Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Exploring COVID-19 transmission patterns and key factors during epidemics caused by three major strains in Asia

Han Tu, Xia Wang*, Sanyi Tang

School of Mathematics and Statistics, Shaanxi Normal University, Xi’an 710119, PR China

ARTICLE INFO

Keywords:
COVID-19
Hierarchical clustering
K-shape clustering
Vaccination
Preventive and control measures

ABSTRACT

The COVID-19 epidemic has lasted for more than two years since the outbreak in late 2019. An urgent and challenging question is how to systematically evaluate epidemic developments in different countries, during different periods, and to determine which measures that could be implemented are key for successful epidemic prevention. In this study, SBD distance-based K-shape clustering and hierarchical clustering methods were used to analyse epidemics in Asian countries. For the hierarchical clustering, epidemic time series were divided into three periods (epidemics induced by the Original/Alpha, Delta and Omicron variants separately). Standard deviations, the Hurst index, mortality rates, peak value of confirmed cases per capita, average growth rates, and the control efficiency of each period were used to characterize the epidemics. In addition, the total numbers of cases in the different countries were analysed by correlation and regression in relation to 15 variables that could have impacts on COVID-19. Finally, some suggestions on prevention and control measures for each category of country are given. We found that the total numbers of cases per million of a population, total deaths per million and mortality rates were highly correlated with the proportion of people aged over 65 years, the prevalence of multiple diseases, and the national GDP. We also found significant associations between case numbers and vaccination rates, health expenditures, and stringency of control measures. Vaccinations have played a positive role in COVID-19, with a gradual decline in mortality rates in later periods, and are still playing protective roles against the Delta and Omicron strains. The stringency of control measures taken by a government is not an indicator of the appropriateness of a country’s response to the outbreak, and a higher index does not necessarily mean more effective measures; a combination of factors such as national vaccination rates, the country’s economic foundation and the availability of medical equipment is also needed. This manuscript was submitted as part of a theme issue on “Modelling COVID-19 and Preparedness for Future Pandemics”.

1. Introduction

The COVID-19 epidemic has lasted for more than two years since the outbreak in late 2019. At the beginning of the outbreak, Asian countries treated the epidemic significantly more seriously than European and American countries, and most of the Asia countries took timely, active, and effective preventive measures so that the epidemics were effectively controlled (Yi et al., 2020). However, new mutant strains emerged and there are significant differences in the transmissibility and virulence of different strains (Flaxman et al., 2020; Hart et al., 2022; Nyberg et al., 2022). Many Asian countries, such as India and Turkey, were hit by the Delta and Omicron variants leading to new epidemic waves. Meanwhile, because the high intensity of long-term prevention and control measures have had major impacts on societies and economies, and because the pathogenicity of some new mutant strains has been reduced compared to the Original strain (Bakhshandeh et al., 2021; Wu et al., 2022), some countries began to treat COVID-19 as just an ordinary influenza-like disease with negligible effects on health they gradually abandoned the implementation of strict control measures (Chen, n.d.). In addition, from the end of 2020, various countries have successively promoted vaccination plans, which have also had a substantial impact on the development of the pandemic.

In addition, Asia, as the world’s largest continent, has a geographical area of more than 44.5 million square kilometers and a population of more than 4.5 billion, accounting for nearly 60 % of the total global population. Due to Asia’s vast territory and large population, there are huge social, economic and cultural differences between countries. For example, Japan, Korea, Singapore, and Israel are developed countries,
while most of the remaining countries, including China, India, Turkey, and Iran, are developing countries. Singapore is the country with the largest population density, reaching 7915.731 people per km², while Mongolia is the country with the lowest population density, with a density of 1.98 people per km². Differences in population structures, socio-economic conditions, medical conditions, control measures and so on, may also account for differing trends in how COVID-19 epidemics develop in different countries. Therefore, the development of COVID-19 in Asia is very complex, and we urgently need to understand the similarities and differences in the development patterns of COVID-19 in Asian countries and the key factors that influence the different characteristics, so that we can target the implementation of epidemic prevention measures by region (Xie et al., 2021).

Cluster analysis is a useful tool to improve understanding of the similarities and differences amongst epidemic development patterns of, which has already been applied to the study of COVID-19 (Tang et al., 2020; Wu et al., 2020; Roques et al., 2020; Sun et al., 2020; Yang and Wang, 2021; Zhu and Li, 2020; He et al., 2020; James and Menzies, 2020). Yi et al. used hierarchical clustering to cluster 34 provincial administrative regions of China according to three scales: number of cases, number of deaths, and number of recoveries (Yi et al., 2020). A series of studies by Zarikas et al. introduced a clustering algorithm specifically for clustering countries according to the concept of stratified analysis based on active COVID-19 cases, but only clustered countries according to the number of new cases per population (Zarikas et al., 2020). Li et al. clustered countries based on prevalence in provincial administrative regions of China, using elbow cluster analysis and performed a descriptive analysis of the number of infectious regenerations, R₀, in areas within the resulting cluster classes (Li et al., 2020). All of these studies have only clustered countries with a single variable (Yi et al., 2020; James and Menzies, 2020; Zarikas et al., 2020; Li et al., 2020), while there is also currently little literature on clustering based on the shape of epidemic time series.

In the analysis of influencing factors, Siddiqui et al. explored the relationship between confirmed, suspected and fatal cases, and temperature profiles. They used the K-means method to cluster different regions of China, and analyzed the effect of temperature on the outbreak (Siddiqui et al., 2020). Carrillo et al. analyzed country-level variables based on numbers of COVID-19 confirmed cases and developed models defining clusters in 155 countries, showing that models with five or six clusters could stratify countries according to the number of confirmed cases, but they only compared these clusters in terms of four dimensions: number of confirmed cases, number of deaths, case fatality rate and the order in which countries reported their first cases (Carrillo-Larco et al., 2020). Farseev et al. studied economic and health factors influencing the spread of COVID-19 disease, resulting in four national clusters, but only clustered and correlated the influencing factors without considering the overall trend in the development of the epidemic (Farseev et al., 2020). Xie et al. used the spectral clustering method to cluster individual countries based on their basic statistical characteristics, time-domain characteristics, and epidemic-specific characteristics to reveal the patterns and characteristics that exist in each country’s epidemic sequence. Three relatively important factors were selected for analysis in their causal analysis section (Xie et al., 2021). Rizvi et al. clustered the countries using social, economic, health, and environment-related indicators that influence the spread of the disease, but that study only evaluated confirmed cases and deaths as parameters to study factors that are closely related to the spread of the disease (Rizvi et al., 2021). How to systematically evaluate the epidemic situation in different countries, comprehensively consider the impact of society, economics, population structure, virus variation and policy adjustment, and identify the key influencing factors is still an urgent and challenging problem.

In order to solve the above problems, we clustered Asian countries according to characteristic variables and the shape of epidemic time series. In order to explore the influence of virus variation on epidemics, those induced by the Original/Alpha, Delta and Omicron strains were studied separately. When considering the impact of social, economic, demographic and other factors on the COVID-19 epidemic, two factors, vaccination rate and stringency index, were added to help in-depth understanding of the virus’s transmission characteristics, judge epidemic trends, identify the similarities and differences amongst epidemic development models in different countries, identify the leading factors behind different characteristics, and provide theoretical support for taking active and effective epidemic prevention measures.

