

# Dự đoán tình hình COVID-19 tại Việt Nam bằng mô hình SIR

Vũ Công Thành<sup>1</sup>, Ngô Trường Huy<sup>1</sup>, and Phạm Bá Cường<sup>1</sup>

{19120374, 19120242, 19120183}@student.hcmus.edu.vn

<sup>1</sup>Chương trình Cử nhân tài năng  
Khoa Công nghệ Thông tin  
Trường Đại học Khoa học Tự nhiên  
Đại học Quốc gia TP.HCM, Việt Nam

**Abstract.** Đại dịch COVID-19 là một dịch bệnh truyền nhiễm do virus SARS-CoV-2 cùng với các biến thể của nó gây ra trên phạm vi toàn cầu, Việt Nam cũng là một trong các quốc gia chịu nhiều ảnh hưởng về kinh tế và đời sống. Chính vì vậy chúng tôi đã sử dụng mô hình SIR để dự đoán các ca mắc, ca điều trị khỏi và người có nguy cơ lây nhiễm dựa trên dữ liệu từ Bộ Y tế. Kết quả cho thấy mô hình có thể dự đoán ngắn hạn được xu hướng bùng phát dịch bệnh, tuy nhiên chưa chính xác khi dự đoán số lượng do mô hình chưa khớp đủ tốt với dữ liệu.

## 1. Giới thiệu

SIR là một mô hình toán học cơ bản và kinh điển về dịch bệnh được giới thiệu trong bài báo từ 1927 của Kermack và McKendrick [1]. Đây là một trong những mô hình đơn giản nhất để mô tả sự thay đổi của 3 nhóm  $S$ ,  $I$  và  $R$  theo thời gian  $t$ . Mô hình này có thể sử dụng để dự đoán sự lan truyền dịch bệnh lây nhiễm từ người sang người, có thể sử dụng để nghiên cứu và phân tích đối với nhiều bệnh khác nhau như sởi, SARS, MERS hay gần đây nhất là đại dịch COVID-19.

Đại dịch COVID-19 ảnh hưởng đến nhiều mặt trong đời sống của mỗi người do đặc tính lây lan nhanh giữa người với người và sự nguy hiểm của nó nếu mắc phải. Để đưa ra được một chiến lược tối ưu trong công tác phòng và chống dịch COVID-19 để sớm đưa cuộc sống của mọi người về trạng thái bình thường mới, việc dự đoán xu hướng bùng phát dịch là một công việc hết sức cần thiết trong giai đoạn căng thẳng này.

Trong nội dung của bài báo này chúng tôi sẽ trình bày về việc áp dụng mô hình SIR để dự đoán tình hình bùng phát dịch bệnh dựa trên số liệu có được tại Việt Nam. Dữ liệu được chúng tôi thu thập từ trang Dữ liệu COVID-19 <sup>1</sup>

Bài báo này sẽ có cấu trúc như sau: mục 2 sẽ là phần trình bày phương pháp đề nghị để giải quyết bài toán, mục 3 mô tả quá trình thực nghiệm của chúng tôi trên dữ liệu thu thập được tại khu vực Việt Nam. Cuối cùng là mục 4, chúng tôi sẽ tổng kết và đưa ra các đề xuất cho các nghiên cứu trong tương lai

## 2. Phương pháp đề nghị

### 2.1. Mô hình SIR

Mô hình SIR xác định trạng thái của một người trong quần thể theo 3 nhóm, dự đoán của mô hình cho biết sự thay đổi trạng thái giữa các trạng thái theo thời gian:

<sup>1</sup>Trung tâm Giám sát An toàn Không gian mạng Quốc gia (nguồn từ Bộ Y tế): <https://covid19.ncsc.gov.vn/dulieu>

- **Susceptible:** Những người có nguy cơ mắc bệnh
- **Infected:** Những người đang bị nhiễm bệnh và có thể lây lan sang người khác
- **Recovered / Removed:** Những người không còn khả năng mắc bệnh

