Artificial Intelligence for Precision Viral Surveillance of Emerging Infection Disease (EID): The Metaverse-envisioned Study

Ting-Yu Lin, Ming-Fang Yen, Li-Sheng Chen, Chen-Yang Hsu, Yen-Po Yeh, Tony Hsiu-Hsi Chen

Submitted to: Journal of Medical Internet Research
on: November 15, 2024

Table of Contents

Artificial Intelligence for Precision Viral Surveillance of Emerging Infection Disease (EID): The Metaverse-envisioned Study

Ting-Yu Lin[1]; Ming-Fang Yen[2]; Li-Sheng Chen[2]; Chen-Yang Hsu[1]; Yen-Po Yeh[3]; Tony Hsiu-Hsi Chen[1]

[1]National Taiwan University Taipei TW
[2]Taipei Medical University Taipei TW
[3]Changhua County Public Health Bureau Changhua TW

Corresponding Author:
Tony Hsiu-Hsi Chen
National Taiwan University
Room 533, No. 17, Xuzhou Road, Zhongzheng District, Taipei City, Taiwan
Taipei
TW

Abstract

Background: Precision containment measures guided by dynamic viral shedding are essential for effective control of emerging infectious diseases (EID) though the methodology is complex. The advent of metaverse applications in healthcare introduces a data-driven digital twin model—integrating physical and virtual spaces through augmented reality (AR) and mixed reality (MR)—as a promising solution to enhance disease control strategies.

Objective: To develop an AI-based viral surveillance model for monitoring EID through dynamic viral shedding data and to evaluate the effectiveness of precision contact tracing, isolation, and quarantine schedules within the metaverse.

Methods: A digital twin thread design, consisting of a temporal data pipeline, was proposed to support various twin functions. We began with the physical twin, containing dynamic cycle threshold (Ct) data of viral shedding collected through repeated RT-PCR tests. This data trained parameters governing the infectious disease process using Markov machine learning techniques. In virtual reality (VR), an avatar was created to represent these digital threads, forming the virtual thread cohort. Analytical twins were further developed through AR to overlay virtual data onto the physical twin. Finally, a series of decision twins, implemented in MR, were proposed to evaluate the effectiveness of immersive, precision-guided contact tracing, isolation and quarantine schedules. This metaverse-envisioned surveillance model was illustrated using data from COVID-19 community-acquired outbreaks of the Alpha and Omicron VOCs in Changhua, Taiwan.

Results: Based on physical twin data from 269 COVID-19 cases infected with the Alpha VOC, a virtual thread cohort of 1,000,000 cases were spawned. Analytic twins enabled through AR provided information not only from the physical twin but also captured real-time daily changes in the virtual twin that were unavailable from the physical twin. Accordingly, adjacent transitions such as from normal to Ct?18 pre-symptomatic state occurred more frequently in the analytic twin than the physical twin. Using the first cluster of Alpha VOC infection from the analytic twins to calculate a series of indicators related to the spread of Alpha VOC. A series of decision twins identified optimal days for Ct-guided precision contact tracing and to evaluate the effectiveness of infection control. For individuals with Ct values between 18-25, the optimal days ranged from 7 days of retrospective tracing for reaching 30% effectiveness, 13 days for 60%, and 24 days for reaching 90%. For Omicron VOC, among boosted individuals, 77% were effectively protected after three days of quarantine, and rising to 94% after seven days while those without a booster shoed 39% effectiveness after three days and 76% after seven days.

Conclusions: A Ct-guided, metaverse-envisioned surveillance model demonstrates potential for timely and precise containment of EID. This approach has significant implications for extending 4P medicine into metaverse healthcare, enhancing precision surveillance and containment for EID in the future.

(JMIR Preprints 15/11/2024:68835)
DOI: https://doi.org/10.2196/preprints.68835

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Artificial Intelligence for Precision Viral Surveillance of Emerging Infection Disease (EID): The Metaverse-envisioned Study

Ting-Yu Lin, PhD[1], Amy Ming-Fang Yen, PhD[2], Sam Li-Sheng Chen, PhD[2], Chen-Yang Hsu, PhD[1,3], Yen-Po Yeh, PhD[4], Tony Hsiu-Hsi Chen, PhD[1,*]

[1] Institute of Health Data Analytics and Statistics, College of Public Health, National Taiwan University, Taipei, Taiwan

[2]School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

[3]Dachung Hospital, Miaoli, Taiwan

[4]Changhua County Public Health Bureau, Changhua, Taiwan

***Correspondence and reprint request:**

Professor Tony Hsiu-Hsi Chen

ORCID: 0000-0002-5799-6705

Institute of Health Data Analytics and Statistics, College of Public Health, National Taiwan University

Address: Room 533, No. 17, Xu-Zhou Road, Taipei, 100, Taiwan

TEL: +886-2-33668033

E-mail: ntuhhchen@gmail.com; chenlin@ntu.edu.tw

Word Count: 4652

ABSTRACT

BACKGROUND Precision containment measures guided by dynamic viral shedding are essential for effective control of emerging infectious diseases (EID) though the methodology is complex. The advent of metaverse applications in healthcare introduces a data-driven digital twin model—integrating physical and virtual spaces through augmented reality (AR) and mixed reality (MR)—as a promising solution to enhance disease control strategies.

OBJECTIVE To develop an AI-based viral surveillance model for monitoring EID through dynamic

viral shedding data and to evaluate the effectiveness of precision contact tracing, isolation, and quarantine schedules within the metaverse.