2. Methods

2.1. Data sources

This paper focuses on 45 countries in Asia that have experienced COVID-19 epidemics (excluding North Korea, Turkmenistan, and Northern Cyprus). We collected data on the numbers of COVID-19 confirmed cases and deaths from 23 January 2020 to 15 April 2022, on the Our World in Data website (WHO, n.d), we attempted to overcome the inconsistency of data sampling by using a combination of data sources, and the data are assumed to be real and reliable.

To better explain the clustering results, variables that may have an impact on COVID-19 were also collected from the website Our World in Data (WHO, n.d) and from the World Bank (World Bank Open Data, n.d.).

These variables were selected because they are closely related to the COVID-19 pandemic from a clinical and public health perspective. Diabetes is common in patients with COVID-19, tuberculosis shares a common pathological structure with acute respiratory diseases like COVID-19, and the likelihood of an impaired immune response due to HIV/AIDS is high (Carrillo-Larco et al., 2020). So, one chronic non-communicable disease (diabetes) and two communicable diseases (tuberculosis and HIV/AIDS) were selected. Indicators such as public health facilities and health expenditures were also selected, which may affect the probability that a person will take preventive measures and acquire appropriate treatment when necessary. Besides, the vaccination rates and the stringency of control measures may also have important impacts on the development of the epidemic. Therefore, data on 15 variables were collected, including population density (x₁), the proportion of people over 65 (x₂), life expectancy (x₃), percentage of female cigarette smokers (x₄), percentage of male cigarette smokers (x₅), GDP (x₆), extreme poverty index (x₇), hospital beds per thousand (x₈), public health facilities (x₉), health expenditure (x₁₀), diabetes prevalence (x₁₁), HIV prevalence (x₁₂), tuberculosis prevalence (x₁₃); stringency index (x₁₄) and vaccination rates (x₁₅) (See Table S1–S3 in Appendix S1 for details).

2.2. Research methodology

Due to differing socio-economic conditions and cultures, countries’ authorities do not all pay the same amount of attention to their epidemic situations, so may implement different control measures. Besides, the epidemic developments are affected by many other factors, including medical conditions, population structure, economic development level, underlying diseases, control measures and so on. Therefore, in order to analyze the epidemic development trend of Asian countries more clearly and the potential causes affecting epidemics in various countries, we analysed shape-based epidemic time series and feature variable-based methods for cluster analysis.

In order to consider whether the epidemic curves of Asian countries have similar shapes, similar to the data used in the Characterizing the Spatial Heterogeneity of the COVID-19 Pandemic through Shape Analysis of Epidemic Curves (Srivastava and Chowell, 2021), a time series of newly confirmed cases reflecting the epidemiological dynamics of each country was chosen. We adopted the k-shape clustering algorithm based on shaped-based distance (SBD) to conduct shape-based time series cluster analysis. The appropriate number of Asian countries to be classified is determined by calculating the cluster validity index. After that,
clustering is performed and the shapes curves of different categories are discussed.

When studying the composition of characteristic variables of epidemics in Asian countries, and taking account of the different strains involved, we separated the data by taking the beginning as 23 January 2020, and the cut-off at 1 July 2021 and 1 January 2022, we selected the data from three periods for analysis. Six features which can show the similarity and global structure of the time series, including the standard deviation, Hurst index, mortality rate, peak value of confirmed cases per capita, average growth rate and control efficiency, were chosen for hierarchical cluster analysis.

When interpreting the classified epidemic trends in Asian countries, the correlation coefficients of 15 variables that may have impacts on COVID-19, such as the proportion of people over 65, health expenditures, tuberculosis prevalence, stringency index of responses to the epidemic, and vaccination rates, were calculated for each country. We also performed a regression analysis to investigate the significance of the variables, and finally the results obtained from these three components were combined for analysis and discussion.

2.2.1. Shape clustering based on K-shape partitioning clustering algorithm

(1) Shape-Based Distance

When studying time-series data, it is relevant sometimes to consider the overall trend of the time-series curves, and the Shape-Based Distance (SBD) metric is a way of determining whether the curve patterns are similar. SBD can ignore factors such as time series amplitude and displacement and it can also eliminate the effects of amplitude and displacement on this metric by normalizing and moving the series, respectively.

(2) K-shape clustering algorithm

K-shape, as a new time series clustering algorithm, is similar to the K-means algorithm (Paparrizos and Gravano, 2016) and is based on a measurable iterative optimization process, but K-shape uses a different distance measure and cluster center calculation method compared with K-means. K-shape tries to maintain the shape when comparing time series.

(3) Clustering validity index

How to judge the goodness of clustering results is also an important part. The results of clustering follow a principle: the intra-class similarity is as large as possible, while the inter-class similarity is as small as possible (Jin, 2014), so many evaluations of clustering algorithms are based on this principle. In order to evaluate the merits of clustering algorithms, 5 Cluster Validity Indexes (CVI), including Davies-Bouldin index (DB), Davies-Bouldin+ index (DB+), Silhouette index (SIL), COP index (COP), and Calinski-Harabasz index (CH) (Liu, 2017; Arbelaitz et al., 2013), are used to measure the effect of clustering (See Appendix S2 for details).

2.2.2. Hierarchical clustering based on feature variables

Since the outbreak of the COVID-19 pandemic in late 2019, the SARS-CoV-2 virus has mutated many times with significant affects on the epidemiology of the disease. The Delta variant was first discovered in India in October 2020, and as time progressed, the Delta strain continued to mutate and it emerged in many regions around the world. The Delta strain is characterized by a short incubation period, rapid transmission, high viral load, and is more likely to lead to critical illness (Wang and Zhao, 2021). On 2 July 2021, WHO held a press conference to state that the Delta variant of COVID-19 was becoming the cause of a major epidemic in many countries (Ma, 2021). On 26 November 2021, the WHO defined Omicron as the fifth “strain of concern”, stating that the overall global risk assessment of the Omicron strain was “very high” and that it was likely to spread widely around the world. As of 4 January 2022, 128 countries and territories had reported the presence of the Omicron strain. For Asia, the proportion data of virus strains in some countries can be found on the website of our world in data (as shown in Fig. S5). It shows that in July 2021, the major strains of 11 out of 15 countries (73 %) were Delta strains, while in June 2021, the major strains of only 3 out of 10 countries (30 %) were Delta strains. In January 2022, the major strains of 19 out of 20 countries (95 %) were Omicron strains, while in December 2021, the major strains of only 1 out of 18 countries (6 %) were Omicron strains. So, to illustrate the difference between the Delta, Omicron and previous strains more clearly, in this paper, the major epidemic strain of COVID-19 during 23 January 2020 and 30 June 2021 was designated as the Original/Alpha strain, the period from 1 July 2021 to 31 December 2021 is defined as Delta strain dominant, and the period from 1 January 2022 to 15 April 2022 is defined as Omicron strain dominant. Finally clustering was performed separately based on the feature variables during the three periods of study.

(1) Selection of feature variables

This paper studies the development trend of the epidemic, so structural features describing the global organization and internal variation of the time series were selected for clustering. Among them, standard deviations and the Hurst index are usually used to show the similarity of change and global structure of an epidemic time series (Xie et al., 2021). Furthermore, as explained below, the mortality rate, peak value of confirmed cases per capita, average growth rate, and prevention and control efficiency of measures used against the epidemic can characterize the time series development of the COVID-19 epidemic in Asia.

