with COVID-19, including both symptomatic and asymptomatic cases, identified through self-reporting or testing. This figure plays a crucial role for health policymakers in assessing the ongoing progression of the disease. It can be determined by subtracting the number of deaths and recoveries from the total confirmed cases, which is important to early diagnosis of commutable diseases as first pathway of prevention (Solanki et al., 2022).

Given the importance of active case dynamics in epidemic control, this study aims to investigate the effectiveness of statistical and mathematical models in estimating the transition rates from active cases to recoveries and deaths across six distinct waves of COVID-19 in Iran. The findings may provide valuable insights for policymakers to enhance disease prevention and control strategies.

2 Materials and methods

2.1 Data cxollection

COVID-19 case data were obtained from official Iranian health sources, including daily counts of infected, active, recovered, and death cases. The dataset spans six epidemic waves from February 27, 2020, to March 19, 2022, and reflects national-level reporting across all provinces.

2.2 Data preprocessing

To ensure model reliability, the dataset was preprocessed by handling missing values using linear interpolation. For each wave, the data were split into training (80%) and testing (20%) subsets to evaluate model generalizability. Time series normalization and smoothing techniques were applied where necessary to reduce noise and improve convergence in machine learning models.

2.3 Modeling techniques

To estimate and compare the dynamics of infected, active, recovered, and death cases, a combination of statistical, machine learning, and mathematical models was employed:
• **Statistical Models:**
– *ARIMA* (AutoRegressive Integrated Moving Average)
– *Holt-Winter* exponential smoothing
– *TBATS* (Trigonometric, Box-Cox, ARMA, Trend, and Seasonal components)
– *Prophet* (developed by Facebook for time series forecasting)
• **Machine Learning Models:**
– *MLP* (Multi-Layer Perceptron)
– *Hybrid-e* (ensemble-based hybrid model)
– *BSTS* (Bayesian Structural Time Series)
– *NNETAR* (Neural Network AutoRegressive model)
– *ELM* (Extreme Learning Machine)
• **Mathematical Models:**
– *Rang-Kutta Felberg method* for numerical solution of differential equations,
– *IACR minimum optimization model*, a modified three-dimensional SIR-based framework representing (Infected–Active Case–Recovered) dynamics. This model captures the transmission behavior of SARS-CoV-2 by estimating key transition rates between compartments.

2.4 Performance metrics

Model performance was evaluated using three standard metrics:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2},$$

$$\begin{aligned} MAE &= \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \\ MAPE &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|y_i - \hat{y}_i|}{y_i} \times 100\right), \end{aligned}$$

where y_i represents the actual value and $\hat{y}_i$ the predicted value at time i. These metrics were computed separately for training and testing datasets to assess model accuracy and robustness.

A brief overview of these models is provided:

1. **Neural Network Auto Regression Model (NNETAR**): A statistical learning and a parametric non-linear model named Neural Network Auto Regression Model that used in forecasting problems (Sena and Nagwani, 2016).

2. **Auto-Regressive Integrated Moving Average Model (ARIMA):** The ARIMA method, introduced by Box and Jenkins, is renowned and extensively applied in time series modeling (Yonar et al., 2020). ARIMA offers two advantages: The first is its high interpretability, and the second is maximize prediction accuracy (Papastefanopoulos et al., 2020). These models consist of three combination components: the auto-regressive (AR) model, the moving average (MA) model, and a white noise process.

3. **Hybrid model**: In R software, there are suitable functions for ensemble forecasts. Within the 'forecast Hybrid' package, default ensemble forecasts are created by combining forecasts from auto.arima(), ets(), thetaf(), nnetar(), stlm(), tbats(), and snaive() using equal weights. Alternatively, weights can be determined based on in-sample errors as proposed by Bates and Granger (1969) or through cross-validation. Cross-validation is employed to assess model accuracy and is compatible with user-defined models and forecasting functions.

4. **Holt-Winter (HW):** The Holt-Winter model, also known as triple exponential smoothing, is employed for forecasting in univariate time series data (Papastefanopoulos et al., 2020). This model is simple and easily automated that does not need high data-storage requirements. The HW is well-suited for short-term forecasting and utilizes the maximum likelihood function to estimating parameters (Awajan et al., 2018).

5. **Bayesian Structural Time-Series (BSTS)**: BSTS constructs analytical models by incorporating prior distribution (prior experience) and observed data (likelihood function). The final Bayesian model is derived by multiplying the prior distribution with the likelihood function to form the posterior distribution (Jun, 2019).