METHODS A digital twin thread design, consisting of a temporal data pipeline, was proposed to support various twin functions. We began with the physical twin, containing dynamic cycle threshold (Ct) data of viral shedding collected through repeated RT-PCR tests. This data trained parameters governing the infectious disease process using Markov machine learning techniques. In virtual reality (VR), an avatar was created to represent these digital threads, forming the virtual thread cohort. Analytical twins were further developed through AR to overlay virtual data onto the physical twin. Finally, a series of decision twins, implemented in MR, were proposed to evaluate the effectiveness of immersive, precision-guided contact tracing, isolation and quarantine schedules. This metaverse-envisioned surveillance model was illustrated using data from COVID-19 community-acquired outbreaks of the Alpha and Omicron VOCs in Changhua, Taiwan.

RESULTS Based on physical twin data from 269 COVID-19 cases infected with the Alpha VOC, a virtual thread cohort of 1,000,000 cases were spawned. Analytic twins enabled through AR provided information not only from the physical twin but also captured real-time daily changes in the virtual twin that were unavailable from the physical twin. Accordingly, adjacent transitions such as from normal to Ct≤18 pre-symptomatic state occurred more frequently in the analytic twin than the physical twin. Using the first cluster of Alpha VOC infection from the analytic twins to calculate a series of indicators related to the spread of Alpha VOC. A series of decision twins identified optimal days for Ct-guided precision contact tracing and to evaluate the effectiveness of infection control. For individuals with Ct values between 18-25, the optimal days ranged from 7 days of retrospective tracing for reaching 30% effectiveness, 13 days for 60%, and 24 days for reaching 90%. For Omicron VOC, among boosted individuals, 77% were effectively protected after three days of quarantine, and rising to 94% after seven days while those without a booster shoed 39% effectiveness after three days and 76% after seven days.

CONCLUSIONS A Ct-guided, metaverse-envisioned surveillance model demonstrates potential for timely and precise containment of EID. This approach has significant implications for extending 4P medicine into metaverse healthcare, enhancing precision surveillance and containment for EID in the future.

INTRODUCTION

Metaverse health care has gained traction in recent years[1,2] with the application of the extended reality (XR) —including virtual reality (VR), augmented reality (AR), and mixed reality (MR) across various medical fields such as radiology, oncology, training, and more to enhance healthcare services[3,4] between the real and virtual worlds. It should be noted that most metaverse applications rely on sensors or wearable smart devices to realize XR [5.6.7]. Few studies, however, have highlighted the application of a data-driven metaverse that merges virtual and physical data to create AR-enabled information, facilitating interaction between decision-makers and policy under different mixed-reality scenarios. Developing such a data-driven digital twin model within the metaverse would be highly beneficial for real-time precision healthcare[8], especially in the face of emerging infectious diseases (EID) while a series of previous research papers have highlighted the link between the digital twin model and personalized heath care[9,10,11,12,13].

The COVID-19 pandemic demonstrated that viral load levels are associated with the transmissibility of SARS-CoV-2, disease severity, hospitalization rates, and the effectiveness of antiviral therapy[14,15]. A metaverse-based surveillance model could be applied to monitor dynamic viral shedding, measured through cycle threshold (Ct) values obtained via reverse transcriptase-polymerase chain reaction (RT-PCR) tests. Such viral surveillance is crucial for planning precision contact tracing and setting scheduled quarantine and isolation strategies.

The advantages of a data-driven metaverse model are twofold. First, Ct dynamics during the pre-symptomatic or asymptomatic phases of infectious diseases are often undetectable with conventional cross-sectional RT-PCR tests[16,17]. Second, even with repeated Ct measurements, the natural infectious course may remain incomplete. A Ct-guided surveillance model, based on a data-driven metaverse, has the potential to address these challenges by capturing the complete trajectory of Ct-based infectious courses up to recovery. This approach leverages both VR and AR within the digital twin technique, the core of the metaverse. Furthermore, the application of MR facilitates the

evaluation of immersive precision containment measures, adjusting Ct-guided contact tracing, isolation, and quarantine schedules.

The objective of this study, therefore, was to develop a data-driven, metaverse-envisioned model with potential applications for precision viral surveillance. Such a model is aimed at designing immersive, Ct-guided contact tracing, quarantine, and isolation strategies to address emerging infectious diseases like COVID-19.

METHODS

Definitions on Artificial Intelligence Viral Surveillance in the Metaverse

Before introducing the methodology of the metaverse-envisioned model, we need a glossary to define domain knowledge on viral surveillance for EID, as well as relevant terminologies in the metaverse and their combined applications.

Susceptible-Pre-symptomatic and Symptomatic Infection-Recovery (representing the pinnacle of physical-virtual co-existence**) Process**: The S-PSI-R process is defined as the trajectory of a natural infectious process used in the following Markov machine learning algorithms to delineate when a susceptible subject is exposed to infection through contact with suspected infectives, progresses through the pre-symptomatic phase, and advances to the symptomatic phase upon the appearance of symptoms, continuing until recovery or death, as shown in **Figure 1**.

Ct-guided States: Ct-guided states are defined as stages of viral shedding upon infection, classified into discrete levels such as low, moderate, and high viral load, based on continuous Ct values obtained via RT-PCR tests, as shown in Figure 1.

AI Viral Surveillance: AI viral surveillance is defined as the monitoring of EID using machine learning algorithms to model the complete trajectory of the S-PSI-R process, overlaid with Ct-guided states, as shown in **Figure 1**.