- The standard deviation ($\sigma$) characterizes the time series and can reveal the dynamics of the numbers of daily confirmed cases over time in Asian countries, calculated as:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$

Where $x_i$ is the number of daily confirmed cases of COVID-19 on day $i$, $\bar{x}$ is the mean of the numbers of daily confirmed cases, and $n$ is the length of the time series of the numbers of daily confirmed cases.

- The Hurst index ($H$) describes the pattern of time series over time and can visually describe the relationship between the development trend after the epidemic has begun and the current trend in Asian countries. The larger the $H$ value, the longer that a time series trend will last. The Hurst index (Zhang et al., 1997) is calculated in this section using the repolarization method with the following formula:

$$\log\left(\frac{R}{S}\right) = \log(K) + H\log(t).$$

Where $R$ is the extreme difference of each time series segment, $S$ is the standard deviation of each time series segment, $t$ is the length of the time series segment, and $K$ is the intercept. It runs from the beginning of each VOC.

- The peak value of confirmed cases per capita ($p$) gives an indication of the severity and the level of control of the outbreak in the country. According to the previous studies, population density is highly correlated with COVID-19, and there is a large difference in population size between Asian countries. Considering only the peak of confirmed cases will ignore the influence of population and other factors, and peak value of confirmed cases per capita has a greater impact on the adequacy of medical resources (Ferguson et al., 2020), so in order to remove the influence of demographic factors (Qi et al., 2020), we calculated it as...
P = \frac{\max(x_i)}{N} \quad (3)

Where \( N \) is the total population of the country, the population size is taken from the World Bank (World Bank).

(2) The mortality rate (\( y_1 \)) is a core indicator of the development of the epidemic, it is very closely related to the severity of the epidemic and the level of medical care in the country, etc. It is calculated by the formula.

\[ y_3 = \frac{d}{c} \times 100\%. \quad (4) \]

Where \( d \) is the cumulative number of deaths, and \( c \) is the cumulative number of confirmed cases during the study period.

(3) The average growth rate (\( \overline{G} \)) describes the average growth rate from the initial value to the peak, which can indicate the speed of the spread of the epidemic and whether the national preventive and control measures are timely and effective. It is calculated as.

\[ \overline{G} = \sqrt{\frac{\max(x_i)}{x_0}} \quad (5) \]

Where \( x_0 \) is the value of the first number of daily confirmed cases in the country that is not zero, and \( T \) is the length of time between the initial number of daily confirmed cases and the maximum number of confirmed cases during the study period.

(4) Control efficiency (\( \varepsilon \)) describes the degree of reduction of the number of daily confirmed cases in a single day and can indicate the overall effectiveness of epidemic prevention and control in Asian countries. It is calculated as.

\[ \varepsilon = \frac{\max(x_i)}{x_{\text{ave}}} \times 100\%. \quad (6) \]

where \( \varepsilon \) is the prevention and control efficiency and \( x_{\text{ave}} \) is the average of the last three days of the number of daily confirmed cases.

These six time series feature indicators for each Asian country were calculated and the data set was then standardized to remove the effect of the magnitude.

(2) Hierarchical clustering method

Hierarchical clustering is a method that is often used for classification of small datasets (Sun et al., 2008), which is implemented in the following way: each object forms a class by itself at the beginning, and then the two most similar classes are merged each time, the Euclidean distance between the new class and the other classes is recalculated after the merger, and this process continues until all objects are grouped into one class. The process of merging classes can be described by a spectral clustering diagram.

Hierarchical clustering often has the advantages of an easy definition of similarity, the number of clusters does not need to be defined in advance and the hierarchical relationship between classes is relatively intuitive, so this paper adopts the analysis method of hierarchical clustering. The Euclidean distance between samples is chosen as the distance measure, and the sum of squared deviations between classes is chosen as the distance between classes. Then the correlations between 15 variables, including the proportion of people over 65, health expenditures and tuberculosis prevalence in Asian countries, and the total numbers of cases per million of population and the total numbers of deaths per million were calculated for the explanations of the classified epidemic trends in Asian countries, and the Spearman rank correlation coefficient was chosen as the metric.

In this paper, correlation analysis studies the closeness between two variables, but it is not possible to obtain a specific relationship between them. Therefore, next we introduce multiple regression, which not only determines the regression model between the dependent variable and several independent variables, giving the regression equation, but also evaluates the contribution of the independent variables to the dependent variable. Finally, we can use the obtained regression equation, given the value of the independent variables, to predict and control the dependent variable. The variables were first selected to avoid colinearity before conducting the regression analysis, and then the filtered data were subjected to multiple linear regression analysis, and the regression results obtained are shown in Tables S11–S13 in the Appendix S1. For linear regression models with insignificant results and low explanatory degree, generalized additive models (GAM) were introduced. The GAM provides a preferred solution to study the relationship between multiple independent variables and a dependent variable, while not requiring prior knowledge of the relationship between the dependent and independent variables. And as a flexible statistical model, it can be used to detect the effects of nonlinear regression and to fit the model using a nonlinear smoothing term. Besides, the GAM is often used to assess the interactions between relevant factors and diseases, as well as for predictive early warning of diseases (Tao et al., 2012; Wang et al., 2017; Hu, 2020, Chodick et al., 2021).

And similarly, generalized linear regression was first conducted on the variables after eliminating co-linearity, and the independent variables with a significant linear relationship with the dependent variable in the results continued to be included in the GAM model as linear parameter terms, and the insignificant independent variables were fitted using a nonlinear smoothing term. And the final model was selected based on the higher value of R², a smaller Akaike’s Information Criterion (AIC) and generalized cross validation (GCV).

3. Results

3.1. Epidemic time series shape clustering results

To confirm how many categories should be classified, the 5 CVIs shown in the methods section were calculated (Table 1). According to the principle that DB, DB*, and COP are as small as possible, and that the SIL and CH are as large as possible, where the bolded font in Table 1 represents the best choice, we found that 3 indicators support the classification of Asian countries into 4 categories, determining the number of clusters as 4. Next, the time series of the daily confirmed cases, the shape-centered curve of each category and the clustering maps of Asian countries were obtained (Figs. 1 and 2(a)). As shown in Fig. 1(b), the numbers of daily confirmed cases in countries of the first category show a decreasing trend, reach the lowest valley, fluctuate continuously up to the peak, and then are rapidly controlled, but later the numbers of confirmed cases show a rapidly increasing trend again, such as in Yemen, which has the weakest economic base in Asia. The daily number of confirmed cases of the second category initially increase, gradually decline after reaching a peak and flatten out later, such as in Thailand, which had a high proportion of people over 65 but improved vaccination rates in later stages. The numbers of daily confirmed cases of the third category were relatively stable in the early stage, and have larger outbreaks in the late stages, such as in China, which had an excellent control efficiency in the early stage, but in the later stage the control efficiency was below the average value. The numbers of daily confirmed cases of the fourth category developed smoothly in the early stage, but there was a large outbreak in the middle, and after the epidemic was controlled there was a small increase in the mid to late stage. The number of confirmed cases kept decreasing and

| Table 1: Clustering algorithm evaluation table. |
|-----------------------------------------------|
| SIL(max) | CH(max) | COP(min) | DB(min) | DB*(min) |
|-----------|---------|---------|--------|---------|
| Kshape1(3) | 0.3750 | 21.9555 | 0.2854 | 0.8506 | 0.8997 |
| Kshape2(4) | 0.3882 | 23.0576 | 0.2290 | 1.0650 | 1.2078 |
| Kshape2(5) | 0.2890 | 18.0090 | 0.2311 | 1.0298 | 1.4275 |
Fig. 1. K-shape division clustering and shape-centered result map. Panel (a) is the result of dividing Asian countries into four categories by shape, and Panel (b) is the form-centered curve of the countries in Panel (a).