Số người thuộc các nhóm trạng thái bên trên tại thời điểm  $t$  là  $S(t)$ ,  $I(t)$  và  $R(t)$ . Trong mô hình đơn giản nhất với tổng dân số là hằng số, mối quan hệ giữa  $S(t)$ ,  $I(t)$  và  $R(t)$  tại mọi thời điểm  $t$  là:

$$S(t) + I(t) + R(t) = N = \text{const}$$

Khi  $N$  càng lớn, ví dụ như dân số của một thành phố hay một quốc gia, sự chuyển đổi trạng thái theo thời gian của hệ  $S$ ,  $I$ ,  $R$  càng xấp xỉ gần đến hệ phương trình vi phân:

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Mỗi phương trình trong hệ trên biểu thị độ dốc (tốc độ thay đổi) của các đại lượng  $S$ ,  $I$  và  $R$  tại thời điểm  $t$  theo trạng thái của toàn bộ hệ tại cùng thời điểm đó. Tham số  $\beta$  thể hiện tỉ lệ lây nhiễm theo đầu người, tức là xác suất trung bình để một người chuyển đổi từ trạng thái  $S$  sang trạng thái  $I$  (từ khỏe mạnh trở thành người mắc bệnh). Còn tham số  $\gamma$  biểu thị tỉ lệ hồi phục, cho biết khoảng thời gian mắc bệnh trung bình (thời gian ở trong trạng thái  $I$ ) là  $\frac{1}{\gamma}$ .

Khi xét một mô hình dịch bệnh như SIR, ta cần quan tâm đến đại lượng  $R_0$  là hệ số lây nhiễm cơ bản. Với mô hình SIR cơ bản nhất,  $R_0 = \frac{\beta N}{\gamma}$  trong đó  $\beta N$  là trung bình số người khỏe mạnh bị một người mắc bệnh lây nhiễm lây cho trong khoảng thời gian mắc bệnh  $\frac{1}{\gamma}$ . Khi  $R_0 < 1$ , dịch sẽ không thể bùng phát do người bệnh không thể lây cho nhiều hơn một người trong khoảng thời gian  $\frac{1}{\gamma}$  nên số lượng người nhiễm bệnh (trạng thái  $I$ ) ngày càng giảm dần, ngược lại nếu  $R_0 > 1$  tức là người bệnh có thể lây nhiễm cho nhiều hơn một người trong khoảng thời gian mắc bệnh, bệnh dịch có thể bùng phát do số lượng người mắc bệnh sẽ tăng theo cấp số nhân.