Metaverse: Following the definition of the Metaverse coined by Stephenson in 1992 [18,19], the metaverse applied in this study consists of the physical space, loaded with real-time data on the Ct-guided S-PSI-R process, and the virtual space, containing virtual digital thread data.

Digital Twin Thread: The digital twin thread is defined as the application of the digital twin concept to the digital thread, with its temporal data pipeline capturing Ct value dynamics. Note that the term "thread," in relation to the temporal data pipeline on Ct-guided states, will be used for various types of twins throughout this study.

Virtual Twin: The virtual twin is defined as the virtual counterpart of the physical twin, representing

the digital thread cohort of the same individual based on the S-PSI-R process, including dynamic Ct data. In this study, the virtual twin focuses on a data-driven digital thread twin framework.

Extended Reality (XR): In the metaverse context, ER encompasses both AR and MR. AR enable real-time interaction between virtual and physical data in viral surveillance, while MR facilitates various immersive precision viral surveillance strategies within the virtual world.

Analytic Twins: Analytic twins are defined by applying the concept of AR to overlay the virtual twin thread onto the physical twin thread within the metaverse. This integration generates complete trajectories of the Ct-loaded S-PSI-R process for the physical twin thread, which would otherwise lack information on Ct dynamics prior to AR application.

Decision Twins: Decision twins are designed to make automated decisions and judgments, identifying the optimal Ct-guided surveillance approach given a series of immersive containment measures under the concept of MR within the metaverse.

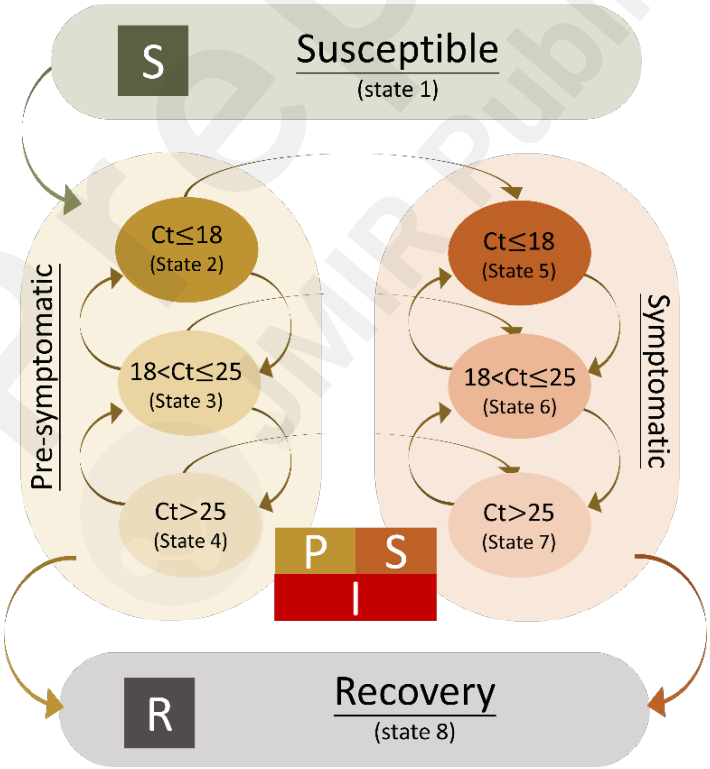

Figure 1. S-PSI-R model

The Metaverse-envisioned Digital Twin Model

Figure 2 shows a five-step procedure for constructing a metaverse-envisioned model based on the digital twin framework for AI viral surveillance of EID, illustrated by SARS-CoV-2 infection.