6. **TBATS model**: The acronym BATS represents the key components of the method, including Box–Cox transform, ARMA errors, Trend, and Seasonal components (Papastefanopoulos et al., 2020). The TBATS model is created by substituting the seasonal components in the BATS model with a trigonometric seasonal formulation, (De Livera et al., 2011). This model has some advantages. Firstly, it is computationally efficient for maximum likelihood estimations and can effectively decompose complex seasonal

time series. Secondly t encompasses a wide parameter space, leading to improved forecasting capabilities and the ability to handle nonlinear characteristics present in real-world time series data (Papastefanopoulos et al., 2020).

7. **Prophet**: This time series forecasting model is specifically tailored to address business time series challenges. It incorporates the three primary components of time series data (trend, seasonality, and holidays) within a decomposable framework to enhance forecasting accuracy. Generally, the model employs a saturating growth model and a piecewise linear model to predict trends effectively (Papastefanopoulos et al., 2020). The related forecasting tool is available in R that called prophet.

8. **Multi-Layer Perceptron (MLP)**: The MLP, a neural network model inspired by the biological neuron model, is structured with a minimum of three layers comprising inputs, weights, biases, and an activation function that generates the output (Parhizkari et al., 2020).

9. **Extreme Learning Machines (ELM)**: This algorithm is known for its rapid learning capabilities in single hidden layer feed-forward neural networks. This method has two benefits, the first is high learning speed due to reducing the training time and the second is improving the generalization performance (Pontoh et al., 2020). The initial weights and biases of the hidden layer are randomly assigned without fine-tuning, and the least squares method resolves the global optimization to determine the output weights of the hidden layer (Lai et al., 2020).

10. **Susceptible-Infected-Recovered (SIR):** The SIR compartmental model is a mathematical representation of infectious disease dynamics that partitions the population into three distinct groups. Susceptible (S) denotes individuals who have not been infected by COVID-19 or are susceptible to reinfection. Infected (I) represents those currently infected, and Recovered (R) includes individuals who have either recovered or succumbed to COVID-19. The model is characterized by a set of differential equations in Supplement 1. The SIR model operates under the assumption of a uniform and consistent population, with its behavior dictated by the parameters β (transmission rate) and γ (recovery rate). In this model, each infected individual comes into contact with β people. As only susceptible individuals can contract the infection, the likelihood of encountering a susceptible person is determined by the proportion of susceptible individuals in the total population, denoted as $N = S + I + R$. Additionally, at each time step, a fraction λ of the infected population transitions to the recovered state within the system (Vega et al., 2022).

11. **Infected-Active Case-Recovered (IACR)**: It is an adjusted model of SIR (by defining $\alpha = 0$ due to S (t) being unknown) that active and death cases are added in. Active case (AC) is the current number of people confirmed to be infected by COVID-19 but not recovered or dead yet and their situation is indeterminate. It represents an important metric for Public Health and Emergency response authorities, when assessing hospitalization needs versus capacity. Also, death (D) is patients died from the COVID-19. In this model is assumed that the recovered people have an infection possibility

by COVID-19 at any time. The differential equations of this model is defined due to Figure 1 and demonstrated in Supplement 1, where η, $\lambda$and δ parameters is related to infected, active case, and death rates respectively.

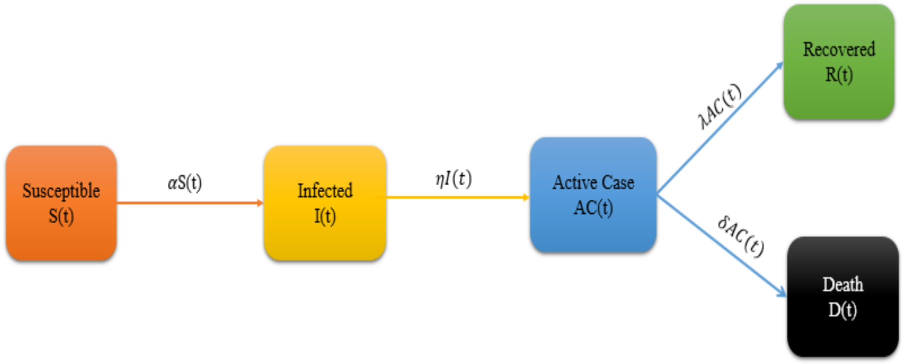

Figure 1: diagram of IACR dynamic model.

