Appendix A. Supplementary Information for the main text

Appendix A.1. The basic reproduction number with a disease-free equilibrium

In our mathematical model, there is an equilibrium point at which the booster vaccination rate α_1=0 when the vaccination program is essentially over. At this time, it can be assumed that no infected individuals exist in the population, and the susceptible population is divided into three groups due to cross-immunity from previous infections: susceptible to both strains, susceptible to strain 1 (Delta strain), and susceptible to strain 2 (Omicron strains). Therefore, $E_0 = I_0 = E_V = I_V = L_0 = L_V = 0$, and the disease-free equilibrium $e_0 = \left(S^*, S_1^*, 0, 0, 0, S_0^*, 0, 0, 0, S_V^*\right)$, where S^*, S_1^*, S_0^*, S_V^* are both non-negative numbers. Using the next generation method, we can obtain the basic reproduction number in this disease-free equilibrium. The basic reproduction number R_0 is obtained by solving the following equation: $R_0 = \rho(FV^{-1})$ [1]. F are the derivatives of the new infections, V is the transition matrix of the flow between compartments and ρ is the spectral radius.

The matrix of new infections $\mathscr{F}$ is given by

$$\mathscr{F} = \begin{bmatrix} \beta\frac{I_0}{N}(S + S_V) \\ 0 \\ \frac{\beta_1 I_V(S+S_0)+\beta_2 I_V(S_1-\alpha_1 S_V)}{N} \\ 0 \end{bmatrix}.$$

While the transition matrix $\mathscr{V}$ is given by

$$\mathscr{V} = \begin{bmatrix} k_1 E_0 \\ -k_1 E_0 + (\gamma_1 + \delta_1) I_0 \\ k_2 E_V \\ -k_2 E_V + (\gamma_2 + \delta_2) I_0 \end{bmatrix}.$$

Hence, the Jacobian matrices are given by

$$F = \begin{bmatrix} 0 & \frac{\beta S^* + \beta S_V^*}{N^*} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\beta_1 S^* + \beta_1 S_0^* + \beta_2 (S_1^* - \alpha_1 S_V^*)}{N^*} \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} k_1 & 0 & 0 & 0 \\ -k_1 & (\delta_1+\gamma_1) & 0 & 0 \\ 0 & 0 & k_2 & 0 \\ 0 & 0 & -k_2 & (\delta_2+\gamma_2) \end{bmatrix}.$$

Therefore,

$$FV^{-1} = \begin{bmatrix} \frac{\beta S^* + \beta S_V^*}{N^*(\delta_1+\gamma_1)} & \frac{\beta S^* + \beta S_V^*}{N^*(\delta_1+\gamma_1)} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\beta_1 S^* + \beta_1 S_0^* + \beta_2 (S_1^* - \alpha_1 S_V^*)}{N^*(\delta_2+\gamma_2)} & \frac{\beta_1 S^* + \beta_1 S_0^* + \beta_2 (S_1^* - \alpha_1 S_V^*)}{N^*(\delta_2+\gamma_2)} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The characteristic polynomial of FV^{-1} is given by $\lambda(\lambda - D)(\lambda - O)$, then the basic reproduction number is given by $R_0 = max\{|D|, |O|\}$.

$$R_0^D = \frac{\beta S^* + \beta S_V^*}{N^*(\delta_1 + \gamma_1)}. \tag{A.1}$$

$$R_0^O = \frac{\beta_1 S^* + \beta_1 S_0^* + \beta_2 (S_1^* - \alpha_1 S_V^*)}{N(\delta_2 + \gamma_2)}. \tag{A.2}$$