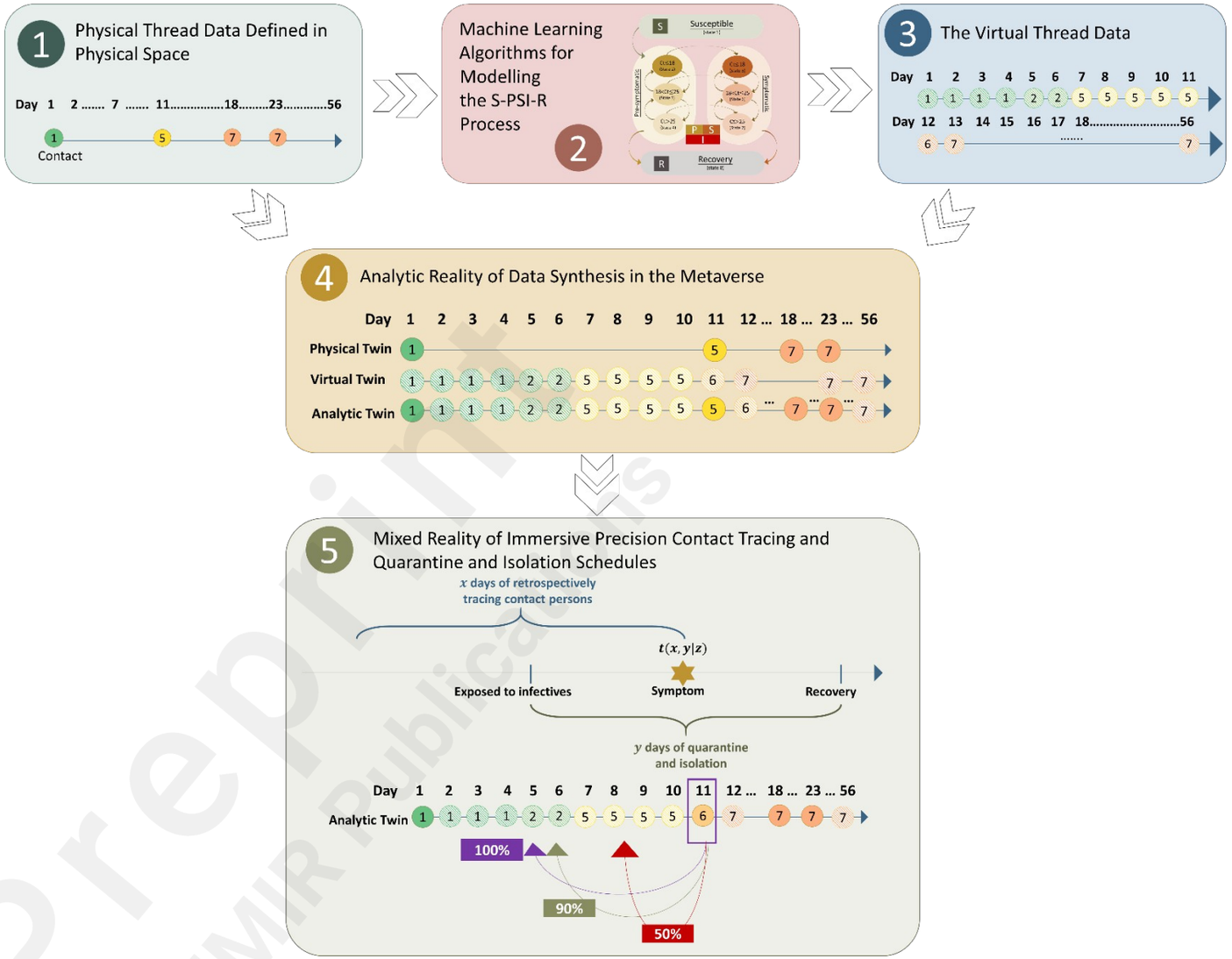

Figure 2. The five-step procedure for the metaverse-envisioned digital twin model

1. The Real-world Data of the Physical Twin Thread

The first step is to establish the physical twin based on real-world data on the dynamics of Ct levels, as shown in the left panel of **Figure 2**. This data pertains to the dates of symptom onset and scheduled RT-PCR tests, which measure viral shedding by Ct value over time—from initial contact with infective individuals, through quarantine and isolation, and, if confirmed as a moderate or severe COVID-19 case, to hospitalization and ultimately recovery. It should be noted that the trajectory of real-world data is only partially observable, particularly regarding the exact pre-symptomatic period and complete Ct dynamics. The unobservable aspects include the pre-symptomatic period in symptomatic cases, where the first positive Ct result is detected only at the time of the initial RT-PCR test, and cases that remain asymptomatic until recovery. Since RT-PCR testing is scheduled during the pre-symptomatic phase, the resulting Ct dynamics may not fully reflect real-time data throughout quarantine and isolation.

2. Machine Learning Algorithms for Modelling the S-PSI-R Process

The second step involves developing probabilistic machine learning algorithms to estimate the transition parameters that govern the complete real-time trajectory of the S-PSI-R process associated with Ct dynamics. These parameters are learned by

training the real-world data using machine learning algorithms based on the Markov process, as previously developed[20].

3. The Virtual Thread Data

The third step is to utilize the developed Markov machine learning algorithms, along with the transition parameters that govern the Ct-guided S-PSI-R process, to perform a series of in-silico simulations, thereby creating the virtual digital thread cohort. The right panel of **Figure 2** displays the simulated virtual world data as the digital thread cohort, showing complete trajectories of the susceptible-pre-symptomatic-symptomatic infection-recovery process with Ct-defined state dynamics over 60 days. This trajectory begins with the time of exposure to infection, progresses through the pre-symptomatic phase and the onset of symptoms, and continues until recovery, overlaid with three Ct levels of viral shedding.

4. Augmented Reality for Creating Analytic Twins:

Under the digital twin thread platform, we apply the concept of AR to merge real-world and virtual-world entities, creating analytic twins that reflect the autonomous evolution of virtual entities in the metaverse. This integration continuously generates Ct-guided information for the physical entity.

Figure 2 also illustrates a symptomatic case in which an analytic twin thread is

completed in the metaverse. Virtual Ct-defined real-time data is transferred to fill in unavailable Ct-defined states within the physical data. Similar procedures are applied to other COVID-19 case patterns, including persistent asymptomatic cases and pre-symptomatic (non-persistent) cases.

5. Mixed Reality of Immersive Precision Contact Tracing and Quarantine and Isolation Schedules

The concept of MR is further applied to assess immersive experiences in precision contact tracing, quarantine, and isolation schedules, based on autonomous real-time twins in the metaverse. The metaverse-envisioned immersive Ct-guided strategies, shown in the bottom box of **Figure 2**, include precise retrospective tracing of suspected contacts and tailored quarantine and isolation durations (in days), guided by the Ct levels of the infected individual.