The purpose of current study is to compare infected, recovered, and death rates in COVID-19 pandemic active cases in Iran estimated by statistical, machine learning and mathematical models. This pandemic is consisted of six waves so far, which the first one started from 27th February 2020 and the sixth one ended on 19th March 2022. One of the most useful mathematical models in study epidemics is IACR model, faced every population nowadays. There are two methods to estimate IACR components that are Rang-Kota Felberg method and minimum optimization.

In the First estimation way, we shall apply Rang-Kota Felberg method (Butcher, 2016) to obtain the solution of IACR differential equation in Supplement 1 by a simulation in R (R Core Team, 2026) with deSolve package. This package resolves differential equations' initial value problems and is available in CRAN database (Soetaert et al., 2010a,b). The obtained answer of IACR differential equation is considered as an estimate of its components. In the second method, the parameters of IACR differential equation are estimated by minimum optimization problem in Supplement 4 that the solution is the parameters η, λ, δ. The optimization is done by the constroptim() function of the stats package that installed in the R software base automatically (Lange, 2001). This function minimizes a subject function to linear inequality constraints using an adaptive barrier algorithm in Supplement 4. The solution of this optimization problem is presumed as components of IACR.

In this study, a dataset is calculated from worldometers database (https://www.worldometers.info/coronavirus/country/iran) to survive the status of COVID-19 pandemic, contained of infected, dead, recovered and active cases caused by this infection in Iran. There are nine models is explanted in Section 2 and their results will be compared with Rang-Kota Felberg method and IACR minimum optimization estimation. These outputs will be demonstrated and interpreted in the Section 3 (Al-Qaness et al., 2020).

3 Results

The analysis was conducted using Iran's COVID-19 dataset spanning six epidemic waves between February 27, 2020, and March 19, 2022. The dataset was split into training (80%) and testing (20%) subsets. Various statistical and machine learning models were evaluated based on performance metrics including RMSE, MAE, and MAPE.The optimal model for each wave and parameter (infected, active, recovered, death) was selected based on the lowest values of RMSE, MAE, and MAPE. These models were then compared with two mathematical approaches: The Rang-Kutta Felberg method and the IACR minimum optimization model.

Table 1: Comparative performance of statistical, machine learning, and mathematical models across six COVID-19 waves

			Training Data			Testing Data		
Wave	Variables	Models	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	Infected	MLP	159.55	125.78	10.23	155.21	155.21	11.19
		Felberg	685.44	2226.10	184.04	318.87	508.39	42.58
		Optim	598.98	1968.94	137.05	462.19	607.50	40.84
	Active	BSTS	662.89	543.79	7.71	980.11	650.12	4.72
		Felberg	5342.74	14906.37	91.88	2699.30	3922.44	26.75
		Optim	17623.92	13771.46	87.76	4129.10	4352.95	22.74
	Recovered	ARIMA	325.15	241.99	58.07	414.12	382.54	23.41
		Felberg	876.56	1530.49	103.46	274.82	377.13	26.99
		Optim	893.46	1061.08	77.53	257.03	278.17	17.31
	Death	ARIMA	10.93	8.33	12.48	18.743	16.32	18.92
		Felberg	3.09	9.13	11.31	1.49	2.47	2.83
		Optim	38.35	40.99	41.72	11.85	12.13	12.22
2	Infected	HW	157.05	128.09	5.36	177.73	134.37	7.27
		Felberg	864.67	3397.70	141.96	436.43	855.68	35.04
		Optim	508.50	1175.39	47.55	444.20	536.98	21.26
	Active	TBATS	276.26	219.32	0.85	293.50	236.51	0.78
		Felberg	8430.41	33748.34	129.05	4365.03	8603.99	32.43
		Optim	9246.65	34514.56	129.60	4294.30	7674.32	28.50
	Recovered	ARIMA	209.75	153.48	6.57	270.2811	236.76	11.78
		Felberg	788.89	2994.12	132.18	413.6160	786.53	33.26
		Optim	512.14	1857.90	86.25	257.0098	446.88	19.94
	Death	Prophet	14.15	11.12	7.03	18.52	16.26	13.91
		Felberg	5.05	17.94	10.33	2.56	4.47	2.52
		Optim	23.00	81.66	50.36	10.94	18.21	11.06
3	Infected	Hybrid-e	397.96	296.59	4.21	227.37	188.98	2.99
		Felberg	2512.85	11525.69	155.22	1195.38	2744.65	39.56
		Optim	3975.41	19057.41	257.22	2092.20	4574.33	62.68
	Active	Hybrid-e	1650.80	1248.28	0.84	19324.90	14858.46	9.82
		Felberg	35497.98	170963.57	119.45	17102.87	39983.83	29.18
		Optim	149057.83	307075.09	188.47	54451.08	71149.13	48.21
	Recovered	Prophet	528.83	412.11	7.83	1131.82	936.25	14.80
		Felberg	1648.03	7770.82	122.78	789.24	1809.04	30.08
		Optim	5709.36	12097.32	174.31	1833.70	2680.51	41.52
	Death	TBATS	21.33	16.16	6.29	29.04	26.15	31.82
		Felberg	4.62	18.70	6.35	2.31	4.73	1.67
		Optim	118.87	129.91	59.70	17.21	22.21	13.74

Continuation of Table 1.