Fig. 2. Clustering map, where Panel (a) is a map of Asia divided into four categories according to the shape of the epidemic; Panel (b) is a map of Asia divided into two categories according to the feature variables of the Original/Alpha strain epidemic period; Panel (c) is a map of Asia divided into three categories according to the feature variables of the Delta strain epidemic period; Panel (d) is a map of Asia divided into four categories according to the feature variables of the Omicron strain period.
stabilized, such as in Japan, which had the highest proportion of people over 65 in Asia, but with a relatively high GDP.

3.2. Clustering results of epidemic feature variables

Based on the six time series feature indicators during the epidemics induced by the Original/Alpha, Delta and Omicron strains, hierarchical clustering analysis was performed respectively, and the clustering results are shown in Fig. S1. During the Original/Alpha strain epidemic period, it follows from Fig. S1(a) that Brunei, China, India, and Yemen fall into one category each, which is because the feature values of these four countries are very different from the rest. For instance, Brunei has a much higher average growth rate than other countries, China has the best control efficiency, India has a very huge standard deviation in time series of the numbers of confirmed cases and Yemen has the highest mortality rate. To better analyze the results, the four countries with large bias were removed, then specific analysis was performed only for the remaining two categories (A1 and A2). Similarly, during the Delta strain epidemic period, Asian countries are classified into three categories (D1, D2 and D3) after deleting those countries with a large deviation (shown in Fig. S1(b)), and during the Omicron strain epidemic, Asian countries were classified into four categories (O1, O2, O3 and O4, shown in Fig. S1(c)) after removing countries with large deviations based on six time-series characteristic indicators, the specific sources of bias of the feature variables are shown in Table S16.

Based on the results of clustering, the numbers of daily confirmed cases in Asian countries during the three major strain epidemic periods were plotted by category (Fig. S2). The six indicators were then averaged for Asian countries belonging to different categories in all three phases separately (Fig. 3).

As shown in Fig. S2(a and b) and Fig. 3(a), during the Original/Alpha strain epidemic period, the average standard deviation, average growth rate, Hurst index and mortality rate of A2 are higher than those of A1. This indicates that compared with A1 countries, the time series of numbers of daily confirmed cases of countries in A2 fluctuate more violently, grow faster and have stronger continuity. Besides, the peak value of confirmed cases per capita and the control efficiency of countries in A1 are larger than those of countries in A2. This indicates that although the peak value of confirmed cases per capita in type A1 countries is higher than that in type A2 countries, the control effect is good and the mortality rate is low.

As shown in Fig. S2(c–e) and Fig. 3(b), during epidemics induced by the Delta strain, type D1 countries have the highest Hurst index and average growth rate, and the lowest peak value of confirmed cases per capita, indicating that the overall time series of type D1 countries is characterized by a rapid rise and strong continuity, and that the development trend is relatively stable in general. Type D2 countries have the lowest standard deviation, Hurst index, mortality rate, average growth rate, and control efficiency, the highest peak value of confirmed cases per capita. This suggests that the time series of type D2 countries are not sufficiently continuous and that the epidemic is not sufficiently controlled, with a tendency to continue rising at later stage. The standard deviation, mortality rate and control efficiency of type D3 countries are the highest, and the Hurst index is lower than that of type D2 countries, indicating that the overall development of type D3 countries is not very stable, and the time series has a tendency to increase significantly in the later part of the period, and the continuity of the time series is not high.

As shown in Fig. S2(f–i) and Fig. 3(c), during epidemics induced by the Omicron strain, type O1 countries have the highest control efficiency and Hurst index, indicating that type O1 countries have a higher continuity of the time series of confirmed cases and are the most effective in controlling the epidemic. The type O2 countries have the highest mortality rate and the lowest standard deviation, average growth rate, control efficiency, and Hurst index, indicating that the overall development of the type O2 countries is relatively stable and the epidemic is increasing slowly, but the control of the epidemic is inadequate, with high mortality rate, low control efficiency, and poor time series continuity. Type O3 countries have the lowest peak value of confirmed cases per capita and the highest average growth rate, indicating that type O3 countries have the weakest outbreaks compared with the other three types of countries, although the epidemic is rising more rapidly. Type O4 countries have the highest standard deviation and highest peak value of confirmed cases per capita, indicating that type O4 countries have the most unstable and fluctuating epidemic development and the highest outbreak size but the lowest mortality rate.

To explore the reasons for the differences among different types of Asian countries during these periods, we calculated the mean of 15 significant variables which may be correlated with the development of epidemics for each type of country (see Tables S1–S3 in Appendix S1 for details).

During epidemics induced by Omicron strain, type O1 countries have lower mortality rates, slightly higher than those of type O4 countries, and the highest population density, GDP and diabetes prevalence, but the lowest number of hospital beds per thousand; type O2 countries have the highest mortality rate, and the lowest population density, the proportion of people over 65, life expectancy, GDP, public health facilities, health expenditures, HIV prevalence and stringency of control measures, and the highest percentage of smokers; type O3 countries have the lowest total cases per million, but higher mortality rate, second only to type O2 countries, and the highest public health facilities and health expenditures, but also the highest percentage of smokers, HIV and tuberculosis prevalence, and the lowest vaccination rates; Although type O4 countries have the highest total number of cases and deaths per million, they have the lowest mortality rate, type O4 countries have the highest proportion of people over 65, the longest life expectancy, the highest number of hospital beds per thousand, vaccination rates, and strict control measures among the four categories, and the lowest prevalence of both diabetes and tuberculosis. Comparisons of these variables for the other types of countries are shown in Appendix S3.

Fig. 3. Comparative radar plot of six indicators according to strain, where Panel (a) shows the results during the period when the Original/Alpha strain was dominant, and Panel (b) shows the results when the Delta strain predominated, and Panel (c) shows the results when the Omicron strain was predominating.
3.3. Correlation analysis

To understand how these 15 influencing factors would affect the total numbers of cases per million, total deaths per million, and the mortality rates in different categories of Asian countries during the three epidemic periods, we calculated the Spearman correlation coefficients and tested the significance of the correlation coefficients (shown in Fig. 4). The spearman correlation coefficients with absolute values greater than 0.3 are shown in Table 2 and the results with high significance ($p < 0.05$) are in bold.

Results of the Spearman correlation analysis indicate that the total numbers of cases and deaths per million were significantly positively correlated with life expectancy, GDP, hospital beds per thousand and public health facilities. The total numbers of cases per million were also negatively correlated with an extreme poverty index and health expenditures. However, the mortality rate was negatively correlated with life expectancy, GDP, prevalence of diabetes and vaccination rates, and positively correlated with health expenditures (Rizvi et al., 2021).

Specifically, it can be seen from the Spearman correlation coefficient plots for type O1 countries (see Fig. 4(c)) that there were high positive correlations between total numbers of cases per million and life expectancy (0.67), and GDP (0.59) in type O1 countries. There were high positive correlations between total numbers of deaths per million and the proportion of people over 65 (0.62), life expectancy (0.51), and hospital beds per thousand (0.52). Mortality rate was highly positively correlated with health expenditures (0.52), and HIV prevalence (0.50), and highly negatively correlated with vaccination rates (-0.59), GDP (-0.57), and diabetes prevalence (-0.53). We found a high positive correlation between health expenditures and the mortality rate, probably due to the high correlation between health expenditures and HIV prevalence (0.34). Whereas, the negative correlation between diabetes prevalence and mortality rate may be due to the high positive correlation between diabetes prevalence and vaccination rates (0.47). This group of countries, despite having the highest prevalence of diabetes and lower health expenditures, has the best control of the epidemic due to the best economic base, lower prevalence of tuberculosis, higher vaccination rates, and more adequate public health facilities, although it has the second-highest total number of cases and deaths per million among the four groups of countries, but lower mortality rates than O3 and O4 type countries.