where $R_0(D)$, $R_0(O)$ are the basic reproduction numbers of strain 1 and strain 2, respectively. To further evaluate the effect of vaccines as well as cross-immunity on the basic reproduction number. The disease-free equilibrium point was given by $\left(S^*, S_1^*, 0, 0, 0, \frac{\eta_2 L_0^*}{\alpha_1}, 0, 0, 0, \frac{\eta_1 L_V^*}{\alpha_1}\right)$. And we assume that the percentages of S^*, S_1^*, L_0^*, L_V^* in the total population are $\theta^*, {\theta_1}^*, {\theta_0}^*, {\theta_V}^*$ and their percentages can be computed by booster vaccination coverage and the protective power given by cross-immunity. Therefore, the basic reproduction number is given by $R_0 = max\left\{R_0^D, R_0^O\right\}$. we simplified the basic reproduction numbers for each of the two strains as:

$$R_0^D = \frac{\beta(\theta^* + \eta_1 {\theta_V}^*)}{\alpha_1(\delta_1 + \gamma_1)}. \tag{A.3}$$

$$R_0^O = \frac{\beta_1(\alpha_1\theta^* + \eta_2{\theta_0}^*) + \alpha_1\beta_2({\theta_1}^* - \eta_1\theta_V^*)}{\alpha_1(\delta_2 + \gamma_2)}. \tag{A.4}$$

Appendix A.2. Stability of the disease- free equilibrium

The Disease-Free Equilibrium of our system is unstable if $R_0 > 1$ while it is locally asymptotically stable. The Jacobian matrix of our system at the disease-free equilibrium e_0 is given by

$$J_M = \begin{bmatrix} -\alpha_1 - \lambda & 0 & 0 & -\frac{\beta S^*}{N^*} & 0 & -\frac{\beta_1 S^*}{N^*} & 0 & 0 & 0 & 0 \\ \alpha_1 & -\lambda & 0 & 0 & 0 & -\frac{\beta_2(S_1^* - \alpha_1 S_V^*)}{N^*} & 0 & 0 & \alpha_1 & \alpha_1 \\ 0 & 0 & -k_1 - \lambda & \frac{\beta S^* + \beta S_V^*}{N^*} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_1 & -(\delta_1 + \gamma_1) - \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -k_2 - \lambda & \frac{\beta_1 S^* + \beta_2(S_1^* - \alpha_1 S_V^*) + \beta_1 S_0^*}{N^*} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_2 & -(\delta_2 + \gamma_2) - \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_1 & 0 & 0 & -\eta_2 - \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma_2 & 0 & -\eta_1 - \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{\beta_1 S_0^*}{N^*} & \eta_2 & 0 & -\alpha_1 - \lambda & 0 \\ 0 & 0 & 0 & -\frac{\beta S_V^*}{N^*} & 0 & 0 & 0 & \eta_1 & 0 & -\alpha_1 - \lambda \end{bmatrix}.$$

The Jacobian matrix has a null eigenvalue λ_1, five negative eigenvalues $\lambda_2 = \lambda_3 = \lambda_4 = -\alpha_1$, $\lambda_5 = -\eta_1$, $\lambda_6 = -\eta_2$. And the other eigenvalues are given by

$$\begin{bmatrix} -k_1 - \lambda & \frac{\beta S^* + \beta S_V^*}{N^*} \\ k_1 & -(\delta_1 + \gamma_1) - \lambda \end{bmatrix}$$

and

$$\begin{bmatrix} -k_2 - \lambda & \frac{\beta_1 S^* + \beta_2(S_1^* - \alpha_1 S_V^*) + \beta_1 S_0^*}{N^*} \\ k_2 & -(\delta_2 + \gamma_2) - \lambda \end{bmatrix}.$$

Their characteristic polynomials are

$(k_1 + \lambda)(\delta_1 + \gamma_1 + \lambda) - k_1\frac{\beta S^* + \beta S_V^*}{N^*} = 0$

and $(k_2 + \lambda)(\delta_2 + \gamma_2 + \lambda) - k_2\frac{\beta_1 S^* + \beta_2(S_1^* - \alpha_1 S_V^*) + \beta_1 S_0^*}{N^*} = 0.$

That is,

$\left(1 + \frac{\lambda_7}{k_1}\right)\left(1 + \frac{\lambda_8}{\delta_1 + \gamma_1}\right) = R_0^D, \left(1 + \frac{\lambda_9}{k_2}\right)\left(1 + \frac{\lambda_{10}}{\delta_2 + \gamma_2}\right) = R_0^O.$

Thus, when $R_0^D < 1, R_0^O < 1$, which means $\lambda_7, \lambda_8, \lambda_9, \lambda_{10} < 0$, at this moment, the system is stable. When $R_0^D > 1$ or $R_0^O > 1$, there exists at least one $\lambda_i > 0$ $(i = 7, 8, 9, 10)$, Hence, the system is unstable [2, 3].