Waves	Variables	Models	Training Data RMSE	Training Data MAE	Training Data MAPE	Testing Data RMSE	Testing Data MAE	Testing Data MAPE
4	Infected	TBATS	1929.09	1529.38	11.45	2046.17	1755.51	22.21
		Felberg	3190.00	11140.00	73.48	2220.00	3830.00	21.33
		Optim	5990.42	23618.39	171.29	3961.48	6820.91	39.25
	Active	MLP	1811.91	1480.90	0.41	5529.72	4690.15	1.60
		Felberg	57100.00	240000.00	68.20	26810.00	53280.00	15.81
		Optim	89321.88	372662.30	104.18	141696.82	140984.12	39.03
	Recovered	ARIMA	1045.50	790.62	6.09	2160.97	1961.00	12.99
		Felberg	2270.00	9390.00	70.34	1210.00	2230.00	16.73
		Optim	2318.33	9703.40	73.11	3696.54	3651.93	26.96
	Death	ARIMA	37.28	28.34	11.77	45.67	41.72	23.85
		Felberg	0.89	3.04	1.16	0.46	0.78	0.29
		Optim	23.28	90.04	39.19	39.48	41.07	19.50
5	Infected	MLP	2608.27	1899.02	12.33	4142.22	3934.37	55.65
		Felberg	6510.00	23600.00	128.53	4092.00	8435.00	36.13
		Optim	11210.10	19300.67	120.17	11161.52	12652.75	50.43
	Active	BSTS	6850.98	4548.39	0.99	8524.11	7926.513	16.17
		Felberg	103000.00	517500.00	177.40	57260.00	147900.00	39.05
		Optim	206507.70	989497.63	311.98	288157.55	448380.28	109.61
	Recovered	HW	1354.33	1039.18	4.95	3477.06	3081.25	62.21
		Felberg	5340.00	28010.00	181.22	2850.00	7520.00	40.42
		Optim	9603.85	42524.67	257.73	13022.99	19232.26	101.42
	Death	NNETAR	28.81	21.03	6.85	37.65	34.40	61.06
		Felberg	1.56	6.53	2.39	0.98	2.51	0.60
		Optim	73.37	179.70	97.68	151.11	143.92	60.58
6	Infected	MLP	3697.30	2746.44	18.99	5733.23	5271.48	60.96
		Felberg	8120.00	17000.00	73.90	4039.21	4263.00	17.46
		Optim	11711.46	18265.48	87.31	1466.67	1563.99	12.78
	Active	Holt-	5633.11	4572.59	1.90	6632.70	5049.71	2.68
		Felberg	48030.00	131900.00	46.99	23970.36	309280.00	11.08
		Optim	85179.37	183475.47	73.57	111872.67	87536.59	37.86
	Recovered	MLP	3388.49	2666.58	17.89	4128.36	3432.35	32.67
		Felberg	3580.00	8730.00	53.51	1910.05	2287.00	13.07
		Optim	3381.50	7290.54	56.42	4990.41	4270.60	26.56
	Death	TBATS	18.33	14.17	13.92	34.38	31.48	35.00
		Felberg	0.33	0.89	0.58	0.18	0.23	0.14
		Optim	33.11	43.78	27.03	84.02	44.28	26.56

In the first wave:
1. MLP was the best model for estimating infected cases.
2. BSTS performed best for active cases.
3. ARIMA was optimal for recovered and death cases.

Across all waves, the Felberg model consistently outperformed others in estimating death cases. However, for other IACR components (infected, active, recovered), the best-performing model varied by wave due to differences in performance metrics. Figures 2 and 3 illustrate this comparison across epidemic peaks.

