Drivers and forecasts of multiple waves of the coronavirus disease 2019 pandemic: A systematic analysis based on an interpretable machine learning framework

Zicheng Cao1,2 | Zekai Qiu1,2 | Feng Tang1,3 | Shiwen Liang1,4 | Yinhgan Wang1,5 | Haoyu Long1,2 | Cai Chen1 | Bing Zhang1,2 | Chi Zhang1,2 | Yaqi Wang1,2 | Kang Tang1,2 | Jing Tang1,2 | Junhong Chen1,2 | Chunhui Yang1,2 | Yuzhe Xu1,2 | Yulin Yang1,2 | Shenglan Xiao1,2 | Dechao Tian1,2 | Guozhi Jiang1,2 | Xiangjun Du1,2,6

1School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou, P.R. China
2School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen, P.R. China
3Foshan Center for Disease Control and Prevention, Foshan, P.R. China
4Fujian Provincial Center for Disease Control and Prevention, Fuzhou, P.R. China
5Clinical Research Center, Second Affiliated Hospital of Kunming Medical University, Kunming, P.R. China
6Key Laboratory of Tropical Disease Control, Ministry of Education Sun Yat-sen University, Guangzhou, P.R. China

Correspondence
Xiangjun Du, School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, P.R. China. Email: duxj9@mail.sysu.edu.cn

Zicheng Cao and Zekai Qiu contributed equally to this work.

Funding information
National Key Research and Development Projects of the Ministry of Science and Technology of China, Grant/Award Number: 2021YFC2301300; Guangdong Frontier and Key Tech Innovation Program, Grant/Award Numbers: 21202001262000006; 2019B020228001, 2019B11103001; Natural Science Foundation of Guangdong Province, China, Grant/Award Number: 2021A1515011592; Shenzhen Science and Technology Program, Grant/Award Number: KQTD20180411143323605

Abstract
Coronavirus disease 2019 (COVID-19) has become a global pandemic and continues to prevail with multiple rebound waves in many countries. The driving factors for the spread of COVID-19 and their quantitative contributions, especially to rebound waves, are not well studied. Multidimensional time-series data, including policy, travel, medical, socioeconomic, environmental, mutant and vaccine-related data, were collected from 39 countries up to 30 June 2021, and an interpretable machine learning framework (XGBoost model with Shapley Additive explanation interpretation) was used to systematically analyze the effect of multiple factors on the spread of COVID-19, using the daily effective reproduction number as an indicator. Based on a model of the pre-vaccine era, policy-related factors were shown to be the main drivers of the spread of COVID-19, with a contribution of 60.81%. In the post-vaccine era, the contribution of policy-related factors decreased to 28.34%, accompanied by an increase in the contribution of travel-related factors, such as domestic flights, and contributions emerged for mutant-related (16.49%) and vaccine-related (7.06%) factors. For single-peak countries, the dominant ones were policy-related factors during both the rising and fading stages, with overall contributions of 33.7% and 37.7%, respectively. For double-peak countries, factors from the rebound stage contributed 45.8% and policy-related factors showed the greatest contribution in both the rebound (32.6%) and fading (25.0%) stages. For multiple-peak countries, the Delta variant, domestic flights (current month) and the daily vaccination population are the three greatest contributors.
1  |  INTRODUCTION

An emerging infectious disease, coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2, has rapidly spread globally. On March 11, 2020, the World Health Organization declared COVID-19 as a global pandemic. Until 11 February 2022, there were 405,688,274 COVID-19 cases globally (Looi, 2020; Roser et al., 2020). The number of new cases continues to increase every day, and COVID-19 remains a global concern after 2 years of circulation.

Since July 2020, there has been a second wave of infections in many countries (Cacciapaglia et al., 2020; Wise, 2020; Xu & Li, 2020; Zhu et al., 2020), especially in Europe and Asia. Vaccine development was also ongoing in many countries during that period. In early 2021, countries around the world began to implement extensive vaccination programs with the aim of ending the COVID-19 pandemic (Awadassieid et al., 2021). However, new virus variants with greater transmissibility have emerged, combined with control measures, and caused multiple waves of record-breaking numbers of new infections (Mohapatra et al., 2022; Sonabend et al., 2021). The ongoing pandemic has become a long-term battle, not just for public health, as governments are required to balance economic recovery, political votes and control of the COVID-19 pandemic (Di Domenico et al., 2020; Ferrante et al., 2020; Leung et al., 2020). As a result, decision-makers have adjusted intervention policies dynamically, while considering many types of factors, including the status of the COVID-19 pandemic (Panovska-Griffiths et al., 2020; Qiu et al., 2020; Sebhatu et al., 2020), the characteristics of new virus variants, vaccination status, socioeconomic conditions (Kandel et al., 2020) and medical resources (Brauner et al., 2020), to contain the pandemic.