The Spearman correlation coefficient plot for O2 countries (Fig. 4(f)) shows that there was a high negative correlation between the total numbers of cases per million and health expenditures (-0.74). The total deaths per million had a high positive correlation with the proportion of people over 65 (0.36), GDP (0.36), vaccination rates (0.35) and stringency of control measures (0.35), and negative correlations with health expenditures (-0.46), percentage of smokers (-0.40) and extreme poverty index (-0.37). The mortality rate was highly negatively correlated with vaccination rates (-0.40) and hospital beds per thousand (-0.58). We found that the total numbers of cases and deaths per million in this category were positively correlated with public health facilities and vaccination rates, similar to the D1 type countries, both of which to
some extent reflect the importance that a country attaches to the epidemic, with higher detection rates of confirmed cases and deaths as the importance increases. As can be seen from Table S3 in Appendix S1, the economic base of this type of country is weak, although the total numbers of cases and deaths per million are the lowest, this may be caused by the low economic level and the low detection rate for the epidemic, the type O2 countries did not pay enough attention to the epidemic, the public health facilities, health expenditures and the stringency of control measures are the lowest, and the vaccination rates are lower than those of O1 and O4 countries, which results in high mortality rates and poor control efficiency and continuity in this group of countries.

To analyze the causes of the rapid average growth rates and small sizes of outbreaks in O3 countries, the correlation coefficients between the total numbers of cases per million and mortality rate in O3 countries and each of the 15 important influencing variables were calculated (Fig. 4(g)). The total numbers of cases per million had a high positive correlation with GDP (0.83), life expectancy (0.72), hospital beds per thousand (0.53), and vaccination rates (0.47), as well as a high negative correlation with GDP (-0.55), health expenditures (-0.53), and percentage of smokers (-0.47). Mortality rate had a high positive correlation with the percentage of smokers (0.75), a high negative correlation with the vaccination rates (-0.75), and GDP (-0.55). The higher positive correlation between the total cases per million and GDP, hospital beds per thousand and vaccination rates may be due to the strong positive correlation of all three variables with life expectancy (0.67, 0.35, 0.38) and these three variables are closely related to the importance given to COVID-19 and may have a positive effect on the detection rate. The highest rate of epidemic growth in this group of countries may be due to the highest percentage of smokers, HIV prevalence, and tuberculosis prevalence, and the lowest vaccination rates in O3 countries. At the same time, these countries have adequate public health facilities and higher health expenditures, as well as more strict control measures and better control outcomes (only lower than type O1 countries), so that overall the size of the outbreak in type O3 countries was the smallest.

Since there are only 3 sample points in type O4 countries, the precision of correlation analysis for them is not high. In type O4 countries, the mean values of important influencing variables (Table S3 in Appendix S1) were analyzed for type O4 countries, and it was found that the level of GDP in this category is lower than that in type O1 and O3 countries, and the level of aging is severe, which may lead to the unstable development and large outbreak size in this category. Type O4 countries have the highest total number of cases and deaths per million but the lowest mortality rate, probably because type O4 countries have the lowest prevalence of both diabetes and tuberculosis, lower prevalence of HIV than type O1 and O3 countries, and the highest number of hospital beds per thousand, vaccination rates, and stringency of control measures among the four categories.

A comprehensive analysis leads to the following: For some countries with high GDP and good economic base, they tend to face the situation of serious aging and high diabetes prevalence, so this will cause higher total cases per million, and correspondingly higher total deaths per million, but these countries have good medical conditions and high vaccination rates, so the mortality rate has decreased; Vaccine has a significant role in preventing serious illness and death (Pan, 2021; Nyberg et al., 2022), and observing the table, it can be found that the Spearman correlation coefficients between the vaccination rates and mortality rates are mostly significant and negative; observing the Spearman correlation coefficients between the total numbers of cases per million and the vaccination rates, it can be seen that they are mostly positive, and there is also a negative correlation but it is not significant. So, it is possible that the vaccine is not very effective in preventing the infection, and that the vaccination is a kind of comfort to the person, which will produce a relaxed mentality and lead to more infections. The control measures had a significant effect on the total numbers of cases and deaths per million for the first two strains, but did not show a significant effect in the Omicron phase, probably because the existing measures were no longer effective in controlling the outbreak as the strains continued to spread rapidly. Unlike other countries, control measures have a significant negative correlation with the total numbers of cases and deaths per million in type D2 countries, and according to

---

**Table 2**

Spearman correlation coefficient table.

| x1  | x2  | x3  | x4  | x5  | x6  | x7  | x8  | x9  | x10 | x11 | x12 | x13 | x14 | x15 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A1  | 0.67| 0.69| 0.34| 0.31| 0.34| 0.31| 0.54|     |     |     |     |     |     |     |     |
| A2  | -0.35| 0.34| 0.33| -0.37| 0.36| -0.34| 0.71| 0.62| 0.38|     |     |     |     |     |     |
| D1  | 0.50|     | 0.56| -0.46| 0.36| 0.50| 0.31|     |     |     |     |     |     |     |     |
| D2  |     |     | 0.49| 0.40| -0.47| -0.73| -0.56|     |     |     |     |     |     |     |     |
| O1  | 0.31| 0.67| 0.31| 0.33| 0.59| 0.45| 0.35| -0.43| 0.36| 0.35| 0.53|     |     |     |     |
| O2  | -0.52| 0.72|     | -0.47| 0.83| 0.53| -0.53| 0.33| -0.87| 0.47|     |     |     |     |     |
| O3  | 0.32| 0.72|     |     |     |     |     |     |     |     |     |     |     |     |     |

**Table 2 (continued)**

| x1  | x2  | x3  | x4  | x5  | x6  | x7  | x8  | x9  | x10 | x11 | x12 | x13 | x14 | x15 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A1  | 0.54| 0.53| 0.40| 0.43| 0.47| 0.77|     |     |     |     |     |     |     |     |     |
| A2  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| D1  | -0.31| 0.30| -0.51| 0.59| -0.72| -0.89|     |     |     |     |     |     |     |     |     |
| D2  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| O1  | 0.62| 0.51| 0.43| 0.34| -0.31| 0.52|     |     |     |     |     |     |     |     |     |
| O2  | 0.36| 0.33| -0.40| 0.36| -0.37| -0.46|     |     |     |     |     |     |     |     |     |
| O3  | 0.53| 0.40| 0.33| -0.54| 0.68| 0.54| -0.76| 0.53| -0.57|     |     |     |     |     |     |

**Table 2 (continued)**

| x1  | x2  | x3  | x4  | x5  | x6  | x7  | x8  | x9  | x10 | x11 | x12 | x13 | x14 | x15 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A1  | -0.35|     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| A2  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| D1  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| D2  | -0.55|     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| O1  | -0.36| 0.46| -0.41|     |     |     |     |     |     |     |     |     |     |     |     |
| O2  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| O3  | -0.48| 0.75| -0.55| -0.42|     |     |     |     |     |     |     |     |     |     |     |     |

**Table 2 (continued)**

| x1  | x2  | x3  | x4  | x5  | x6  | x7  | x8  | x9  | x10 | x11 | x12 | x13 | x14 | x15 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A1  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| A2  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| D1  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| D2  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| O1  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| O2  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| O3  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

---
Table S2 in Appendix S1, type D2 countries have the highest GDP and vaccination rates, the lowest proportion of people over 65, but the least strict and poor control measures, which indicates that with the increase of control measures, the total numbers of cases and deaths per million in type D2 countries will decrease, and the epidemic can be better controlled.