Appendix A.3. Fitting real data from Malaysia, Iceland, and Bahrain

The numerical simulation results for three countries, South Korea, Denmark and Spain, are shown in the Fig A.1 The simulations include the daily cases as well as cumulative cases. The results show that our model is well-fitted to the real reported data. In addition, we fit the Omicron sequence percentages for a number of countries besides those in the main text. In Fig A.2, our model fits the proportion of Omicron sequences for each country separately very well, by comparing and fitting the

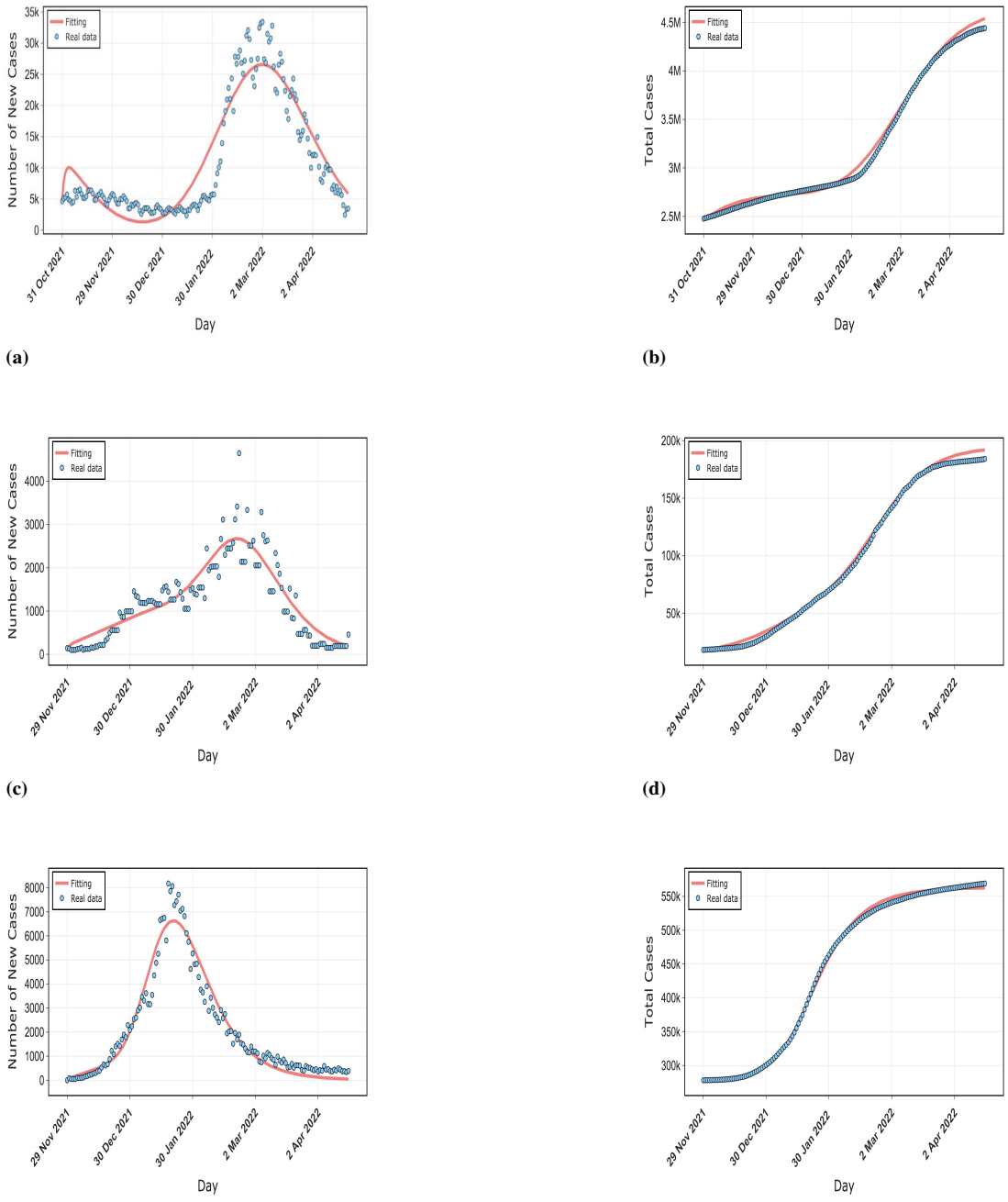

Fig. A.1. Comparison of our model results with real data on the epidemic in Malaysia, Iceland, and Bahrain. (a),(d),(e) shows the fit of the model to the daily new cases. (b) shows the fit of the model to the cumulative cases.

data for multiple countries, we obtain the timing of the replacement of previous strains by Omicron strains and the competitive advantage of each strain, as shown in section 4 of the main text.

Appendix A.4. Comparison of the predictive performance of the models

This subsection aims to compare the prediction performance from our model (multi-strain model), The traditional transmission SEIR model, machine learning and deep learning techniques on COVID-19 dataset for seven countries including Denmark, Malaysia, Spain, Iceland, Korea, Bahrain, and South Africa all over the globe. LSTM, Stacked LSTM,GRU, and Bidirectional LSTM are taken as deep learning models for sequential predictions. The training set is 120 days and the test set is 7, 30, 60, 120 days respectively. Both LSTM,GRU and BiLSTM have a three-layer network with 256, 256 and 30 neurons per