The combined effects of various control measures are complicated, and the effect could be context dependent for a specific country. Those understandings are important for precise control and more quantitative studies are needed. Various non-pharmaceutical interventions (Flaxman et al., 2020; Haug et al., 2020; Hsiang et al., 2020; Russell et al., 2021; Sun et al., 2020) implemented by many countries have been shown to be useful to contain and control the spread of the pandemic. Without restrictions, population migration, which may increase the risk of contact with infected people, may lead to the rapid spread of the disease (Cao et al., 2020; Kubota et al., 2020; van Oosterhout et al., 2021). This has been extensively studied based on data from many countries (Duhon et al., 2020). Soft power at the national level, including social and economic support and the reasonable optimization and mobilization of medical resources, has the potential to slow down and affect the spread of COVID-19 (Haider et al., 2020; Li et al., 2020). In addition, many studies have found that climate and other environmental factors may also play roles in the spread of COVID-19 (Price et al., 2019; Qu et al., 2020). Overall, the spread and control of COVID-19 is a complex process involving various complex heterogeneous factors that are intertwined (Han et al., 2022). This nonlinear system becomes even more complex with the emergence of new virus variants, the uncertain effect of vaccination on the human population (Sowa et al., 2021) and the heterogeneity of regional developmental levels. A well-developed interpretable machine learning framework (Ayoub et al., 2021; Murri et al., 2021) may adapt to this complex scenario and provide reasonable interpretation and predictive capabilities, which are critical for a greater understanding of the mechanisms underlying the effectiveness of various COVID-19 control measures. Ultimately, this will facilitate the development and implementation of targeted and precise prevention and control strategies for the ongoing COVID-19 pandemic.

In this study, factors and datasets related to the spread and control of COVID-19 were collected as much as possible from all over the world and used for further analysis. The effects of various factors on the spread of COVID-19 were studied based on a nonlinear interpretable machine learning framework. Contributing factors were identified and characterized for different phases of the pandemic, before and after vaccine implementation, and for countries with or without rebound waves. The driving factors and potential mechanisms were explored, and forecasting efforts were made.

2  |  MATERIALS AND METHODS

2.1  |  Study design

This study was designed as a quantitative study of the contribution of multi-dimensional factors to the spread of COVID-19, using the effective reproductive number as an indicator, and to the occurrence of multiple waves of infections. Specifically, this study focused on the following aspects: (1) the quantification of the overall contribution of multi-dimensional factors, before and after vaccine implementation; (2) a comparative analysis of the factors influencing the rising and
fading stages of the pandemic, in countries with or without rebound waves, and before and after vaccine implementation; and (3) rebound risk forecasting for the spread of COVID-19.

### 2.2 | Data

Daily numbers of COVID-19 cases were collected from 39 countries up to 30 June 2021. These 39 countries were selected based on the following principles: (1) they represented countries from five continents, (2) the total number of cases in these countries was among the highest during the study period and (3) data for multi-dimensional factors during the same period were available. The COVID-19 pandemic was divided into two phases—the pre-vaccine era and the post-vaccine era—using the date when vaccination began in each country as the dividing date. Of the 39 countries selected, seven single-peak countries without a rebound wave and 13 double-peak countries with a rebound wave were identified in the pre-vaccine era (Figure S1), while 36 multi-peak countries with multiple rebound waves (excluding China, Australia and New Zealand, which had no rebound wave) were identified in the post-vaccine era. A rebound wave was identified when the number of new COVID-19 cases exceeded the average value of the peak in the pre-vaccine era for three consecutive weeks.

The factors analyzed were divided into the following groups: policy-related factors (including 13 specific non-pharmaceutical interventions in the previous week and at the current time point), travel-related factors (including domestic and foreign air passenger flow in the previous month and the current month), socioeconomic factors (population, population density, per capita income, aging ratio and sex ratio), environmental factors (maximum temperature, minimum temperature, average temperature, relative humidity and absolute humidity) and medical factors (number of doctors per 1000 residents, number of nurses, number of pharmacists, number of beds, medical expenditure quota, cardiovascular death rate and diabetes prevalence). Additionally, the number of daily vaccinations was considered as a vaccine-related factor, and information regarding existing virus variants was included in the mutant-related factor category. All of these data were analyzed at the national level.

**Effective reproductive number:** The daily effective reproductive number \( R_t \) was used as an indicator of the spread of COVID-19. \( R_t \) indicates the average number of secondary cases caused by one infected person per day (Fernández-Naranjo et al., 2021), and it was calculated based on the Susceptible-Infected-Recovered model (Arroyo-Mariol et al., 2021; Kermack & Mckendrick, 1927). When \( R_t \leq 1 \), maintaining the current prevention and control measures is expected to gradually control an infectious disease epidemic. In contrast, when \( R_t > 1 \), the infectious disease will continue to spread, suggesting that the prevention and control measures need to be optimized and strengthened. The \( R_t \) values for COVID-19 in the 39 countries included in the analysis were obtained from Our World in Data (https://ourworldindata.org/covid-cases).

**Policy-related factors:** These data were obtained from the Oxford COVID-19 Government Response Tracker (OxCGRT, https://github.com/OxCGRT/covid-policy-tracker), which includes data on the level of specific measures (strictness of implementation) taken by governments during the outbreak. Thirteen specific measures were analyzed, as detailed in Table S1. These classification variables were converted into a score in the range of [0,100]. The details of this calculation are presented in the Materials and Methods section of the Appendix.