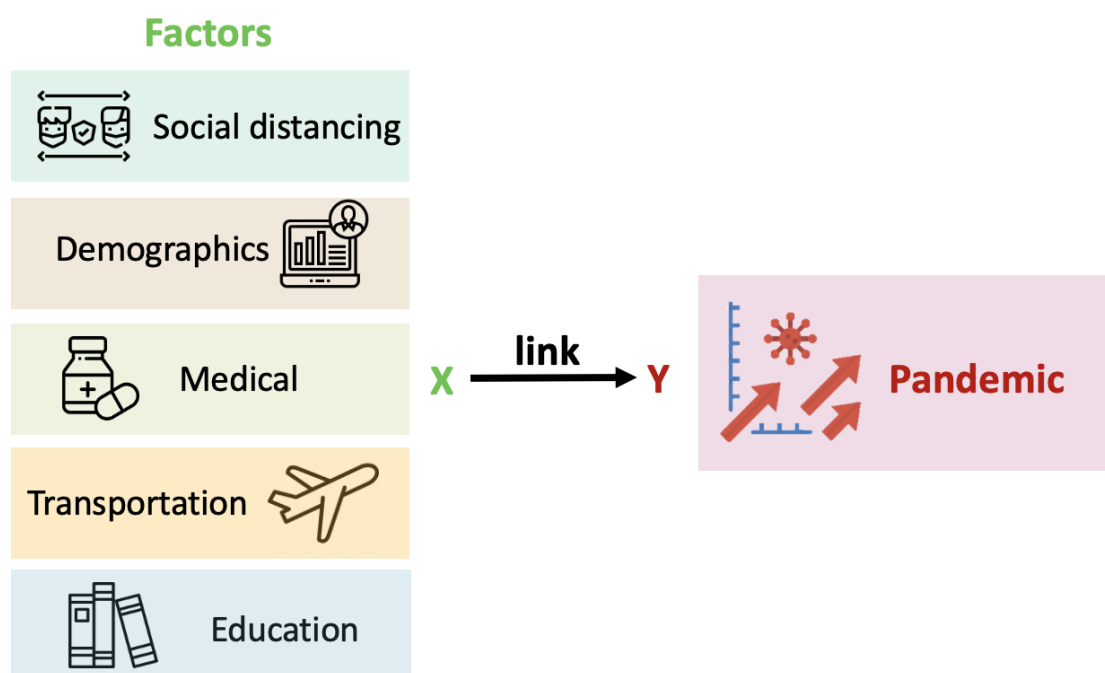


Hình 1. Sơ đồ biểu diễn sự thay đổi trạng thái trong mô hình SIR [2]

## 2.2. Ứng dụng mô hình SIR để nghiên cứu về COVID-19 ở Việt Nam

Việc dự đoán những biến động về tình hình hình dịch bệnh ở một khu vực dân cư đòi hỏi rất nhiều yếu tố, trong đó có các thông tin về ca nhiễm, ý thức của cộng đồng, tỉ

lệ tiêm vaccine cùng với lối sống tại mỗi khu vực đều ảnh hưởng đến kết quả dự đoán [3]. COVID-19 còn có khả năng gây bệnh cho người đã từng nhiễm bệnh, nên tùy vào từng vùng mà các kết quả thu được có thể tốt hoặc không (nếu người có khả năng nhiễm hoặc đã từng nhiễm bệnh tránh tiếp xúc người khác và có các biện pháp phòng tránh như đeo khẩu trang nơi công cộng cũng sẽ làm giảm đáng kể khả năng nhiễm mới hoặc tái mắc bệnh). Mô hình SIR không thể đưa ra kết quả chính xác nhất đối với các khu vực có tỉ lệ tái nhiễm cao, khi đó ta cần một mô hình phức tạp hơn như SIRS. Ngoài ra, do người bị nhiễm COVID-19 có thời gian ủ bệnh khá dài, trong thời gian ủ bệnh có thể vô tình làm lây lan dịch bệnh mà không lường trước được, với trường hợp này có thể ta sẽ cần mô hình như SEIR để theo dõi sự thay đổi theo thời gian của những người phơi nhiễm bệnh.



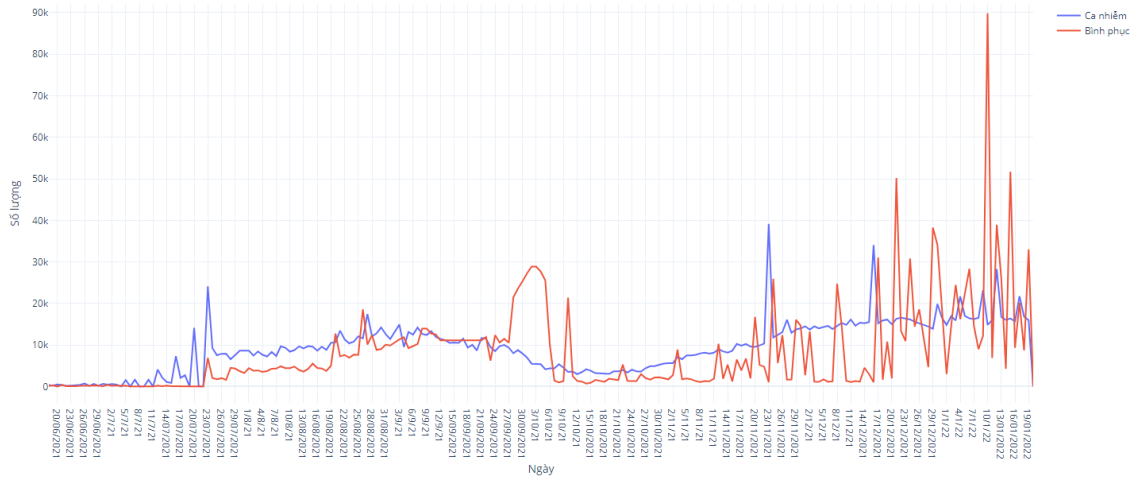
**Hình 2. Các yếu tố ảnh hưởng đến việc lây lan dịch bệnh COVID-19 [4]**