3.4. Regression analysis

In order to determine the specific relationships between the number of confirmed cases, deaths per million and the mortality rate respectively, and 15 important variables, and to make predictions and explanations based on the obtained relational equations, regression analyses were next done for categories with relatively more countries, A1, D1, and O1. It can be found that during epidemics induced by the Original/Alpha strain, the selected variables explained 0.73 for the total numbers of cases per million, 0.80 for the total deaths per million, and 0.81 for the mortality rates in type A1 countries, and the p-values were less than 0.05, indicating that the linear regression equation was significant (Table S11). The regression equations are as follows:

\[ y_1 = 0.40x_1 + 0.55x_2 + 0.42x_{11} + 0.42x_{15}, \]  
\[ y_2 = -0.53x_1 + 0.88x_2 + 0.42x_3 - 0.90x_6 + 0.53x_7 + 0.57x_8 + 0.36x_{10} + 1.47x_{11} + 0.37x_{15}, \]  
\[ y_3 = -0.41x_5 - 0.41x_7 + 0.75x_{10} - 0.41x_{14} - 0.34x_{15}. \]

From the regression results, it can be seen that the total numbers of cases per million rises with population density, the proportion of people over 65, diabetes prevalence, and the vaccination rates. The total numbers of deaths per million rises with higher variables such as the proportion of people over 65, extreme poverty index and prevalence of diabetes, and falls with higher GDP. Mortality rate decreases with higher percentage of male smokers, extreme poverty index, vaccination rates, and stringency index, and increases with higher health expenditures. Overall, it appears that during the Original/Alpha strain stage, for A1 countries with high control efficiency, low standard deviation, and low mortality rate, the proportion of people over 65, economic situation, and diabetes prevalence have a more significant impact on the prediction of COVID-19’s severity.

As shown in Table S12 in Appendix S1, the selected variables have a low degree of explanation for the total numbers of cases, deaths per million and mortality rates in type D1 countries. The results of the generalized additive model fitting selected are shown in Table S14 in the Appendix S1, and the specific expressions are as follows.

\[ y_1 = 817.9x_5 + 3711.7x_{11} + ns(x_1, df = 2) + ns(x_2, df = 4) + ns(x_{14}, df = 4), \]
\[ y_2 = 23.85x_2 + 10.36x_5 + 9.48x_9 + 7.81x_{10} + 34.89x_{11} + ns(x_{14}, df = 3), \]

Fig. 5. Non-linear relationships between population density (a), the proportion of people over 65 (b), and stringency of control measures (c) and the total number of cases per million; non-linear relationships between the stringency of control measures (d) and the total number of deaths per million; non-linear relationships between the proportion of people over 65 (e), GDP (f), hospital beds per thousand (g), health expenditures (h), and stringency of control measures (i) and the mortality rate in type D1 countries. Where the black line represents the curve of the smoothed regression, the shaded part represents the 95% confidence interval.
The regression results show that the explanatory variables account for 92.9% of the variation in the total numbers of cases per million and that the total numbers of cases per million have a significant linear relationship with the percentage of smokers and the prevalence of diabetes, and total numbers of cases per million increases with increases of these two variables. The total numbers of cases per million had a significant non-linear relationship with population density, the proportion of people over 65, and the stringency of control measures (Fig. 5). The total numbers of cases per million rises with increasing population density when the population density is high, and both too strict and too relaxed control measures lead to an increase in the total numbers of cases per million. The generalized additive model for the total numbers of deaths per million has an explanatory power of 83%, and the total numbers of deaths per million has a significant linear relationship with the proportion of people over 65, the percentage of smokers, public health facilities, health expenditures, and the prevalence of diabetes, with a positive relationship. There is a significant non-linear relationship with the stringency of control measures (Fig. 5), which shows that an appropriate level of control is needed to effectively reduce the total numbers of deaths per million. The generalized additive model for mortality rate has an explanatory degree of 93.2%, and mortality rate shows a significant linear relationship with the extreme poverty index, public health facilities, and diabetes prevalence, with mortality rate decreasing as these three variables increase. The proportion of people over 65, GDP, hospital beds per thousand, health expenditures and stringency of control measures also have a significant non-linear relationship with mortality rate. From Fig. 5, it is clear that a higher proportion of people over 65 leads to an increase in mortality rate, which decreases when GDP and hospital beds per thousand are sufficiently high, with the appropriate control strength required to avoid serious consequences of the epidemic. Both too low and too high stringency of control measures are not recommended. Overall, it appears that the stringency of control measures also plays a significant role in predicting the severity of COVID-19 during the Delta period, for type D1 countries with fast-rising, high continuity.

From Table S13 in Appendix S1, it can be seen that the filtered variables explained 83% for the total numbers of deaths per million in type O1 countries, and the p-value was less than 0.05, indicating that the linear regression equation was significant, but the linear regression equation for total numbers of cases per million and mortality rates in type O1 countries was not significant. Similarly, we adopted the generalized additive model, when the $y_1$ itself was fitted with the model, the interpretability was 41.5% and the independent variables was not significant, while the fitness was significantly higher at 86.6% after the logarithmic transformation of $y_1$. So, the final selected model fitting results are shown in Table S15 in Appendix S1. The specific expressions of the regressions were as follows:

$$y_3 = -0.0017 x_7 - 0.0006 x_9 - 0.0038 x_{13} + n s(x_2, df = 1) + n s(x_6, df = 4) + n s(x_8, df = 2) + n s(x_{10}, df = 1) + n s(x_{14}, df = 2).$$

(12)

The regression results from the generalized additive model of total numbers of cases per million and the explanatory variables showed that the explanatory power reached 86.6%, and the total numbers of cases per million had a significant linear relationship with the percentage of smokers, public health facilities and vaccination rates, and as these three variables increased, the total numbers of cases per million increased. There was a significant non-linear relationship with the prevalence of the extreme poverty index, HIV and tuberculosis prevalence (Fig. 6). The total numbers of deaths per million increases with the proportion of people over 65, life expectancy, and the extreme poverty index, and decreases with population density and tuberculosis prevalence. This is somewhat different from the results of the correlation analysis, and considering whether there is an interaction effect on the variables, the new regression results (Table S15 in Appendix S1), show that there is an interaction effect of tuberculosis prevalence with both population density and the extreme poverty index, and that with the inclusion of these two interaction terms, the total numbers of deaths per million rises with population density and tuberculosis prevalence, which can be seen as an interaction effect hiding the positive effect of these two variables. The generalized additive model for mortality rate has an explanatory degree of 81.7%, with significant nonlinear relationships between mortality rate and population density, the proportion of people over 65, hospital beds per thousand, health expenditures and the stringency of control measures. From Fig. 6, we can see that a higher proportion of people over 65 leads to an increase in mortality rate, which decreases when health expenditures are high enough. Overall, it seems that the development of the epidemic is more complicated during the Omicron strain epidemic for type O1 countries with high control efficiency and continuity, but a reasonable allocation of health care resources can still play a positive role.