**Travel-related factors:** Data on air passenger flow, based on monthly statistical measures at the national level, were collected from the Official Aviation Guide. These data included global passenger flow to the countries (foreign flights) and intra-country air passenger flow (domestic flights). A 1-month lag time was also included (Cao et al., 2020) for these data.

**Socioeconomic factors:** The number of residents (population), population density (density), per capita national income (GDP), percentage of the population aged \( \geq 65 \) years (old%) and sex ratio (the ratio of men to woman, sex ratio) were collected from Our World in Data (https://ourworldindata.org/covid-cases).

**Medical factors:** Health resource data were obtained from the World Bank (https://data.worldbank.org). These data included the number of doctors (doctor), nurses (nurse), pharmacists (pharmacist) and beds (beds) and the health expenditure per 1000 people (health expenditure). Additional health data, such as the cardiovascular death rate (cardiovascular death rate) and diabetes prevalence (diabetes prevalence), were obtained from Our World in Data (https://ourworldindata.org).

**Environmental factors:** The daily maximum temperature, minimum temperature, average temperature, relative humidity and absolute humidity during the study period were collected from the National Oceanic and Atmospheric Administration (https://www.noaa.gov/).

**Mutant-related factors:** The five major virus strains (alpha, beta, delta, gamma and lambda) were considered, and the temporal proportions referred to Alpha for Beta, Delta, Gamma and Lambda in each region were included as free variables in the analysis. These proportions were based on the weekly time scale in 39 countries obtained from the Global Initiative on Sharing Avian Influenza Data (https://www.gisaid.org/hcov19-variants/).

**Vaccine factor:** The number of daily COVID-19 vaccinations (daily vaccination population) in the 39 countries was obtained from Our World in Data (https://ourworldindata.org/covid-cases).

All time scales were unified as daily scales using the pandas time-series resampling method (https://pandas.pydata.org/). Moreover, considering the latent period of COVID-19, lag effects for policy- and travel-related factors were also considered and incorporated into the model (see the Materials and Methods section in the Appendix for more details).

### 2.3 | Machine learning framework

Interpretable machine learning frameworks were used to determine the nonlinear contribution of systematic factors to the \( R_t \) value. Specifically, multiple candidates nonlinear regression models, including random forest, support vector machine and extreme gradient boosting...
(XGBoost) (Chen & Guestrin, 2016) models, were used, and important factors were selected using the sequential floating selection (SFS) algorithm (Fif et al., 1994; Pudil et al., 1994). The overall framework is shown in Figure S2. SFS is a sequential feature selection method based on the greedy search that can remove functions from all feature sets and evaluate error functions. If the error reaches the optimal level, the combination of the remaining features is regarded as the best feature combination. A Bayesian search algorithm was used to identify the root mean square error (RMSE) of the parameter sets by fitting the model. The effect of generalization was evaluated by 10-fold cross-validation (S.M. Lundberg et al., 2020). Two temporally non-overlapping datasets were built based on the pre-vaccine era (a training set up to 31 October 2020, and a testing set from 1 November to 30 November 2020) and post-vaccine era (a training set from January 2021 to 31 May 2021, and a testing set from 1 June to 30 June 2021) datasets. The model with the lowest fitting error was selected based on two training sets, and this model was used in subsequent analyses.

2.4 Ranking contribution of the factors

Shapley additive explanation (SHAP) is an interpretable analysis method (S.M. Lundberg et al., 2020) for the outputs of machine learning models. It treats all features as “contributors” and generates a predictive value (SHAP value). For each predicted sample, the value of each influencing factor, known as the feature value, for the sample was uniformly assigned. The SHAP value (S. Lundberg & Lee, 2017) was calculated as follows:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{iN}),$$

where the $i$th sample of the $N$th feature is defined as $x_{iN}$, the reference value of the model (the mean value of the sample target variable) is $y_{base}$ and $f(x_{iN})$ is the SHAP value of $x_{iN}$. A positive value indicates that the feature has a positive contribution. The predicted value of the effective reproductive number $y_i$ here is the linear combination of $y_{base}$ and $f(x_{iN})$. The SHAP value was obtained based on the best model and $f(x_{iN})$ was used to measure the contribution of factors selected by the SFS algorithm, which was further converted into a percentage contribution value (Pan et al., 2020).