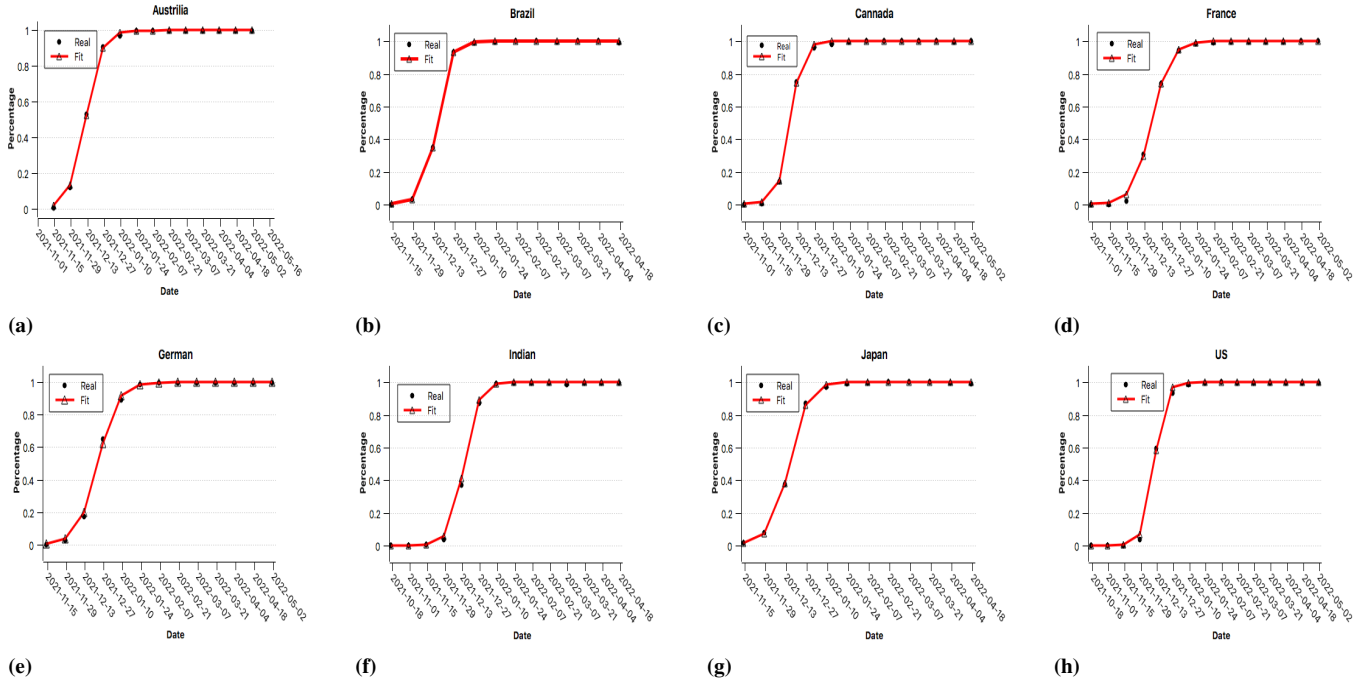

Fig. A.2. Comparison of Omicron sequence percentage fitting results with real data for each country. The red solid line indicates the simulation result of the model, and the black hollow triangle indicates the real data.

Table A.1: Validation Metrics for Total cases COVID-19 forecasting using SEIR, LSTM, BiLSTM, ST-LSTM, GRU,and Multi-strain models.

Predicted days	Models	R^2	MSE	$RMSE$	MAE	$MAPE$	$SMAPE$
7 *Days*	**Multi-strain model**	0.3617	4.85×10^{11}	**696686.5704**	**670379.6835**	**8.5701**	**8.1699**
	LSTM	0.3094	2.56×10^{12}	1600733.6841	1297212.3170	16.4966	15.1604
	BiLSTM	**0.8444**	1.33×10^{12}	1156398.5685	950167.6898	11.5816	11.8603
	ST-LSTM	0.6268	2.03×10^{12}	1426037.6088	1161184.6082	14.7856	13.7933
	GRU	0.5877	2.14×10^{12}	1463502.1113	1191411.2367	15.1512	14.1130
	SEIR	-0.9014	1.44×10^{12}	1202543.9112	1113804.1814	13.2166	14.2628
30 *Days*	**Multi-strain model**	**0.0.9481**	3.66×10^{11}	**605610.3559**	**585892.3154**	**5.3040**	**5.1380**
	LSTM	0.3108	1.62×10^{13}	4031824.8036	3334720.9012	27.5387	30.8842
	BiLSTM	0.6216	1.38×10^{13}	3726051.7885	3071526.3982	26.0797	27.9793