### 4. Discussion

We analyzed the trends and characteristics of the epidemic in various Asian countries, taking into account the key factors affecting the epidemic trends, the number of infections, and the number of deaths in each Asian country under different periods and strains to provide recommendations on epidemic prevention measures for different categories of Asian countries. Asian countries were classified based on the time-series shape and feature values of the epidemic respectively, in addition to considering the composition structure of similar countries’ feature variables, we also obtained the curve dynamics of the epidemic. When considering the development of the epidemic in Asian countries, Asian countries were finally classified into four categories according to the similarity of shapes by adopting the K-shape clustering algorithm based on SBD distance. When studying the composition structure of the epidemic in Asian countries, six feature values for each of the three strains during the study period were selected, using the hierarchical clustering method for clustering, and Asian countries were classified into two categories during epidemics induced by the Original/Alpha strain, three categories during epidemics induced by the Delta strain, and four categories during epidemics induced by the Omicron strain, using 23 January 2020 as the start date, while using 1 July 2021 and 1 January 2022 as the cut-off dates. Fifteen variables were considered and analyzed for the explanation of epidemiological trends in the classified Asian countries, including vaccination rates and a stringency index. By comparing different countries in the same period and different periods in the same country, respectively, we also summarized the experience of countries in preventing and controlling the epidemic in the past two years, revealing the weaknesses in the prevention and control of the epidemic in each category of countries, which helps to understand the spread pattern of the epidemic and provides theoretical support on how to take efficient measures against the epidemic.

Asian countries were classified into four categories according to the similarity of shapes. We selected the representative Asian countries of Japan, Thailand, Singapore, and Yemen, then analyzed and compared the characteristics of these five Asian countries at different periods to make feasible suggestions for the next epidemic prevention in Asian countries. Japan, the most mature country in Asia, with the most
adequate number of hospital beds and a more developed economy, took more relaxed control measures during the Original/Alpha strain period and had a higher mortality rate, but as Japan took stricter control measures than before during the Delta strain and Omicron strain periods, accompanied by higher vaccination rates, the mortality rate gradually decreased. Japan was classified into category 4 by shape, indicating that the epidemic is now under control there after experiencing rapid growth in the mid-term. Thailand has a high proportion of people over 65, but public health facilities are more than adequate, and in the initial stages of the epidemic, strict control measures were not taken and vaccination rates were low, but in the later stages, vaccination rates increased. The results of clustering by shape revealed that Thailand belongs to the second category, i.e. the outbreak was more rapid in the early stages, but was better controlled in the later stages without a massive outbreak. Singapore, as the most densely populated Asian country with a good economic base and a relatively high degree of aging, had high vaccination rates and a relatively low mortality rate attributable to the Original/Alpha strain, but took more relaxed control measures at the Delta strain stage, resulting in a higher-than-average peak value of confirmed cases per capita, although the stringency of control measures increased at the Omicron stage, the peak value of confirmed cases per capita was still higher in the Omicron phase. Singapore was assigned to the fourth category based on shape clustering, indicating that the epidemic is now developing relatively smoothly after a larger outbreak in the medium term. Yemen, as the poorest economically based country in Asia, has lower health expenditures, fewer public health facilities, and hospital beds per thousand, as well as the lowest vaccination rates and more relaxed control measures. Due to the highest mortality rate, which deviates significantly from other Asian countries, Yemen has been in a category of its own when clustering based on characteristic variables. According to the shape, Yemen was assigned to the first category, and it can be found that such countries have recurrent outbreaks in the later stages of the epidemic and are still in an unstable stage of development. For some countries with a risk of re-emergence of the epidemic at a later stage, such as those classified 1 and 3 by shape, an appropriate increase in the stringency of control measures can be considered. For countries classified 2 and 4 by shape, where the epidemic is better controlled at a later stage, the original anti-epidemic policy can be continued. For countries with relatively lax epidemic control and low efficiency in prevention and control, but with an excellent economic base, after which it is recommended to increase spending on the epidemic and to increase the stringency of control measures and continue to improve vaccination. For countries with very poor economic levels, shortage of hospital beds and public health facilities, and high mortality rate, it is recommended to invest more in health facilities to save patients from critical cases of the diseases.

The cluster analysis results based on features show that during the Original/Alpha strain epidemic period, Asian countries were clustered into two main groups, A1 and A2. The type A1 countries characterized by high control efficiency, low standard deviation, and low mortality rate, the type A2 countries fluctuates sharply, increases rapidly and is highly continuous. The total numbers of cases per million in type A1 countries has a higher correlation with variables such as GDP, life expectancy, and vaccination rates, while the
total numbers of cases per million in type A2 countries has a higher correlation with the prevalence of diabetes, and the stringency of control measures, and GDP: A1 < A2, life expectancy: A1 < A2, vaccination rates: A1 > A2, the prevalence of diabetes: A1 < A2, and the stringency index: A1 > A2, which may make the trends of epidemic development different in the two groups of countries.

During the Delta strain epidemic period, Asian countries were clustered into three major groups, D1, D2, and D3, with type D1 countries characterized by a rapid rise in epidemic development and high continuity, but type D1 countries had the lowest total number of cases and deaths per million. Type D2 countries have an overall epidemic trend of poor continuity and control but low mortality rate, with the highest total cases per million in this type of country. Type D3 Countries with high mortality rate and unstable development, but with high control efficiency, have the highest total deaths per million, the lowest population density, life expectancy, GDP, hospital beds per thousand and public health facilities, but have the strictest control measures than type D1 and D2 countries, health expenditures are higher and slightly lower than in D1 countries. The mortality rate in both D1 and D2 countries were highly correlated with the vaccination rates. It can be found that the stringency of control measures: D3 > D1 > D2, vaccination rates: D2 > D3 > D1, GDP: D2 > D1 > D3, health expenditures: D1 > D3 > D2, and control efficiency: D3 > D1 > D2, which indicates that increasing vaccination rates can reduce mortality rates (Hart et al., 2022), but too relaxed control measures can have the consequence of unstable epidemic development, while taking proper control measures help to improve control efficiency.

During the Omicron strain epidemic period, Asian countries were classified by clustering into four major groups, O1, O2, O3, and O4. Among them, type O1 countries have high control efficiency and continuity, these countries have a lower mortality rate, slightly higher than type O4 countries. Type O2 countries are characterized by low standard deviations, low average growth rates, poor control efficiency, and continuity, these countries have the highest mortality rate. Type O3 countries have the highest average growth rate, small outbreak size, and the lowest total cases per million, but higher mortality rate, second only to type O2 countries. Type O4 countries have unstable development and large outbreak sizes. These countries have the highest total cases and deaths per million, but the lowest mortality rate, the highest proportion of people over 65, the longest life expectancy, and the highest number of hospital beds per thousand, vaccination rates, and stringency index among the 4 types of countries, and the lowest prevalence of both diabetes and tuberculosis. The total numbers of cases per million in O1 countries were highly correlated with GDP and life expectancy, while the total numbers of cases per million in O2 countries were highly correlated with health expenditures, while the total numbers of cases per million in O3 countries were highly correlated with important variables such as life expectancy, GDP, and tuberculosis prevalence. We found that GDP: O1 > O3 > O4 > O2, life expectancy: O4 > O1 > O3 > O2, vaccination rates: O4 > O1 > O2 > O3, health expenditures: O3 > O4 > O1 > O2, and tuberculosis prevalence: O3 > O2 > O1 > O4, which indicates that in the Omicron period, countries with a good economic base were also more effective in controlling the epidemic, while countries with severe aging, poor economic base and inadequate vaccination will face high mortality rates. In addition, the control measures had a significant effect on the total numbers of cases and deaths per million for the first two strains, but did not show a significant effect in the Omicron phase. This is probably because the existing measures were no longer effective in controlling the outbreak as the strains continued to spread more rapidly, which is similar to the previous study that interventions might be less effective if transmission of the virus occurs quickly (Hart et al., 2022).