2.5 Model evaluation and short-term pandemic rebound risk prediction

A rebound risk model was built by considering the daily increasing ($R_t > 1$) or decreasing ($R_t \leq 1$) trend as the subject (binary) and all factors selected as covariates. The RMSE and $R^2$ values were used to evaluate the model fit to $R_t$ using 10-fold cross-validation for the training set. The accuracy, precision, recall, F1 score and area under the curve were used to evaluate the rebound risk.
contribution of policy-related factors in several countries, such as Sweden, Mexico and Bulgaria, being as high as 70%. However, the contribution of travel-related factors in many countries, such as Spain, the United States and Cameroon, was 30%. Moreover, the contributions of socioeconomic, environmental and medical factors were low in all countries evaluated. Of note, there were noticeable differences in the contributions of these factors in different countries. For example, the contribution of socioeconomic factors in Colombia, Kenya and Chile was 16.17%, 14.89% and 14.64%, respectively; the contribution of environmental factors in Argentina, Ecuador and Colombia was 15.67%, 13.60% and 13.21%, respectively; and the contribution of medical factors in Algeria and Canada was 11.27% and 9.89%, respectively. In the post-vaccine era, a significant decrease in the dominance of policy-related factors was seen (Table S4), and the contribution of vaccine factor was relatively high in some countries, such as Bangladesh, Canada and South Africa, which showed contributions of 15.03%, 12.54% and 10.66%, respectively. At the same time, the contribution of variant viral strains in these countries was also prominent, reaching 32.67%, 25.92% and 28.16%, respectively.

3.2 Factors related to the rebound of COVID-19

As COVID-19 continued to spread, different countries responded differently, leading to different outcomes, with rebound waves occurring in many countries. In the pre-vaccine era, 13 of the 39 countries analyzed had a clear rebound wave with double peaks in the number of new cases, whereas seven countries had no rebound wave, but a clear single peak (Figure S1). In addition, the first peak was further divided into the rising and fading stages, and the second peak was classified as the rebound stage. In the post-vaccine era, 36 of the 39 countries analyzed experienced multiple rebound waves, with more than one peak. Models were then built for the seven double-peak countries with rebound waves, 13 single-peak countries without a rebound wave in the pre-vaccine era and 36 countries with multiple rebounds in the post-vaccine era (Table S5). These models had RMSE values of 0.634, 0.298 and 0.021 and $R^2$ values of 0.824, 0.712 and 0.988, respectively (Figure 2).

Overall, factors in the rising and fading stages had similar effects on the spread of COVID-19 in single-peak countries (Figure 2a). Policy- and travel-related factors were dominant during the rising stage (contributions of 33.7% and 18.7% in the rising and fading stages, respectively), but policy-related factors were predominant in the fading stage, reaching a contribution of 37.7% (Figure 2a). However, approximately half (45.8%) of the contribution to the spread of COVID-19 in double-peak countries was from factors in the rebound stage. Policy-related factors showed contributions of 32.6% and 25.0% during the rebound and fading stages, respectively (Figure 2b). In the post-vaccine era, the results of the model including countries with multiple rebounds were similar to the results of the model that included all 39 countries (Figure 1b), with policy- and travel-related factors accounting for more than half of the contribution, although the contribution from policy-related factors significantly decreased (Figure 2c). Mutant-related and vaccine factors contributed 17.7% and 7.3%, respectively, to the spread of COVID-19 in the post-vaccine era. The contributions from environmental and medical factors increased to 9.6% and 8.7% in the post-vaccine era.

In single-peak countries (Figure 2d), the contribution of domestic flights (previous month) in the rising stage was 18.71%, with effective SHAP values of 0.23, indicating a positive contribution to the spread of COVID-19. Other important factors in single-peak countries were policy-related factors, such as contact tracing (contribution, 9.90%; effective SHAP value, –1.14) and the cancellation of public events.
FIGURE 2 Contribution of the 20 most important factors and their effective SHAP values (sum of SHAP values) for $R_t$ at different stages of the pandemic in single-peak (a, d), double-peak (b, e) and multiple-peak (c, f) countries. The first peak includes the rising and fading stages, and the additional peaks were defined as the rebound stage. The colour for each group is the same as in Figure 1.

3.3 Rebound risk forecasting

After understanding the different contributions from various factors, we sought to determine whether the rebound risk of COVID-19 could be predicted. Predictive models were built based on data from 20 selected countries in the pre-vaccine era and 36 countries in the post-vaccine era. A high risk of rebound was defined as $R_t > 1$, and a low risk of rebound was defined as $R_t \leq 1$. Daily forecasting was then performed for the following 30 days, based on actual data for all of the contributing factors in the model. COVID-19 rebound was defined as $R_t > 1$ for more than half of the 30 days. The accuracy of the risk forecast was 0.78 and 0.81 in the pre-vaccine and post-vaccine eras, respectively (Table 1 and Table S6).

4 DISCUSSION

On the basis of available datasets, the contribution of systematical factors to the spread of COVID-19 was studied through an interpretable machine learning framework. Globally, policy-related and travel-related factors (Bielecki et al., 2021; Chinazzi et al., 2020; Jia et al., 2020) play essential roles in the spread of the COVID-19 epidemic during the stage when the vaccine was not yet widely available (Figure 1a). Compared with double-peak countries, the spread of COVID-19 in single-peak countries was mostly driven by domestic migration and policy-related factors (Figure 2d). Neglecting travel restrictions (both foreign and domestic flights) in the earlier stage and relaxing intervention policies in the later stages may increase the likelihood of pandemic resurgence (Figure 2e). This remained true for the post-vaccine era, during which unfavourable policy implementation and new virus variants with increased infectivity caused multi-peak rebounds, even after vaccine implementation (Figure 2f). A forecasting model based on this knowledge gave reasonable predictions for the risk of rebound. Our study sheds light on quantitatively understanding the mechanisms underlying the global spread of COVID-19 and will facilitate the precise and cost-effective control of the ongoing COVID-19 pandemic (Reddy et al., 2021).
## TABLE 1  Rebound risk forecasting for 39 countries