Transmission rates were calculated using mathematical models only. Based on performance:
1. The Felberg model was selected to estimate the transmission rate from active to death cases (δ).
2. The IACR minimum optimization mode was used to estimate transmission rates from active to recovered (λ) and infected to active cases (η). For example, in the first

wave:
1. $\eta = 0.1$: 10% of infected individuals transitioned to active cases daily.
2. $\lambda = 0.04$: 4% of active cases recovered daily.
3. $\delta = 0.008$: 0.8% of active cases resulted in death daily.

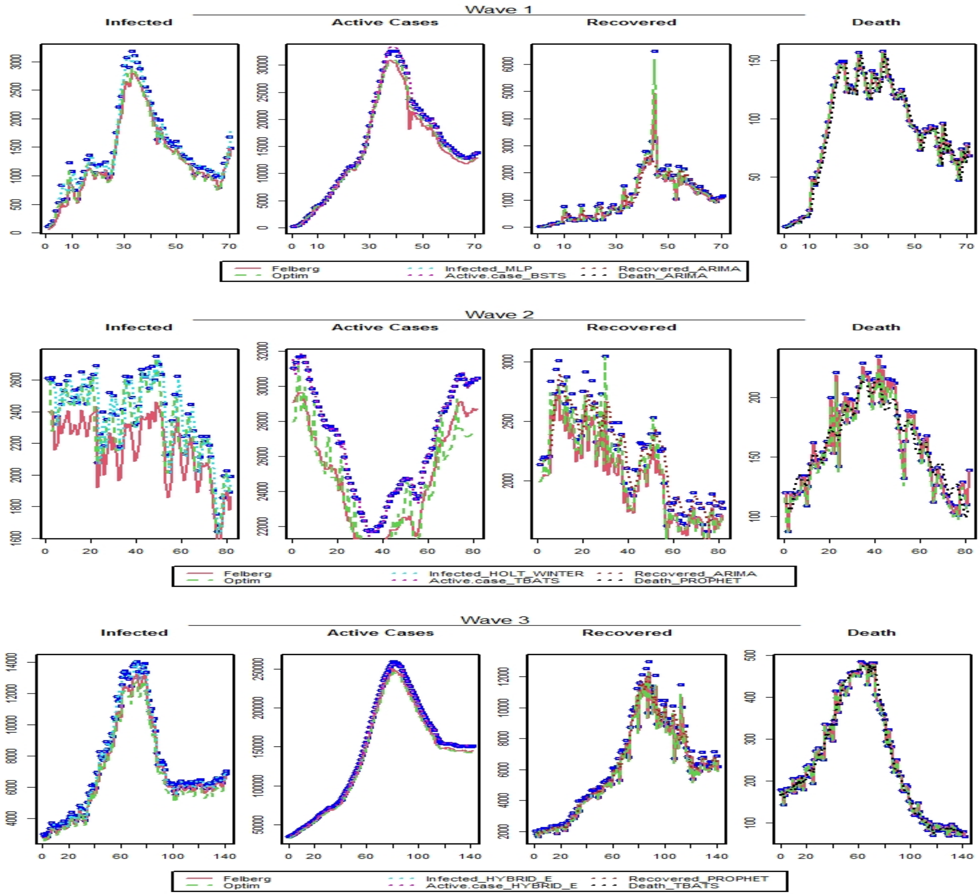

Figure 2: The estimation comparison of best statistical and mathematical model with Rang-Kota Felberg and IACR minimum optimization models in the sixth wave: Waves 1-3.

Table 2 summarizes transmission rates across all six waves. The third and fourth waves exhibited the highest (η) rate (9%), which declined significantly following vaccination efforts. In contrast, the sixth wave showed a resurgence in cases due to reduced compliance with health protocols. The impact of vaccination is evident in the fifth wave, where (λ) increased and (δ) decreased, reflecting improved recovery and reduced mortality. Additionally, virus type and incubation period influenced transmission dynamics, particularly in asymptomatic cases.