Data Sources of Physical Entity

Alpha VOC

We illustrate the aforementioned metaverse-envisioned, Ct-guided precision surveillance using two SARS-CoV-2 variants: Alpha VOC and Omicron VOC. Data were obtained from a community-acquired COVID-19 outbreak in Changhua, a central city in Taiwan. The Changhua Health Bureau followed the conventional surveillance system to identify COVID-19-positive patients through intensive contact tracing across various settings. Digital contact tracing during the COVID-19 pandemic was implemented and has been fully described elsewhere[21]. A total of 269 COVID-19 patients from community-acquired Alpha VOC outbreaks between May and July 2021, along with their 7,749 close contacts, were identified. All individuals identified as close contacts of suspected infectives were quarantined and isolated, undergoing scheduled RT-PCR testing to measure Ct values and monitor for symptom onset. Of 269 recorded infections from contact tracing, 24 cases (8.9%) were asymptomatic and 245 cases (91.1%) were symptomatic. Out of these, repeated RT-PCR measurements were performed on 963 cases. In addition to demographic information, physical real-time data were collected, including the date of exposure to infectives, date of symptom onset, duration of quarantine and isolation, dates of admission to and discharge from hospitalization for confirmed COVID-19 cases, and

time to recovery.

Regarding information on viral load level, among the asymptomatic or pre-symptomatic group, a total of 109 cases had specific Ct values: 21.1% with Ct ≤18, 23.9% with Ct between 18-25, and 55.0% with Ct >25. In the symptomatic group, there were 854 cases, with 19.3% having Ct ≤18, 20.7% between 18-25, and 60.0% with Ct >25.

Index Cluster of Alpha VOC outbreak

Of the 269 COVID-19 cases infected with the Alpha VOC, we aimed to demonstrate how to use analytic twins to directly estimate the basic reproductive number (R_0), generation time, and serial interval without relying on conventional statistical models. This demonstration used the first large cluster outbreak in Changhua, which originated from a transmission chain involving a fruit wholesaler. The index case, a fruit wholesaler, contracted COVID-19 after contact with a confirmed case while conducting business in New Taipei City. Upon returning to Changhua, the individual unknowingly spread the virus to family members, friends, and others within their social and business networks, resulting in a widespread cluster infection in the region. This transmission chain led to over 100 confirmed cases and several fatalities.

Omicron VOC

In the original Omicron variant epidemic in the same county, 1,118 cases out of a larger population of 1,288,561 were documented from April 1 to May 16, 2022, marking the early Omicron outbreaks. In contrast to the Alpha VOC period, which followed a zero-COVID-19 policy, providing the optimal quarantine and isolation schedule for identified infectives became the main containment measure to prevent transmission and large-scale epidemic spread.

In addition to information on physical real-time data collected like Alpha VOC, vaccination status for the Omicron VOC were also collected. The vaccination program was the primary difference between the Alpha and Omicron response protocols in Taiwan. The vaccination status of each individual was used to evaluate appropriate precision quarantine and isolation protocols to reduce transmission from identified infective individuals.

Statistical Machine Leaning Algorithms

A nine-state Markov process was utilized to model the conventional S-PSI-R process, augmented by Ct dynamics. This continuous-time Markov process accommodates a series of progressions and regressions across three Ct levels (≤18, 18–25, >25) within and between the pre-symptomatic and symptomatic phases,

ultimately leading to recovery, with a set of transition parameters as outlined in the first column of **Table 2**. The usefulness of this Markov process for assessing time to recovery has been fully described elsewhere[22]. Here, we applied this Markov process to generate virtual real-time data, allowing for a complete projection of the Ct dynamics trajectory—from remaining in the pre-symptomatic phase to the onset of symptoms and eventual recovery—based on physical data, including relevant symptom onset dates, multiple Ct readings, and hospital discharge details.

Ct-guided Nine-State Markov Process

Let $Y(t)$ represent a random process for delineating the Ct-guided model, including susceptible, pre-symptomatic state, and symptomatic status with low ($Ct \leq 18$), middle ($18<Ct\leq 25$) and high ($Ct \geq 25$) Ct-value, respectively. The state space of $Y(t)$, denoted as Ω, comprises the real values 1-9 denoted as $\Omega=\{1,2,3,4,5,6,7,8,9\}$, defined as 1: uninfected or the un-notifiable asymptomatic 2-4: Pre-symptomatic with low ($Ct\leq 18$), medium ($18<Ct\leq 25$) and high ($Ct\geq 25$) Ct-value in sequential order; 5-7: Symptomatic with low ($Ct\leq 18$), medium ($18<Ct\leq 25$) and high ($Ct\geq 25$) Ct-value in sequential order; 8: Recovery with asymptomatic; 9: Recovery after symptom. To further delineate the instantaneous rate of each transition**,** an intensity matrix is denoted as follows

$$Q = \begin{array}{c} \\ Previous \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\lambda_2-\lambda_3 & \lambda_2 & 0 & \lambda_3 & 0 & 0 & 0 & 0 \\ 0 & \lambda_4 & -\lambda_4-\lambda_5-\lambda_6 & \lambda_5 & 0 & \lambda_6 & 0 & 0 & 0 \\ 0 & 0 & \lambda_7 & -\lambda_7-\lambda_8-\lambda_9 & 0 & 0 & \lambda_8 & \lambda_9 & 0 \\ 0 & 0 & 0 & 0 & -\lambda_{10} & \lambda_{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{11} & -\lambda_{11}-\lambda_{12} & \lambda_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{13} & -\lambda_{13}-\lambda_{14} & 0 & \lambda_{14} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{l} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array}$$

Current

where

λ_1: the transition rate from uninfected/non-notifiable to pre-symptomatic with low Ct-value (Ct≤18)

λ_2: the transition rate from pre-symptomatic with low Ct-value (Ct<18) to medium Ct-value (18<Ct≤25)