| Country       | Pre-vaccine era | Post-vaccine era |
|---------------|-----------------|------------------|
|               | Rebound (%)     | Rebound prediction (%) | Correct prediction | Rebound (%)     | Rebound prediction (%) | Correct prediction |
| Afghanistan   | 100 (30/30)     | 100 (30/30)       | ✓                 | 93.33 (28/30)   | 100 (30/30)            | ✓                 |
| Algeria       | –               | –                 | –                 | 100 (30/30)     | 100 (30/30)            | ✓                 |
| Argentina     | –               | –                 | –                 | 0 (0/30)        | 53.33 (16/30)          | X                 |
| Austria       | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Australia     | 36.67 (11/30)   | 3.33 (1/30)       | ✓                 | –               | –                     | –                 |
| Bangladesh    | –               | –                 | –                 | 100 (30/30)     | 86.67 (26/30)          | ✓                 |
| Bolivia       | –               | –                 | –                 | 100 (30/30)     | 100 (30/30)            | X                 |
| Bulgaria      | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Cameroon      | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Canada        | 100 (30/30)     | 100 (30/30)       | ✓                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Chile         | 26.67 (8/30)    | 3.33 (1/30)       | ✓                 | 26.67 (8/30)    | 63.33 (19/30)          | X                 |
| China         | 73.33 (22/30)   | 86.67 (26/30)     | ✓                 | –               | –                     | –                 |
| Colombia      | –               | –                 | –                 | 93.33 (28/30)   | 86.67 (26/30)          | ✓                 |
| Denmark       | 100 (30/30)     | 100 (30/30)       | ✓                 | 6.67 (2/30)     | 0 (0/30)               | ✓                 |
| Ecuador       | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Finland       | 100 (30/30)     | 100 (30/30)       | ✓                 | 43.33 (13/30)   | 0 (0/30)               | ✓                 |
| France        | 26.67 (8/30)    | 100 (30/30)       | ✓                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Germany       | 100 (30/30)     | 100 (30/30)       | ✓                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Greece        | –               | –                 | –                 | 16.67 (5/30)    | 0 (0/30)               | ✓                 |
| Iceland       | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| India         | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Indonesia     | –               | –                 | –                 | 100 (30/30)     | 26.67 (8/30)           | X                 |
| Ireland       | 0 (0/30)        | 0 (0/30)          | ✓                 | 23.33 (7/30)    | 0 (0/30)               | ✓                 |
| Italy         | 60 (18/30)      | 53.33 (16/30)     | ✓                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Japan         | 100 (30/30)     | 100 (30/30)       | ✓                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| Kenya         | 60 (18/30)      | 100 (30/30)       | ✓                 | 83.33 (25/30)   | 70 (21/30)             | ✓                 |
| Mexico        | –               | –                 | –                 | 100 (30/30)     | 6.67 (2/30)            | X                 |
| New Zealand   | 50 (15/30)      | 6.67 (2/30)       | X                 | –               | –                     | –                 |
| Norway        | 70 (21/30)      | 100 (30/30)       | ✓                 | 6.67 (2/30)     | 0 (0/30)               | ✓                 |
| Pakistan      | 100 (30/30)     | 100 (30/30)       | ✓                 | 6.67 (2/30)     | 36.67 (11/30)          | ✓                 |
| Panama        | –               | –                 | –                 | 100 (30/30)     | 100 (30/30)            | ✓                 |
| Philippines   | –               | –                 | –                 | 43.33 (13/30)   | 16.67 (5/30)           | ✓                 |
| Portugal      | –               | –                 | –                 | 100 (30/30)     | 100 (30/30)            | ✓                 |
| Romania       | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| South Africa  | 83.33 (25/30)   | 60 (18/30)        | ✓                 | 100 (30/30)     | 100 (30/30)            | ✓                 |
| South Korea   | 100 (30/30)     | 100 (30/30)       | ✓                 | 50 (15/30)      | 0 (0/30)               | ✓                 |
| Spain         | 40 (12/30)      | 100 (30/30)       | ✓                 | 73.33 (22/30)   | 0 (0/30)               | X                 |
| Sweden        | –               | –                 | –                 | 0 (0/30)        | 0 (0/30)               | ✓                 |
| United States | 100 (30/30)     | 83.33 (25/30)     | ✓                 | 20 (6/30)       | 0 (0/30)               | ✓                 |