	ST-LSTM	0.4174	1.53×10^{13}	3919467.5193	3238241.1291	26.9380	29.8226
	GRU	0.5878	1.74×10^{13}	4176708.5427	3466143.1701	27.5351	25.7431
	SEIR	0.4278	4.04×10^{12}	2012131.2154	1925828.2464	16.1196	17.6354
60 *Days*	**Multi-strain model**	**0.8780**	1.22×10^{12}	**1108909.7615**	**1001466.1963**	**6.8017**	**6.5484**
	LSTM	-0.2354	3.06×10^{13}	5534794.7374	4626671.2296	32.2277	37.5170
	BiLSTM	0.1258	3.45×10^{13}	5877198.3903	4934520.2774	40.7825	33.0800
	ST-LSTM	0.1346	3.76×10^{13}	6135027.9231	5191191.4167	42.2210	34.7309
	GRU	0.1229	2.73×10^{13}	5228768.4593	4394749.2377	31.0043	28.5822
	SEIR	0.0750	9.32×10^{12}	3054100.1913	2496637.0082	16.9124	16.6585
120 *Days*	**Multi-strain model**	**0.7630**	2.02×10^{12}	**1423417.3546**	**1339816.1121**	**8.0177**	**7.6867**
	LSTM	-0.6379	4.37×10^{13}	6612067.9305	5742102.0704	40.2196	34.4087
	BiLSTM	0.0763	3.58×10^{13}	5987247.3086	4994265.1605	34.3915	32.8880
	ST-LSTM	-0.0773	2.66×10^{13}	5162518.1460	4173290.6045	27.0041	30.34795
	GRU	0.0461	2.55×10^{13}	5058781.0851	4050323.8159	26.4129	29.3212
	SEIR	-12.5084	1.15×10^{14}	10747682.2970	8493327.6885	48.1990	36.4588