λ_3: the transition rate from pre-symptomatic with low Ct-value (Ct<18) to symptomatic with low Ct-value (Ct<18)

λ_4: the transition rate from pre-symptomatic with medium Ct-value (18<Ct<=25) to high Ct-value (Ct≥25)

λ_5: the transition rate from pre-symptomatic with medium Ct-value (18<Ct<=25) to low Ct-value (Ct<18)

λ_6: the transition rate from pre-symptomatic with medium Ct-value (18<Ct<=25) to symptomatic with medium Ct-value (18<Ct<=25)

λ_7: the transition rate from pre-symptomatic with high Ct-value (Ct≥25)

to medium Ct-value (18<Ct<=25)

λ_8: the transition rate from pre-symptomatic with high Ct-value (Ct≥25)

to symptomatic with high Ct-value (Ct≥25)

λ_9: the transition rate from pre-symptomatic with high Ct-value (Ct≥25)

to recovery

λ_{10}: the transition rate from symptomatic with low Ct-value (Ct<18)

to medium Ct-value (18<Ct<=25)

λ_{11}: the transition rate from symptomatic with medium Ct-value (18<Ct<=25) to low Ct-value (Ct<18)

λ_{12}: the transition rate from symptomatic with medium Ct-value (18<Ct<=25) to high Ct-value (Ct≥25)

λ_{13}: the transition rate from symptomatic with high Ct-value (Ct≥25) to middle Ct-value (18<Ct<=25)

λ_{14}: the transition rate from symptomatic with high Ct-value (Ct≥25)

to recovery

Based on the intensity matrix Q, we can get the transition probability matrix $P(t)$ through deriving the ordinal differential equation to estimate the risk of each transition at the follow-up time t.

$$\frac{dP_1(t)}{dt} = -P_1(t) \cdot \lambda_1$$

$$\frac{dP_2(t)}{dt} = -P_2(t) \cdot (\lambda_2 + \lambda_3) + P_1(t) \cdot \lambda_1 + P_3(t) \cdot \lambda_4$$

$$\frac{dP_3(t)}{dt} = -P_3(t) \cdot (\lambda_4 + \lambda_5 + \lambda_6) + P_2(t) \cdot \lambda_2 + P_4(t) \cdot \lambda_7$$

$$\frac{dP_4(t)}{dt} = -P_4(t) \cdot (\lambda_7 + \lambda_8 + \lambda_9) + P_3(t) \cdot \lambda_5$$

$$\frac{dP_5(t)}{dt} = -P_5(t) \cdot \lambda_{10} + P_2(t) \cdot \lambda_3 + P_6(t) \cdot \lambda_{11}$$

$$\frac{dP_6(t)}{dt} = -P_6(t) \cdot (\lambda_{11} + \lambda_{12}) + P_3(t) \cdot \lambda_6 + P_5(t) \cdot \lambda_{10} + P_7(t) \cdot \lambda_{13}$$

$$\frac{dP_7(t)}{dt} = -P_7(t) \cdot (\lambda_{13} + \lambda_{14}) + P_4(t) \cdot \lambda_8 + P_6(t) \cdot \lambda_{12}$$

$$\frac{dP_8(t)}{dt} = P_4(t) \cdot \lambda_9$$

$$\frac{dP_9(t)}{dt} = P_7(t) \cdot \lambda_{14}$$

To train all transition parameters, three types of Ct-loaded aggregate data were used. The first type is the first-identified symptomatic cases with the first Ct value at the initial RT-PCR test. The second is the suspected infected cases but lingering in the pre-symptomatic phase during their quarantine and isolation period with consecutive Ct values through a series of RT-PCR tests. The third type is persistent asymptomatic cases " staying in the pre-symptomatic phase all the way until their recovery.

The likelihood function by each transition mode can be derived as follows.

Likelihood function

L1 for transition between two pre-symptomatic states

$$L1(y|\theta)= \prod_{\substack{n_{ij}, \\ for\, i=\{1,2,3,4\},\, j=\{1,2,3,4,8\}}} \prod_{v=1}^{n_{ij}} P_{ij}(t_v) \quad \textbf{(1)}$$

L2 for transition from pre-symptomatic state to symptomatic state

$$L2(y|\theta)= \prod_{\substack{n_{ij}, \\ for\, i=\{1,2,3,4\},\, j=\{5,6,7\}}} \prod_{v=1}^{n_{ij}} P_{i2}(t_{u1})\lambda_3 P_{5j}(t_{u2}) + P_{i3}(t_{u1})\lambda_6 P_{6j}(t_{u2}) + P_{i4}(t_{u1})\lambda_8 P_{7j}(t_{u2}) \quad \textbf{(2)}$$

L3 for transition between two symptomatic states

$$L3(y|\theta)= \prod_{\substack{n_{ij}, \\ for\, i=\{5,6,7\},\, j=\{5,6,7,9\}}} \prod_{v=1}^{n_{ij}} P_{ij}(t_v) \quad \textbf{(3)}$$

Total Likelihood function

$$L(y|\theta)=L1(y|\theta)\,L2(y|\theta)\,L3(y|\theta) \quad \textbf{(4)}$$

Posterior distribution is therefore derived by combining the likelihood function of (4) with the non-informative prior distributions proposed as follows.