Note: “-”: The country did not meet the selection requirements at this stage. Period of risk forecasting: November 1 to 30, 2020 (pre-vaccine era) and June 1 to 30, 2021 (post-vaccine era).
We found that foreign flights played a dominant role in the global spread of COVID-19 in the pre-vaccine era, with a combined contribution of 18.84% (Figure 1a) (Bo et al., 2021; Liu et al., 2020; Middelburg & Rosendaal, 2020; Wells et al., 2020; Wong et al., 2020). The contribution gradually shifted from foreign migration to domestic flights with time (Figure 1b). The most important policy-related factor was income support (previous and current week), with a combined contribution of 12.59%, implicating the importance of economic support and the complex behaviours of human society during an outbreak of an emerging infectious disease. The contributions of socioeconomic and medical factors were less critical than the dominant policy- and travel-related factors. However, their effects cannot be neglected. For example, countries with the greatest contributions from socioeconomic factors (Colombia, Kenya and Chile; Table S4) were underdeveloped regions, implying that the socioeconomic impact of the pandemic was more pronounced in these regions. During a pandemic, the effect of environmental factors is difficult to tease out due to the greater impact of other factors, but environmental factors have been shown to contribute to transmission (Jia et al., 2020; Malki et al., 2020; Yang et al., 2020). Some countries, such as Argentina, showed a relatively large contribution (15.76%) from environmental factors, although this was not a universal finding (Table S4). Overall, the differential implementation of control policies combined with the emergence of variants with greater infectivity has created additional uncertainties regarding recurrent global outbreaks (Figure 2e,f). The dynamic contribution of different factors together with their time-series data provides a visual description of the drivers and specific mechanisms of pandemic rebound (Figure 2 and Figure S4). In the case of Japan, the rebound in the pandemic (June–August, 2020) was the result of weak policy implementation, whereas Finland showed a rebound because of a combination of policy relaxation and weakened travel restrictions (Figure S4).

A comparative analysis of countries with and without rebound indicated that strict contact tracing and effective control of population movement during the rising stage, while maintaining control during the fading stage, were crucial factors for effective control of COVID-19 in single-peak countries (Figure 2a,d). However, countries that experienced a rebound of COVID-19 showed relaxed control policies in the later stage (Figure 2b,e), along with the promoting effect of population movement (both domestic and international) and relaxed testing policies and public event cancellation. The relaxation of restrictions should be carefully evaluated (The, 2020), especially after the implementation of a vaccination program, and restrictions are essential when considering additional reinforcement factors such as new variants (Figure 2f). Effectively controlling the COVID-19 pandemic and preventing a rebound require a systematic approach (Coccia, 2020; Sarkodie & Owusu, 2020). Therefore, the quantitative estimation of effective measures and their contributions informs precise prevention and control measures and provides important suggestions for policymakers, which have been shown to be crucial for controlling the pandemic (Ferrante et al., 2021; Greer et al., 2020).

The findings of this study also facilitate rebound risk forecasting. For this purpose, 1-month predictions for the selected countries in the pre- and post-vaccine eras were made with an accuracy of 0.78 and 0.81, respectively. The large amount of heterogeneity may explain the incorrect predictions in some countries, as the key contributing factors may not be comprehensive for these countries and their relationships may not be simple and universal. For example, the rebound risk prediction for November 2021 was incorrect for France, likely due to an excess positive contribution of foreign flights in the model, which increased the likelihood of a rebound (detailed in Table S7). This result may have been influenced by more specific travel policies and the implementation of screening at the airport. Risk misidentification was most common in South American countries in the post-vaccine era (e.g., Argentina, Bolivia and Chile), which may be related to the unique circulation pattern of COVID-19 in these countries, in which Gamma and Lambda variants were dominant at a time when the Delta variant was dominant in other countries.

There are several limitations to our study. First, although the representation of the countries included in the analysis was considered, the limited number of countries included in the study due to data quality requirements may have affected the generalization of some of our conclusions. Data from more countries should be collected and analyzed in future studies. Second, multiple machine learning algorithms coupled with an interpretable framework were used to explore the complex interactions between various influencing factors, to try to improve the reliability and robustness of the optimized models. Although machine learning models are well developed, show advantages in handling non-linear problems and are broadly used in other fields, they are less frequently used in the field of policy evaluation. In the future, more comprehensive investigations are needed, coupled with studies based on mechanistic models. A more general framework for the effective evaluation and optimization of control policies is also needed and should be the next priority. Finally, the daily basic reproductive number was the only indicator of the spread of COVID-19 used in this study. This metric captures only one aspect of the disease. There may be more relevant disease burden indices for COVID-19, such as death or hospitalization rates. These outcomes and their influencing factors should be comparatively explored in future studies. The framework proposed in this study was helpful in quantitatively understanding the contributions of various factors to the spread of COVID-19. The patterns observed in this study provide insights to facilitate the development of improved prevention and control strategies for the ongoing COVID-19 pandemic and for future pandemics.

5 | CONCLUSION

Factors that affect the spread of COVID-19 in 39 countries were systematically evaluated and prioritized using an interpretable machine learning framework. Policy- and travel-related factors were found to be the main drivers, with policy-related factors being more dominant (more than 60% of the overall contribution). Travel-related factors played an important role in the earlier stage, whereas policy-related interventions were dominant contributors at the later stage, especially in countries that experienced a rebound. Care should be taken when deciding to relax non-pharmaceutical interventions, even after...
the implementation of a vaccination program, and new mutants with greater infectivity may worsen the situation and cause recurrent outbreaks. A reliable prediction model was built based on the findings of the study, and this model may be used to evaluate potential control strategies. A quantitative understanding of the combinatorial effect of various control measures is needed for precisely and effectively controlling the spread of COVID-19.