Chính vì điểm yếu của mô hình SIR với các trường hợp kể trên nên chúng tôi giả định ở Việt Nam những người từng nhiễm sẽ không tái nhiễm, đồng thời chúng tôi cũng giả định rằng việc lây nhiễm chỉ diễn ra giữa những người quen (chung nơi làm việc, chung gia đình...) trong quá trình sinh hoạt chung, và mọi người đều có ý thức trong việc phòng chống dịch bệnh ở nơi công cộng để tránh lây nhiễm cho người khác. Trong trường hợp tổng quát, chúng tôi giả sử quần thể không thay đổi dân số (bỏ qua tỉ lệ gia tăng dân số tự nhiên).

Dữ liệu mà chúng tôi thu thập được bao gồm các thông tin về số người mắc, số người khỏi bệnh và số người tử vong theo thời gian trong khoảng thời gian từ ngày 01/12/2020 đến nay (thời điểm viết bài báo này là ngày 19/01/2022). Bên cạnh đó nguồn dữ liệu còn cho chúng tôi thêm thông tin về tỉ lệ tiêm vaccine và chỉ số giãn cách, tuy nhiên do bài báo này chỉ khảo sát mô hình SIR cơ bản nên chúng tôi lược bỏ các thông tin khác và chỉ giữ lại các thông tin về số người mắc, số người đã bình phục và số người đã tử vong.

### 3. Thực nghiệm

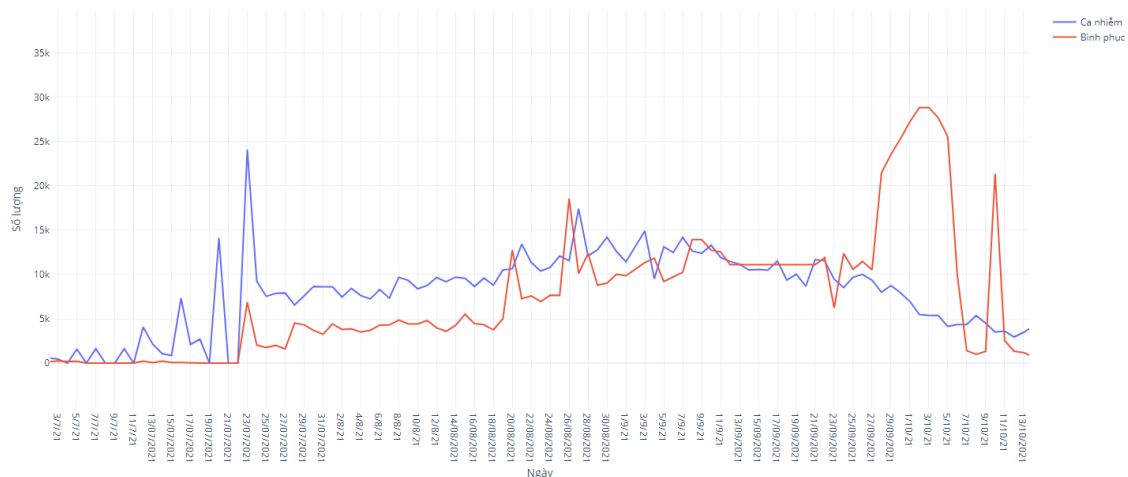
Chúng tôi đã vẽ biểu đồ bên dưới (Hình 3) để phân tích tình hình dịch bệnh trong năm vừa qua và chúng tôi nhận thấy rằng có hai lần bùng dịch tại Việt Nam đạt đỉnh vào khoảng cuối tháng 8 và cuối tháng 12.