Prior distribution

$\pi(\lambda)$ Gamma (0.01, 10000)

Posterior distribution

$$\pi(\lambda|y) \propto L(y|\lambda)\pi(\lambda) \quad \textbf{(5)}$$

Machine Learning with Bayesian Directed Acyclic Graphic (DAG) Model

The Bayesian Monte Carlo Markov Chain (MCMC) method was employed to estimate parameters within the nine-state Markov process, based on the total likelihood derived from the collected real-world data. The detailed sampling algorithms, using random walk Metropolis, are elaborated in the supplementary materials. To ensure robustness and accuracy of this deep Markov learning technique, model efficacy and consistency were validated using a five-fold cross-validation approach, with an 80% training set and 20% test set. Consistency was assessed using the Brier score.

In Silico Simulations for Generating the Virtual Cohort of the Digital Threads

By learning the transition parameters from the Markov model, "In Silico" simulations generate a virtual cohort of 1,000,000 profiles over 60 days as illustrated in the right panel of **Figure 2**, providing dynamic insights into Ct-value fluctuations from pre-symptomatic to recovery stages. Information from virtual and real-world data is expected to be included in the digital thread cohort following the deep Markov machine learning procedures. Digitized threads complete the pre- and symptomatic infection course. This cohort of digital threads can be used to study theoretical infectious epidemiology for illness development.

Patient and public involvement

No patients or members of the public were directly involved in this study.

RESULTS

Transition Parameters Learning from the Physical Twin Thread on Two Mutants

The upper panel of **Figure 3** shows the digital thread of the physical twin for 269 COVID-19 cases, with incomplete temporal pipeline data on Ct-defined states over 60 days, as the majority of Ct levels were unavailable in the physical twin data. Based on the digital thread of the physical twin, the second column of **Table 1** provides a series of transition modes for training the transition parameters, as above.

Following the second step procedure of the digital twin framework in **Figure 2**, a multi-state Markov deep learning approach was used to characterize the dynamics of viral shedding (Ct) across three viral load levels (high, $Ct \leq 18$; medium, $18 < Ct \leq 25$ and low, $Ct > 25$) both before and after symptom onset.

With forward and backward arrows indicating the respective transitions for the Alpha and Omicron VOCs, the estimated instantaneous transition rates (per day) between states, along with their 95% credible intervals, are displayed in **S-Table 1**.

Figure 3. Physical, virtual, and analytic twins

The Results of Virtual Twin Thread

The middle part of **Figure 3** shows the digital thread of the virtual cohort, consisting of 1,000,000 subjects with complete temporal pipeline data on Ct-defined states over 60 days, following in silico simulations based on Markov machine learning algorithms.

The Results of Analytic Twins

The results of analytical twins are diagramed in the lower panel of **Figure 3** after overlying the virtual data on the physical twin. The third column of Table 1 the detailed transition modes of the analytic twins. Compared to the physical twin thread, the analytic twins provided more informative data, as they not only provide physical twin information but also captured real-time daily changes rather than relying on state transitions at fixed intervals. Within the same cohort of 269 individuals, we observed that the transition mode frequencies revealed differences between virtual and physical observations. For example, transitions between adjacent states, such as 1 → 2 and 2 → 3, occurred more frequently in the analytic twin than in the physical twin. Conversely, transitions like 1 → 5 and 1 → 6 were observed more frequently in the physical data. The adjacent transitions in the analytic twin provided more detailed information than the non-adjacent transitions in the physical data. The enhanced

granularity of the analytic twin allows for a more dynamic view of the cohort's infection progression, highlighting differences in state transitions that are crucial for understanding and optimizing surveillance strategies in real-time.

Table 1. Descriptive results of the physical twin in the light of the real-world data by each transition modes

Transition mode	Physical	Analytic
12*	20	212
13	18	17
14	25	5
15	114	29
16	65	2
17	27	0
23*	2	51
24	4	6
25*	3	150
26	1	17
27	10	1
28	2	0
32*	2	13
34*	9	38
35	0	4
36*	0	16
37	11	5
43*	2	8
46	1	0
47*	13	19
48*	22	21
56*	41	188
57	83	40
59	1	0
65*	8	46
67*	119	241
69	1	0
75	1	0
76*	24	64
79*	224	226

*: represent adjacent states

Decision Twins for Optimal Viral Surveillance

By synthesizing virtual and physical twins operated under AR, we can obtain each person's initial viral load at the time of infection and track their daily infection status. This information allows us to determine that if an individual is diagnosed with COVID-19 only upon symptom onset with a Ct value between 18 and 25, the optimal retrospective contact tracing days can be identified with a series of decision twins that were embodied with various immersive strategies given different decision rules of reaching effectiveness of infection control. This means the effectiveness of outbreak containment can be evaluated through interactive immersion of various immersive contact tracing strategies within the framework of mixed reality. For example, 7-day retrospective contact tracing yielded 30% effectiveness, 13-day tracing achieved 60%, and 24-day tracing reached 90%.

A series of decision twins were also formed to evaluate the effectiveness of immersive isolation and quarantine protocols in reducing Omicron VOC transmission, both with and without booster immunization. Among individuals who received a booster, 77% were effectively protected after three days of quarantine, and this figure rose to 94% after 7 days whereas the corresponding effectiveness for those without a booster was 39% after 3 days and 76% after 7 days. For individuals with a booster shot, 90% protection against Omicron VOC was achieved within 5 days, whereas

those without a booster required 11 days for the same level of prevention.