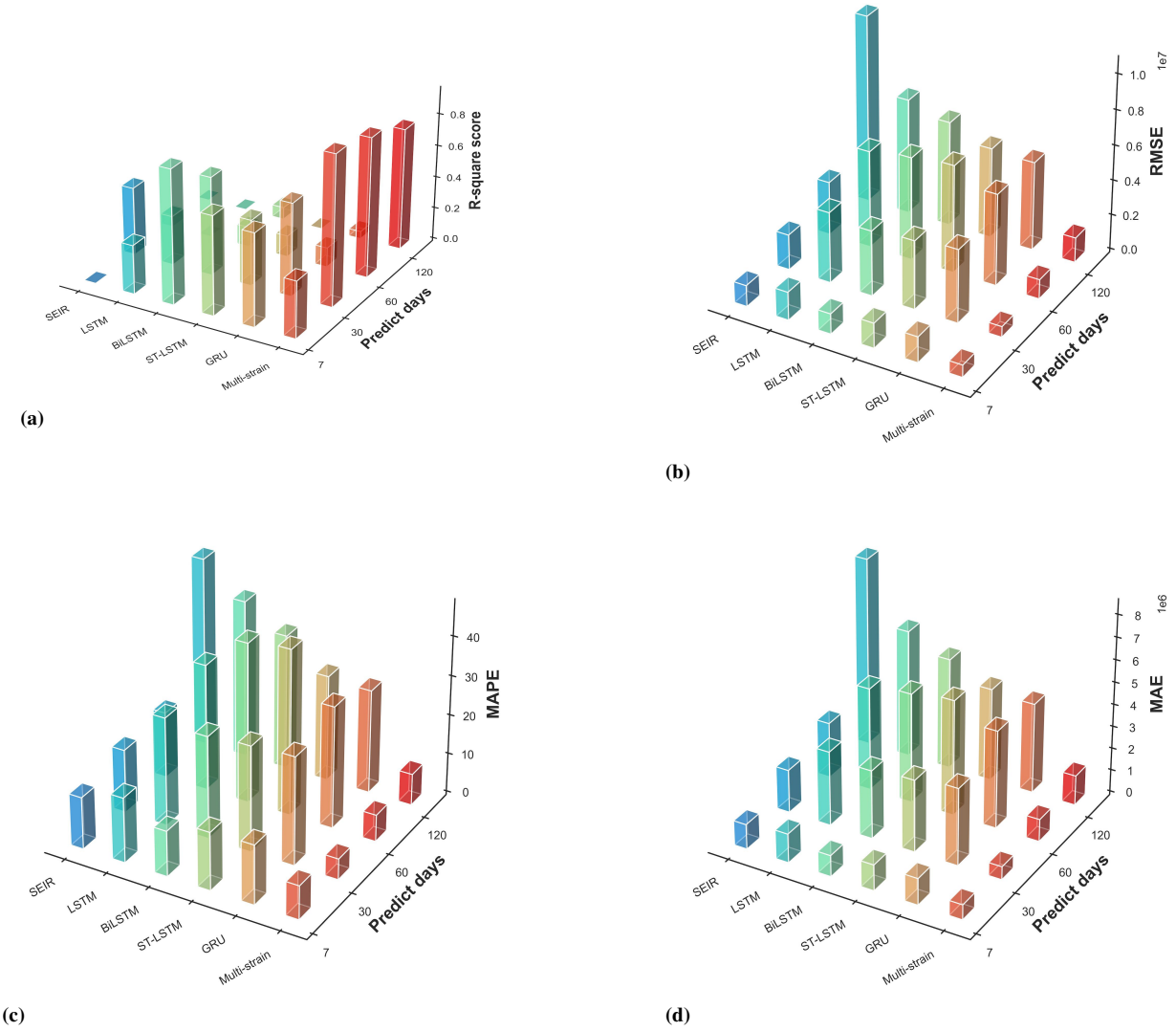

Fig. A.3. Comparison among deep learning techniques ,SEIR model, and our model in terms of different error measures. (a) indicates the coefficient of determination, the closer to 1 means the more accurate the prediction result. (b), (c), (d) indicate Root Mean Square Error, Mean Absolute Percentage Error, and Mean Absolute Error respectively, the lower their values are means the prediction result is better.