**Hình 3. Biểu đồ dữ liệu thực tế theo ngày tại Việt Nam trong khoảng thời gian từ ngày 21/04/2021 đến nay**

Chúng tôi đã thử nghiệm mô hình SIR tối ưu với hàm mất mát được tính toán dựa trên 90.1% sự chênh lệch giữa số liệu thực tế và dự đoán của số lượng người bình phục và 9.9% sự chênh lệch của số lượng người đang nhiễm bệnh

#### 3.1. Đợt bùng dịch đầu tiên (đạt đỉnh cuối tháng 8)

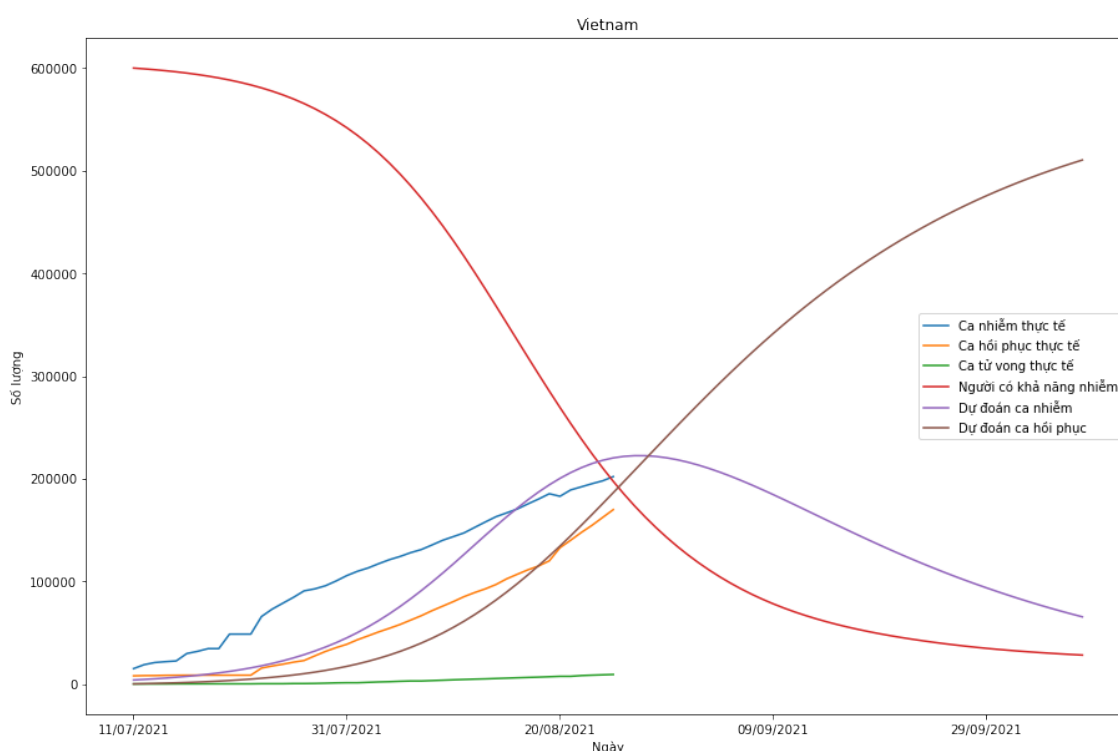


**Hình 4. Biểu đồ thể hiện dữ liệu thực tế theo ngày trong khoảng thời gian từ ngày 03/07/2021 đến ngày 13/10/2021**

Biểu đồ trên (Hình 4) cho ta biết trong đợt bùng dịch đầu tiên, số lượng ca mắc và số lượng ca bình phục theo ngày đều tăng dần đến khoảng cuối tháng 8 - đầu tháng 9, số