ACKNOWLEDGEMENTS
The authors would like to thank the many thousands of Centers for Disease Control and Prevention staff, health workers and data scientists who continuously collect and publicly share data and are dedicated to containing the spread of COVID-19.

CONFLICT OF INTEREST
All authors declare no competing interests.

AUTHOR CONTRIBUTIONS
Xiangjun Du and Zicheng Cao designed the study. Zicheng Cao, Zekai Qiu, Feng Tang and Shiwen Liang collected and analyzed the data. Xiangjun Du, Shenglan Xiao, Dechao Tian and Guozhi Jiang interpreted the data. Xiangjun Du, Zicheng Cao and Zekai Qiu prepared the manuscript. Xiangjun Du, Zicheng Cao, Shenglan Xiao, Dechao Tian and Guozhi Jiang edited the paper. All authors reviewed and approved the submitted manuscript.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available at the Oxford Covid-19 Government Response Tracker (https://github.com/OxCGRT/covid-policy-tracker/) and Our World in Data (https://ourworldindata.org/).

ORCID
Xiangjun Du https://orcid.org/0000-0001-8184-8430

REFERENCES
Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). Coronavirus pandemic (COVID-19). Our World in Data. https://ourworldindata.org/coronavirus
Looi, M. K. (2020). Covid-19: Is a second wave hitting Europe? Bmj, 371, m4113. https://doi.org/10.1136/bmj.m4113
Cacciapaglia, G., Cot, C., & Sannino, F. (2020). Second wave COVID-19 pandemics in Europe: A temporal playbook. Science Reports, 10(1), 15514. https://doi.org/10.1038/s41598-020-72611-5
Wise, J. (2020). Covid-19: Risk of second wave is very real, say researchers. BMJ, 369, m2294. https://doi.org/10.1136/bmj.m2294
Xu, S., & Li, Y. (2020). Beware of the second wave of COVID-19. Lancet, 395(10233), 1321–1322. https://doi.org/10.1016/S0140-6736(20)30845-x
Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., Tan, W. & China Novel Coronavirus Investigating and Research Team. (2020). A novel coronavirus from patients with pneumonia in China, 2019. New England Journal of Medicine, 382(8), 727–733.
Awadassheid, A., Wu, Y., Tanaka, Y., & Zhang, W. (2021). Current advances in the development of SARS-CoV-2 vaccines. International Journal of Biological Sciences, 17(1), 8–19. https://doi.org/10.7150/ijbs.52569
Sonabend, R., Whittles, L. K., Imai, N., Perez-Guzman, P. N., Knock, E. S., Rawson, T., Gaythorpe, K. A. M., Djaafara, B. A., Hinsley, W., FitzJohn, R. G., Lees, J. A., Kanapram, D. T., Volz, E. M., Ghani, A. C., Ferguson, N. M., Baguelin, M., & Cori, A. (2021). Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in England: A mathematical modelling study. Lancet, 398(10313), 1825–1835. https://doi.org/10.1016/S0140-6736(21)02276-5
Mohapatra, R. K., Tiwari, R., Sarangi, A. K., Sharma, S. K., Khandia, R., Saikumar, G., & Dhama, K. (2022). Twin combination of Omicron and Delta variant triggering a tsunami wave of ever high surges in COVID-19 cases: A challenging global threat with a special focus on Indian sub-continent. Journal of Medical Virology, https://doi.org/10.1002/jmv.27585
Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P. Y., & Coizza, V. (2020). Impact of lockdown on COVID-19 epidemic in Ile-de-France and possible exit strategies. Bmc Medicine [Electronic Resource], 18(1), 240. https://doi.org/10.1186/s12916-020-01698-4
Ferrante, L., Steinmetz, W. A., Almeida, A. C. L., Leão, J., Vassão, R. C., Tupinambás, U., Fearnside, P. M., & Duczmal, L. H. (2020). Brazil’s policies condemn Amazonia to a second wave of COVID-19. Nature Medicine, 26(9), 1315. https://doi.org/10.1038/s41591-020-1026-x
Leung, K., Wu, J. T., Liu, D., & Leung, G. M. (2020). First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: A modelling impact assessment. Lancet, 395(10233), 1382–1393. https://doi.org/10.1016/s0140-6736(20)30746-7
Panovska-Griffiths, J., Kerr, C. C., Stuart, R. M., Mistry, D., Klein, D. J., Viner, R. M., & Bonell, C. (2020). Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: A modelling study. Lancet Child Adolesc Health, 4(11), 817–827. https://doi.org/10.1016/s2352-4642(20)30250-9
Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. Journal of Population Economics, 33, 1127–1172. https://doi.org/10.1007/s10014-020-00778-2
Sebhateh, A., Wenneberg, K., Arora-Jonsson, S., & Lindberg, S. I. (2020). Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. PNAS, 117(35), 21201–21208. https://doi.org/10.1073/pnas.2010625117
Kandel, N., Chungong, S., Omaar, A., & Xing, J. (2020). Health security capacities in the context of COVID-19 outbreak: An analysis of International Health Regulations annual report data from 182 countries. Lancet, 395(10229), 1047–1053. https://doi.org/10.1016/s0140-6736(20)30553-5
Brauer, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., Norman, A. J., Monrad, J. T., Besiroglu, T., Ge, H., Hartwick, M. A., Teh, Y. W., Chindelevitch, L., Gal, Y., & Kulveit, J. (2020). Inferring the effectiveness of government interventions against COVID-19. Science, 371(6531), eabd9338. https://doi.org/10.1126/science.abd9338
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., Perez-Guzman, P. N., Schmit, N., Cilloni, L., Ainslie, K. E. C., Baguelin, M., Boonyasiri, A., Boyd, O., Cattarino, L., ... Cattarino, L. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature, 584(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7
Haug, N., Geyerrother, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loero, V., Pinior, B., Thurner, S., & Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. Nature Human Behaviour, 4(12), 1303–1312. https://doi.org/10.1038/s41562-020-01009-0
Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L. Y., Hultgren, A., Krasovich, E., Lau, P., Lee, J., Rolf, E., Tseng, J., & Wu, T. (2020). The effect of large-scale anti-contagion...
outbreak in Chinese provinces: Machine learning approach using novel digital data and estimates from mechanistic models. *Journal of Medical Internet Research [Electronic Resource]*, 22(8), e20285. https://doi.org/10.2196/20285