layer, ST-LSTM stacking another layer of the network between the lstm layers. The Optimizer is Adam with 0.01 learning rate and epochs are 1000. In our model and in the SEIR model, all parameters are set as realistically as possible, as shown in Table **??** and the fitting method is the leastsq. In traditional transmission models such as the SEIR model, the prediction of epidemics is often only for early transmission. Therefore, it is no longer applicable for the current stage. Deep learning techniques can make short-term predictions of time series, but their results are uninterpretable. Table A.1 and Fig. A.3 also show that the above methods no longer have reference value when the prediction time length reaches 30 days. In contrast, our model outperforms in long-term forecasts and significantly improves over traditional models. The reasons and preconditions for the good performance of our model are analyzed in the discussion.

Our model shows well prediction performance on datasets from seven countries, including South Korea, Malaysia, Iceland, Spain, Denmark, South Africa, and Bahrain. It is worth mentioning that better prediction results require the training set to contain the turning points of the epidemic. The training set without the turning points of the epidemic still outperforms the traditional prediction method, but the generated time series is only the general trend of epidemic development.

As we can see in the table A.1, our model achieves the best scores in almost all evaluation metrics compared to the other models, except for the R-square score in 7-day short-term prediction, which is due to the fact that our model focuses more on long-term time series prediction. To better visualize the prediction performance of each model at different time lengths, we selected four evaluation metrics R-square score, RMSE, MAPE and MAE, and plotted their predictive ability in different dimensions in a three-dimensional diagram.

Table A.2: Mean values of model parameters corresponding to the situation of South Korea

Parameter	Description	South Korea	Malaysia	Iceland	Denmark	Bahrain	Spain	South Africa
α_1	Booster vaccination rate	0.0112	0.0091	0.0085	0.0428	0.0464	0.0602	0.0370
β	Transmission coefficient of strain 1	0.3121	0.3246	0.4260	0.3930	1.2632	0.2112	0.1000
β_1	Relative transmission coefficient of strain 2	0.8715	0.9412	0.8607	1.7130	2.111	3.2000	1.6466
β_2	Transmission coefficient after immune escape	0.3123	0.1777	0.3080	0.4215	0.3067	0.3168	0.1056
k_1	Inverse of the latency period of strain 1	0.25	0.25	0.25	0.25	0.25	0.25	0.25
k_2	Inverse of the latency period of strain 2	0.37	0.37	0.37	0.37	0.37	0.37	0.37
δ_1	Death rate due to strain 1	0.0798	0.0514	0.0285	0.0776	0.1000	0.0499	0.0999
δ_2	Death rate due to strain 2	0.009	0.0099	0.0099	0.0023	0.1000	0.0100	0.009
γ_1	Rate of infectious loss of strain 1 infected patients	0.2089	0.3079	0.2251	0.1997	0.3000	0.0300	0.0908
γ_2	Rate of infectious loss of strain 2 infected patients	0.3198	0.3898	0.3310	0.2999	0.3000	0.3260	0.2500
η_1	Rate of loss of cross-immunity to strain 1	3×10^{-4}	4×10^{-4}	3×10^{-6}	6×10^{-8}	3×10^{-4}	5×10^{-8}	10^{-5}
η_2	Rate of loss of cross-immunity to strain 2	0.007	0.005	8×10^{-5}	5×10^{-7}	0.008	9×10^{-4}	0.009
N	Total country population	51780000	32800000	400000	5900000	1700000	47300000	60000000