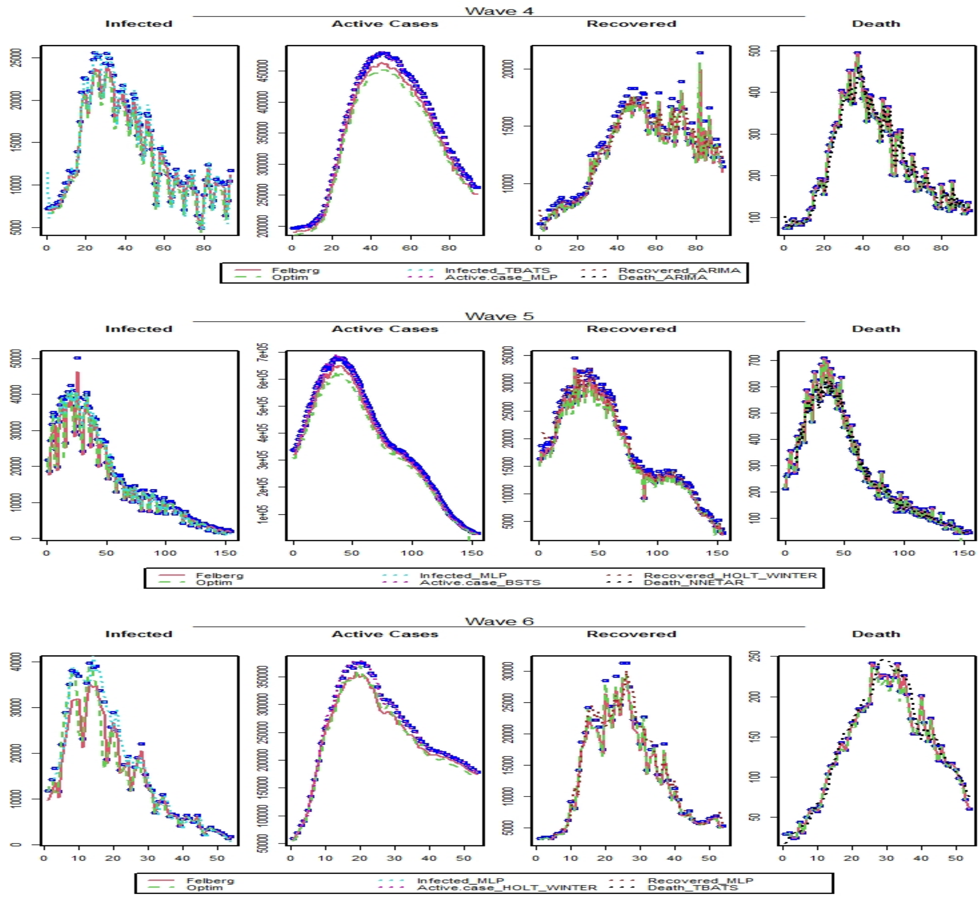

Figure 3: The estimation comparison of best statistical and mathematical model with Rang-Kota Felberg and IACR minimum optimization models in the sixth wave: Waves 4-6.

Table 2: Transmission Rate of parameters in the mathematical models.

Wave	η	λ	δ
1	0.10	0.04	0.008
2	0.02	0.05	0.006
3	0.09	0.05	0.002
4	0.09	0.03	0.0006
5	0.04	0.07	0.0007
6	0.08	0.04	0.0005

4 Discussion

The COVID-19 pandemic has underscored the critical role of predictive modeling in understanding disease dynamics and informing public health strategies. In Iran, the progression of six distinct epidemic waves between 2020 and 2022 provided a unique opportunity to evaluate the performance of statistical, machine learning, and mathematical models in forecasting active cases, recoveries, and fatalities. Recent studies have demonstrated the effectiveness of hybrid modeling approaches. Lotfi et al. (2022) developed a robust optimization model using regression techniques to forecast COVID-19 trends in Iran, achieving superior accuracy compared to traditional smoothing methods. Similarly, Davarian et al. (2025) proposed a multigroup epidemic model incorporating optimal control strategies to reduce asymptomatic and symptomatic transmission, validated using provincial data from Isfahan, Fars, and Khorasan-Razavi. Deep learning models have also shown promise. Abdollahi et al. (2021) compared LSTM, ARIMA, and Prophet models, concluding that seasonal artificial neural networks and LSTM yielded the lowest error rates in predicting new cases and deaths. Zeroual et al. (2020) extended this work by benchmarking six deep learning architectures, including Bi-LSTM and Conv-LSTM, for time series prediction in Iranian and Australian datasets. From a spatial perspective, Pourghasemi et al. (2020) utilized random forest algorithms to generate risk maps and forecast mortality trends, highlighting regional disparities in transmission and healthcare capacity. Ghasemi et al. (2024) applied statistical modeling to mortality data across epidemic phases, identifying ARIMA and cubic regression as optimal predictors.