Middelburg, R. A., & Rosendaal, F. R. (2020). COVID-19: How to make between-country comparisons. *International Journal of Infectious Diseases*, 96, 477–481. https://doi.org/10.1016/j.ijid.2020.05.066

Wells, C. R., Sah, P., Moghadas, S. M., Pandey, A., Shoukat, A., Wang, Y., Wang, Z., Meyers, L. A., Singer, B. H., & Galvani, A. P. (2020). Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *PNAS*, 117(13), 7504–7509. https://doi.org/10.1073/pnas.2002616117

Wong, M. C., Huang, J., Teoh, J., & Wong, S. H. (2020). Evaluation on different non-pharmaceutical interventions during COVID-19 pandemic: An analysis of 139 countries. *Journal of Infection*, 81(3), e70–e71. https://doi.org/10.1016/j.jinf.2020.06.044

Malki, Z., Atlam, E. S., Hassanien, A. E., Dagnew, G., Elhosseini, M. A., & Gad, I. (2020). Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos, Solitons & Fractals*, 138, 110137. https://doi.org/10.1016/j.chaos.2020.110137

Yang, J., Li, J., Lai, S., Ruktanonchai, C. W., Xing, W., Carioli, A., Wang, P., Ruktanonchai, N. W., Li, R., Floyd, J. R., Wang, L., Bi, Y., Shi, W., & Tatem, A. J. (2020). Uncovering two phases of early intercontinental COVID-19 transmission dynamics. *Journal of Travel Medicine*, 27(8), taaa200. https://doi.org/10.1093/jtm/taaa200

The, L. (2020). COVID-19 in India: The dangers of false optimism. *Lancet*, 396(10255), 867. https://doi.org/10.1016/s0140-6736(20)32001-8

Coccia, M. (2020). Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Science of the Total Environment*, 729, 138474. https://doi.org/10.1016/j.scitotenv.2020.138474

Sarkodie, S. A., & Owusu, P. A. (2020). Impact of meteorological factors on COVID-19 pandemic: Evidence from top 20 countries with confirmed cases. *Environmental Research*, 191, 110101. https://doi.org/10.1016/j.envres.2020.110101

Ferrante, L., Duczmal, L., Steinmetz, W. A., Almeida, A. C. L., Leão, J., Vassão, R. C., Tupinambás, U., & Fearnside, P. M. (2021). How Brazil’s President turned the country into a global epicenter of COVID-19. *Journal of Public Health Policy*, 42(3), 439–451. https://doi.org/10.1057/s41271-021-00302-0

Greer, S. L., King, E. J., da Fonseca, E. M., & Peralta-Santos, A. (2020). The comparative politics of COVID-19: The need to understand government responses. *Global Public Health*, 15(9), 1413–1416. https://doi.org/10.1080/17441692.2020.1783340

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

**How to cite this article:** Cao, Z., Qiu, Z., Tang, F., Liang, S., Wang, Y., Long, H., Chen, C., Zhang, B., Zhang, C., Wang, Y., Tang, K., Tang, J., Chen, J., Yang, C., Xu, Y., Yang, X., Xiao, S., Tian, D., Jiang, G., & Du, X. (2022). Drivers and forecasts of multiple waves of the coronavirus disease 2019 pandemic: A systematic analysis based on an interpretable machine learning framework. *Transboundary and Emerging Diseases*, 1–11. https://doi.org/10.1111/tbed.14492