Fig. A.3 shows that when the number of forecast days is 7 days, Almost all models except the SEIR model showed well predictive performance. However, as the number of prediction days reaches 30 days, the prediction results of deep learning models become terrible, with a very small R^2 (for visual representation, when $R^2 < 0$ are noted as 0 in the figure) and a large RMSE, among other evaluation metrics. At this point, all traditional models no longer have reference value. When the prediction time reaches 120 days, the R^2 of other models is already all less than 0 and significantly different from the actual cases. Only our model can match the actual wells on the prediction of such a long time series. The specific fit plots are shown in the section Appendix A.3 and main text. It is worth mentioning that Bi-LSTM has the best performance in deep learning techniques for short-term forecasting and R^2 is also the largest among the models, but it is still our model optimal among other evaluation metrics. Therefore, we can choose Bi-LSTM in combination with our model for shorter-period time series prediction, and firmly choose our model for long-term time series prediction.

the predictive performance of our model performs well in the following conditions.

- Competitive behavior between strains exists.
- The data in the training set already shows a preliminary trend of exponential growth. If the turning point of the epidemic is included, then the predictions will be more accurate. (The 120-day training set contains the turning point of the epidemic.)
- If the general trend of the epidemic development is not a single peak such as in South Africa, changes in competing strains need to be considered.

Appendix A.5. Parameter estimation

The parameters fitted to the real data for each country are shown in Table A.2. The parameters for each country were taken as realistically as possible, and the specific range of values is similar to the evaluation of the parameters in the main text.

Appendix A.6. Competitive advantages of various lineages

We examined the subvariants of Omicron in a more refined way. By fitting the evolutionary trends between various competing variants using the multi-strain model for a total of 15 countries including those mentioned in the Appendix. The time of mutual substitution of each strain and its daily growth of the log-odds rate were evaluated by fitting a Multi-strain mathematical model to the daily proportions of the competing variants. The time taken to increase the proportion of a strain sequence from 5% to 50% was calculated assuming that only two strains existed and using the pairwise difference in their growth rates. Because of the assumed symmetry, the time taken to grow the sequence of a strain from 5% to 95% is twice as long as the time is taken to grow the sequence from 5% to 50%. The table A.3 shows that the shortest time for the Omicron strain to replace the non-Omicron strain as the dominant strain was 14 days, which also means that the competitive relationship between them is strong. The BA.2.12.1 took the longest time to

Table A.3: Comparison of competitive advantage among strains. comparing daily growth rates of Delta, Omicron, and Omicron subvariants and substitution times among strains.

Lineages competing	Growth-rate advantage	Time for growth from 5 to 50% (days)	Time for growth from 5 to 95% (days)
Omicron versus Delta (December 2021 to January 2022)	0.209	14	28
BA.2 versus non-BA.2 Omicron (January to May 2022)	0.105	28	56
BA.2.12.1 versus non- BA.2.12.1 Omicron (May to July 2022)	0.053	50	100
BA.4/ BA.5 versus non- BA.4/ BA.5 Omicron (May to July 2022)	0.105	28	56

replace the non-BA.2.12.1 strain as the dominant strain, about 50 days, which also means that the competitive relationship between them is the weakest. Moreover, BA.4/5 was found to be the most adaptive variant with the strongest competitive advantage at this stage, due to the fact that no strains are currently capable of replacing BA.4/5. What factors determine the outcome of competition between strains? And what factors influence the strength of competition between strains? We will give the answer in the main text.

References

[1] P. Van den Driessche, J. Watmough, Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, Mathematical biosciences 180 (1-2) (2002) 29–48.

[2] Q. Kang, Y. Song, Q. Ding, W. P. Tay, Stable neural ode with lyapunov-stable equilibrium points for defending against adversarial attacks, Advances in Neural Information Processing Systems 34 (2021) 14925–14937.

[3] A. Beavers, Jr, E. Denman, A new solution method for the lyapunov matrix equation, SIAM Journal on Applied Mathematics 29 (3) (1975) 416–421.