The current study builds on these findings by comparing statistical and machine learning models with mathematical frameworks such as the Rang-Kutta Felberg and IACR minimum optimization methods. The Felberg model consistently outperformed others in estimating death cases across all waves, while the IACR model provided more accurate estimates for recovery and infection transitions. These results are consistent with the work of Lotfi et al. (2022), who emphasized the importance of model robustness under parameter uncertainty (Lotfi et al., 2022). Transmission rates- η (infected to active), λ (active to recovered), and δ (active to death) - were computed using mathematical models. The third and fourth waves exhibited the highest η values (up to 9%), reflecting rapid transmission prior to widespread vaccination. In contrast, the fifth wave showed increased λ and reduced δ rates, indicating improved recovery outcomes due to vaccination efforts. This aligns with findings by Ghazanfari et al. (2024), who modeled the burden of COVID-19 and demonstrated the impact of policy interventions on mortality reduction . These quantitative findings on transmission and outcome rates have direct implications for public health policy. The marked drop in δ (active to death) during the vaccination campaign provides strong, model-based evidence for the life-saving impact of vaccination, underscoring the need for sustained and equitable vaccine rollout. Furthermore, the high η (transmission) rates observed prior to vaccination highlight the critical window for implementing stringent non-pharmaceutical interventions during future epidemic waves. Public health officials can use such model-derived parameters to calibrate the timing and intensity of interventions, optimizing resource allocation between treatment centers and preventive measures.

The resurgence in the sixth wave, attributed to reduced compliance with health protocols and the emergence of new variants, highlights the need for adaptive modeling. Studies by Vega et al. (2022) emphasized the importance of integrating epidemiological mechanisms with machine learning approaches to improve predictive model responsiveness. Finally, the integration of multiple modeling approaches-statistical, machine learning, and mathematical-offers a comprehensive toolkit for epidemic analysis. As shown by Papastefanopoulos et al. (2020), hybrid forecasting approaches combining statistical and machine learning models can improve prediction accuracy and support real-time decision-making. Despite the insights provided, this study has several limitations. The accuracy of all models is contingent on the quality and granularity of the underlying epidemiological data, which may be affected by under-reporting or diagnostic capacity variations across different waves and provinces. Furthermore, the mathematical models rely on assumptions of homogeneity and constant parameters within each wave, which may not fully capture complex, real-time human behavioral dynamics or sub-population heterogeneity. Future work would benefit from integrating real-time mobility data, socioeconomic factors, and genomic surveillance data to refine parameter estimation and improve model fidelity.

References

Abdollahi, J., Irani, A.J. and Nouri-Moghaddam, B. (2021). Modeling and forecasting Spread of COVID-19 epidemic in Iran until Sep 22, 2021, based on deep learning. *arXiv preprint arXiv:2103.08178.*

Al-Qaness, M.A., Ewees, A.A., Fan, H. and Abd El Aziz, M. (2020). Optimization method for forecasting confirmed cases of COVID-19 in China. *Journal of clinical medicine,* **9**(3):674.

Awajan, A.M., Ismail, M.T. and Al Wadi, S. (2018). Improving forecasting accuracy for stock market data using EMD-HW bagging. *PLoS One,* **13**(7):e0199582.

Bates, J.M. and Granger, C.W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, **20**(4), 451-468.

Butcher, J.C. (2016). *Numerical Methods for Ordinary Differential Equations.* John Wiley & Sons.

Davarian, N., Heydari, A., Hashemi Mehne, S.H. and Heydari, A.A. (2025). Investigation of stability and optimal control of the multigroup COVID-19 epidemic model in Iran. *Journal of Mathematical Extension,* **19**(1):1–32.

De Livera, A.M., Hyndman, R.J. and Snyder, R.D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association,* **106**(496):1513–1527.

Ghasemi, E., Khorshidi, A., Omidi, M. and Sayehmiri, K. (2024). Statistical modeling of COVID-19 mortality trends in Iran. *Journal of Basic Research in Medical Sciences,* **11**(3):73–83.

Ghazanfari, S., Meskarpour-Amiri, M., Hosseini-Shokouh, S.M., Teymourzadeh, E., Mehdizadeh, P. and Salesi, M. (2024). Designing a model to estimate the burden of COVID-19 in Iran. *BMC Public Health,* **24**(1):2609.

Jun, S. (2019). Bayesian structural time series and regression modeling for sustainable technology management. *Sustainability,* **11**(18):4945.

Lai, J., Wang, X., Li, R., Song, Y., and Lei, L. (2020). BD-ELM: A regularized extreme learning machine using biased dropconnect and biased dropout. *Mathematical Problems in Engineering,* **2020**(1):3604579.

Lange, K. (2001). *Numerical Analysis for Statisticians.* Springer, New York.

Lotfi, R., Kheiri, K., Sadeghi, A. and Babaee Tirkolaee, E. (2022). An extended robust mathematical model to project the course of COVID-19 epidemic in Iran. *Annals of operations research,* **339**(3):1499–1523.