The total numbers of deaths and cases per million were highly correlated with the proportion of people over 65, the prevalence of multiple diseases, and national GDP, but also with vaccination rates, health expenditures, and stringency of control measures. Compared with the mortality of the three periods, the mortality during the Omicron period was significantly lower than that of the previous two periods. This is consistent with the result in Nyberg et al., 2022 that the risk of severe outcomes following SARS-CoV-2 infection is substantially lower for Omicron than for Delta. During the Delta strain epidemic period, the total numbers of cases per million: D2 > D3 > D1 and the total numbers of deaths per million: D3 > D2 > D1, where the GDP of D1 countries is at an intermediate level and the GDP of D2 countries is the highest. It is also clear that countries with high GDP face problems of relatively high population density, high proportion of people over 65, high diabetes prevalence and percentage of smokers, and insufficient attention to epidemic prevention and control interventions, resulting in the highest number of infections, but low mortality rate due to their economic and medical conditions; on the other hand, countries with medium GDP have the lowest total cases and deaths per million due to the high number of hospital beds per thousand, high health expenditures, and low percentage of smokers and diabetes prevalence. Focusing on the Omicron epidemic period’s significant variables and the six feature values, we can find that type O4 countries have the highest proportion of people over 65, but the lowest mortality rate. In contrast, type O3 countries have the second highest ageing and higher mortality rate among the four categories, while type O2 countries have the highest mortality rate despite having the lowest proportion of people over 65. Comparing these three types of countries, type O4 countries have the highest vaccination rates and strict control measures, while O2 and O3 countries have much lower vaccination rates than O4 countries, which again confirms the importance of vaccination to combat COVID-19.

In conclusion, the results obtained on the epidemic trends in Asian countries from 23 January 2020 to 15 April 2022 are not only helpful to understand the epidemic development model of Asian countries during the COVID-19 pandemic, but also reveal the underlying root causes. Therefore, the analysis of key influencing factors of various categories can be used to help countries understand the problems in epidemic prevention and control and learn from the successful experience of other countries. The research methods adopted in this paper can also be used to study epidemics and the factors influencing them in other regions or epidemics. Besides, there are also some limitations. Since the lack of strain proportion data of Asian countries, the substitution process of strains is ignored in this paper. This paper mainly focus on the cumulative infection history and the analysis may provide little insight into how past infection and vaccination dynamics impact future waves. Given the large geographic span of countries in Asia, effects of climatic factors may need to be considered in the future. Moreover, the pattern of epidemic development in Asian countries will always be in a state of change, and it is necessary to continuously monitor the subsequent time series changes and adjust the clustering results according to the real situation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

H. Tu et al.
the work reported in this paper.

Acknowledgements

The authors thank Prof Robert A. Cheke for his kind help and comments. This work was supported by the National Natural Science Foundation of China (NSFCs: 12171295, 12031010) and the Project of Science and Technology Young Star in Shaanxi Province of China (2022KJXX29).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtbi.2022.111336.

References

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., et al., 2013. An extensive comparative study of cluster validity indices[J]. Pattern recognition 46 (1), 243–256.

Bakhshandeh, B., Jahanfrooz, Z., Abbasi, A., et al., 2021. Mutations in SARS-CoV-2: Consequences in structure, function, and pathogenicity of the virus[J]. Microbial Pathogenesis 154, 104831.

Carrillo-Larco, R.M., et al., 2020. Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach[J]. Wellcome Open Research.

Chen J.X. 5 million lives lost, the marathon against the epidemic is far from over. http://epaper.ynet.com/html/2021-11/02/content_386087.htm?div=-1.

Chodick, G., Omer-Gilon, M., Derazne, E., et al., 2021. Adolescent body mass index and changes in pre-pregnancy body mass index in relation to risk of gestational diabetes [J]. E ClinicMedicine 42, 10121.

Farseev, A., Chu-Farseeva, Y.Y., Qi, Y., et al. Understanding economic and health factors impacting the spread of COVID-19 disease [J]. medRxiv, 2020.

Ferguson N M, Laydon D, Nedjati-Gilani, G. et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand[J]. 2020.

Flaxman, S., Mishra, S., Gandy, A., et al., 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe[J]. Nature 584 (7820), 257–261.

Hart, W.S., Miller, E., Andrews, N.J., et al., 2022. Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological analysis[J]. The Lancet Infectious Diseases 22 (5), 603–610.

He, S., Peng, Y., Sun, K., 2020. SEIR modeling of the COVID-19 and its dynamics[J]. Nonlinear dynamics.

Hu, S.K., 2020. Assessment of schistosomiasis transmission level based on ecological niche model[D]. Chinese Center for Disease Control and Prevention.

Siddiqui, M.K., Morales-Menendez, R., et al., 2020. Correlation between temperature and SARS-CoV-2 variants: an epidemiological analysis[J]. The Lancet Infectious Diseases 22 (5), 603–610.

Yi, D.L., Li, G.M., Leng, H.M., 2020. Cluster analysis of regional differences in the epidemiological characteristics and prevention and control studies of a novel Omicron variant of coronavirus[J]. Chinese general medicine 25 (01), 48–51.

Wang Y.X, Zhao Y. Zhong Nanshan and Zhang Boli has the latest research on Strain Delta and vaccination[N]. Science and Technology Daily, 2021-08-23(002).

Wang, Y.H., Liang, T.T., Li, S.H., et al., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study[J]. The Lancet.

Wu, M., Liu, J., Liu, M., Liang, W.N., 2022. The epidemiological characteristics and prevention and control studies of a novel Omicron variant of coronavirus[J]. Chinese general medicine 25 (01), 14–19.

Wu, M., Liu, J., Liu, M., Liang, W.N., 2022. Epidemiological characteristics and prevention and control studies of a novel Omicron variant of coronavirus[J]. Chinese general medicine 25 (01), 14–19.

Xie, C.H., Wu, S.X., Zhang, C., Sun, W.T., He, H.F., Pei, T., Luo, G.P., 2021. Time series characterization of COVID-19 epidemics in countries around the world based on spectral clustering[J]. Journal of Clinical Medicine.

Yang, C., Wang, J., 2021. Modeling the transmission of COVID-19 in the US-A case study[J]. Infectious Disease Modelling.

Yu, D.L., Li, G.M., Leng, H.M., 2020. Cluster analysis of regional differences in the development of novel coronavirus pneumonia epidemic[J]. Journal of Chongqing Medical University.

Zarikas, V., Pouloupolous, S.G., Gareiou, Z., et al., 2020. Clustering analysis of countries using the COVID-19 cases dataset[J]. Data in brief.

Zhang, H.F., Shu, Y.T., Yang, O., 1997. Estimation of Hurst parameter by variance-time plots[J]. Journal of Urban Planning and Development 7, 71–80.