lượng ca nhiễm mới theo ngày đạt đỉnh ở cuối tháng 8. Từ đầu tháng 9 đến tháng 10 số lượng ca bình phục vẫn tăng, còn số lượng ca nhiễm giảm dần.

Để kiểm chứng mức độ hiệu quả của mô hình trong việc dự đoán tình hình dịch bệnh bùng phát, chúng tôi sẽ sử dụng dữ liệu trong khoảng thời gian từ **11/07/2021** đến **25/08/2021** để huấn luyện cho mô hình, sau đó cho mô hình dự đoán **45 ngày** sau đó. Biểu đồ dưới đây (Hình 5) thể hiện số ca nhiễm, số ca hồi phục và số ca tử vong thực tế từ dữ liệu và các dự đoán của mô hình SIR về ca nhiễm, ca hồi phục và người có khả năng nhiễm bệnh.



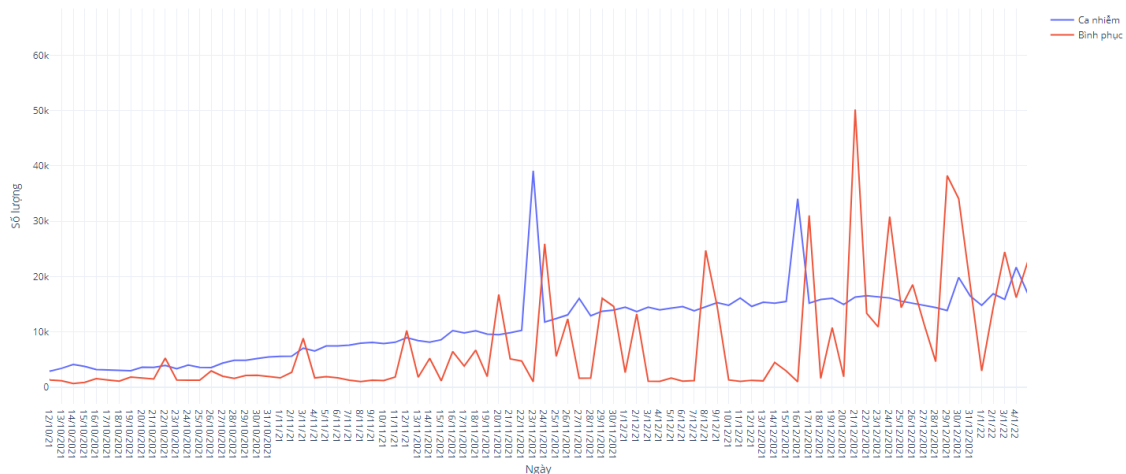
**Hình 5. Biểu đồ thay đổi các trạng thái được dự đoán bởi mô hình SIR áp dụng trên dữ liệu về COVID-19 tại Việt Nam từ ngày 11/07/2021 đến ngày 25/08/2021**

Chúng tôi nhận thấy trong biểu đồ bên trên (Hình 5) các đường dự đoán từ mô hình SIR dự đoán đúng đỉnh dịch vào cuối tháng 8 - đầu tháng 9 và giảm dần sau đó, cùng với đó là số ca hồi phục tăng mạnh. Bên cạnh đó, đường S biểu thị cho số lượng người có khả năng nhiễm cũng giảm mạnh vào lúc bùng phát dịch, đến tháng 10 thì số lượng người có khả năng nhiễm gần như không còn nữa (do chúng tôi đã giả định người sau khi bình phục thì không thể tái nhiễm nên khi gần như toàn bộ dân số đã bị nhiễm thì không còn người có khả năng mắc bệnh nữa).