Ministry of Health and Medical Education of Iran. (2020). *COVID-19 Updates and Guidelines.* Tehran: Ministry of Health and Medical Education.

Papastefanopoulos, V., Linardatos, P. and Kotsiantis, S. (2020). COVID-19: a comparison of time series methods to forecast percentage of active cases per population. *Applied Sciences,* **10**(11):3880.

Parhizkari, L., Najafi, A. and Golshan, M. (2020). Medium term electricity price forecasting using extreme learning machine. *Journal of Energy Management and Technology,* **4**(2):20–27.

Pontoh, R.S., Solichatus, Z., Hidayat, Y., Aldella, R., Jiwani, N.M. and Sukono (2020). COVID-19 modelling in South Korea using a time series approach. *International Journal of Advanced Science and Technology,* **29**(7):1620–1632.

Pourghasemi, H.R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S.R., Babaei, S., Khosravi, R., Etemadi, M., Ghanbarian, G., Farhadi, A., Safaeian, R., Heidari, Z., Tarazkar, M.H., Tiefenbacher, J.P., Azmi, A. and Sadeghian, F. (2020). Spatial modeling and risk mapping of COVID-19 in Iran using machine learning techniques. *International Society for Infectious Diseases,* **98**:90-108.

R Core Team (2026). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Sena, D. and Nagwani, N.K. (2016). A neural network autoregression model to forecast per capita disposable income. *ARPN Journal of Engineering and Applied Sciences,* **11**(22):13123–13128.

Soetaert, K., Petzoldt, T. and Setzer, R.D. (2010a). deSolve: General solvers for initial value problems of ordinary differential equations, partial differential equations, differential algebraic equations and delay differential equations. *R Package Version 1.8.*

Soetaert, K., Petzoldt, T. and Setzer, R.W. (2010b). Solving differential equations in R: Package deSolve. *Journal of Statistical Software,* **33**(9):1–25.

Solanki, R., Varshney, A., Gourishetty, R., Minase, S., Sivadas, N. and Mahajan, A. (2022). Correction in active cases data of COVID-19 for the US states by analytical study. *Disaster Medicine and Public Health Preparedness,* **16**(6):2335–2338.

Talkhi, N., Akhavan Fatemi, N., Ataei, Z. and Jabbari Nooghabi, M. (2021). Modeling and forecasting number of confirmed and death caused COVID-19 in Iran: A comparison of time series forecasting methods. *Biomedical Signal Processing and Control,* **66**:102494.

Vega, R., Flores, L. and Greiner, R. (2022). SIMLR: Machine learning inside the SIR model for COVID-19 forecasting. *Forecasting,* **4**(1):72–94.

World Health Organization. (2020). Statement on the second meeting of the International Health Regulations Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). *WHO Reports.*

World Health Organization. (2020). *Coronavirus disease (COVID-19) pandemic.* Geneva: WHO.

Yonar, H., Yonar, A., Tekindal, M.A. and Tekindal, M. (2020). Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing methods. *Eurasian Journal of Medicine and Oncology,* **4**(2):160–165.

Zeroual, A., Harrou, F., Dairi, A. and Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos, Solitons and Fractals,* **140**:110121.

Appendix

Supplement 1

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

Supplement 2

$$\begin{aligned}\frac{dS}{dt} &= -\alpha S(t),\\ \frac{dI}{dt} &= \alpha S(t) - \eta I(t),\\ \frac{dAc}{dt} &= \eta I(t) - (\lambda + \delta)Ac(t),\\ \frac{dD}{dt} &= \delta Ac(t),\\ \frac{dR}{dt} &= \lambda Ac(t).\end{aligned}$$

Supplement 3

This function minimizes a subject function to linear inequality constraints using an adaptive barrier algorithm. Hence, the constants of subject function are $1 > \psi_k \geq 0$:

$$\min_{\psi_k} \sum_{j=1}^{n} \sum_{i=1}^{4} \left(\frac{X_i(t_j) - X_i(t_{j-1})}{t_j - t_{j-1}} - (F_i(t_j).X(t_j).\psi_k) \right)^2, \qquad k = 1, \ldots, 4,$$

where the state variables is $X(t) = (I(t), AC(t), R(t), D(t))$, $\psi_k = (\eta, \lambda, \delta)$ and $F(t) = (\frac{dI}{dt}, \frac{dAc}{dt}, \frac{dR}{dt}, \frac{dD}{dt})$.