Tuy nhiên mô hình SIR dự đoán số ca nhiễm và ca hồi phục theo ngày còn chênh lệch nhiều so với dữ liệu thực tế, do đó mô hình này chỉ có thể áp dụng để dự đoán xu hướng bùng phát dịch và thời điểm bùng phát chứ không thực sự hiệu quả trong việc dự đoán số lượng các cá thể chuyển đổi trạng thái. Lý do dẫn đến sự chênh lệch lớn là do mô hình SIR bỏ qua nhiều thông tin hữu ích cho việc dự đoán như tỉ lệ tiêm vaccine, thời gian ủ bệnh hay chỉ số giãn cách xã hội.

### 3.2. Đợt bùng dịch thứ hai (đạt đỉnh cuối tháng 12)

Chúng tôi đã thử nghiệm mô hình này trên khoảng thời gian từ **25/10/2021** đến **10/12/2021** (khoảng thời gian bắt đầu bùng dịch lần thứ 2) tuy nhiên mô hình không thể tìm ra các hệ số beta và gamma thích hợp với dữ liệu hiện có, chúng tôi nhận thấy điều này xảy ra có thể là do dữ liệu ca hồi phục không ổn định (có thể do dữ liệu không được cập nhật đầy đủ theo từng ngày).



Hình 6. Biểu đồ minh họa dữ liệu thực tế trong khoảng thời gian từ 12/10/2021 đến 04/01/2021

## 4. Tổng kết

COVID-19 được xem là thảm họa tồi tệ nhất mà thế giới phải chống chọi kể từ khi chiến tranh thế giới lần thứ 2 kết thúc, phân tích và dự đoán những sự thay đổi trạng thái là một trong những công việc cần thực hiện trong công tác phòng chống dịch. Việc sử dụng mô hình SIR cơ bản cũng là một trong những cách để dự đoán xu hướng chuyển đổi trạng thái bệnh phục vụ cho dự đoán bùng phát dịch để hệ thống y tế có sự chuẩn bị, giảm áp lực cho hệ thống y tế.

Tuy nhiên với tính chất của COVID-19 có thể tái nhiễm với người đã bình phục và có thời gian ủ bệnh, cùng với đó còn phụ thuộc vào các yếu tố con người nên mô hình SIR không cho kết quả tốt trong việc dự đoán số lượng người chuyển đổi giữa các trạng thái. Để đạt được kết quả tốt hơn với dịch bệnh COVID-19, ta cần một mô hình phức tạp hơn, với nhiều dữ liệu đầu vào hơn để tăng độ chính xác như SEIR, SIS hay SIRS. Bên cạnh đó việc tinh chỉnh mô hình cũng có thể giúp cải thiện độ chính xác cho các dự đoán tùy theo mỗi khu vực dân cư khác nhau.

## Tài liệu

- [1] W. O. Kermack and A. G. McKendrick, “Contributions to the mathematical theory of epidemics part I,” *Proceedings of the Royal Society of Edinburgh*, pp. 700–721, 1927.
- [2] S. Gounane, Y. Barkouch, A. Atlas, M. Bendahmane, F. Karami, and D. Mesquine, “An adaptive social distancing SIR model for COVID-19 disease spreading and forecasting,” *Epidemiologic Methods*, 2021.

- [3] D. Smith and L. Moore, “The SIR Model for Spread of Disease - The Differential Equation Model.” <https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model>, 2004. Trực tuyến; Truy cập: 17/01/2022.
- [4] Y. Li, “Feature Engineering and Integration of COVID-19 Panel Data.” <https://towardsdatascience.com/feature-engineering-and-integration-of-covid-19-panel-data-f97aaabc8017>, 2020. Trực tuyến; Truy cập: 17/01/2022.