Results of Direct Estimation of R_0, Serial interval, and Generation Time Illustrated by Index Cluster of Alpha VOC

Using the daily reconstructed infection statuses on the first cluster of Alpha VOC (see data resource in the method section) derived from analytic twin, we directly estimated 6.7 of the basic reproductive number R_0 and around 3 of the effective reproduction number R_t.

Additionally, the serial interval—the time between symptom onset in index and secondary cases—was calculated to be about 9 days, while the generation time, or the infection-to-infection interval between index and secondary cases, averaged around 5 days. These quantitative insights, enabled by the analytic twin's continuous, data-driven thread, underscore the advantages of this approach over traditional, assumption-based models, allowing for more precise and timely epidemiological assessments.

DISCUSSION

Data-driven Digital Twin Model in the Metaverse

With the application of extended virtual reality in the metaverse[5,6,7], this study marks the first attempt to implement a dynamic, data-driven digital twin model by overlaying virtual twin data onto physical twin data, creating an AR-formulated synthetic database. Through this database, any scenario derived from the physical twin can interact with various immersive scenarios to determine the optimal strategy —in this case, precision surveillance of EID. This new approach is therefore termed a data-driven digital twin model in the metaverse. The proposed approach was successfully demonstrated using dynamic viral surveillance for an emerging infectious disease, with COVID-19 as an example. Combining this new artificial intelligence technology with viral surveillance for EID within the metaverse framework provides an exemplary application, titled “Artificial Intelligence for Precision Viral Surveillance in the Metaverse,” which is the focus of this study.

Functions of the types of digital twins in the metaverse

One unique characteristic of the digital twin model proposed here is its ability to create different types of twins, each with functions adaptable to extended reality within the metaverse as used in the previous application to neuroscience[23]. These functions are discussed below.

The virtual twin is generated through in silico simulation of parameters learned

from physical twin data using an artificial intelligence model with machine learning algorithms. Since the virtual twin essentially mirrors the physical twin, the artificial intelligence model is highly dependent on the domain knowledge embedded in the physical twin data. In our example of viral surveillance, this process is stochastic, defined by time and states, including disease status and Ct levels, as shown in **Figure 1**. We employed a Markov process with Bayesian acyclic graphical machine learning algorithms to learn these transition parameters. This flexible approach allows for the creation of the virtual twin through in silico simulation using various artificial intelligence and machine learning models. The primary function of the virtual twin is to provide a complete data journey, as demonstrated in our 60-day dynamic Ct results, in contrast to the physical twin, which only contains a partial data journey (**Figure 2**).

Once the virtual twin is established, the analytic twin follows. The analytic twin is central to our data-driven digital twin model in the metaverse for two reasons. First, physical twin data, encoded with real-world data, often lacks complete information on various infection and disease states, particularly those marked by dynamic viral load. For instance, as shown in **Figure 3**, symptomatic cases first identified with symptoms and a Ct value measured by RT-PCR lack observable trajectories for exposure dates and transitions to the pre-symptomatic state with dynamic Ct values of viral shedding. Second, while retrieving this unobservable information is essential, overlaying virtual

twin data onto the physical twin to form the analytic twin through AR is even more valuable. The data encoded in the analytic twin serves as a valuable asset, stored in the cloud for future clinical applications. In our case, the analytic twin further supports a series of decision twins, building a precision decision-support system for viral surveillance in EID control.

Analytic and decision twins pursuant to 4P medicine

More importantly, these two types of digital twins provide an analysis and evaluation portal to fulfill the goals of 4P medicine: personalized, predictive, preventive, and participatory[8]. Building an analytic twin in line with AR principles plays a vital role in personalized viral surveillance. After merging physical data with virtual twin data to create an analytic twin for each individual, the resulting model is expected to be closer to real-world scenarios than a virtual twin alone. In this sense, the analytic twin also serves as a tool for individual prediction, not only for specific outcomes of interest but for the entire disease trajectory. In our example, any individual's trajectory of dynamic viral shedding can be predicted, allowing assessment of viral shedding progression from the date of infection until recovery. Decision twins, based on the principles of mixed reality, are chosen after the interaction between the analytic twin and physical twin, guided by a series of immersive preventive measures. This approach enables the creation of a clinical

decision-support system in the metaverse to evaluate the effectiveness of precision prevention measures, facilitating shared decision-making between patients and clinicians. The decision twin approach offers a valuable solution for studies with multiple interventions beyond traditional randomization designs, particularly in N-to-1 trials. In our viral surveillance example, many immersive precision strategies, such as retrospective contact tracing and personalized quarantine and isolation schedules, can be evaluated. Leveraging decision twins to assess effectiveness in advance allows decision-makers and the public to prepare for EID containment.

The decision twin approach presented here aligns with the preventive and participatory aspects of 4P medicine. Overall, the proposed analytic and decision twins in the metaverse not only enable personalized risk prediction models but also support program evaluators and health decision-makers in evaluating a series of preventive measures beyond randomized designs.

Limitations of data-driven digital twin model in the metaverse

It should be noted that our study differs from current metaverse studies, which often rely on various smart devices—such as hand-mouth displays, haptic devices, HoloLens, and voice control—when operating AR and MR in extended reality within the metaverse. Instead, we focused on a data-driven interface that follows the same principles of extended reality. Although this data-driven digital twin model in the

metaverse is novel, it is currently limited in providing a real-time immersive experience for interactions between reality and virtuality. In our case, this means that real-time artificial intelligence for viral surveillance has not yet been fully achieved. In the future, we hope that real-time operations of the analytic twin with AR and the decision twin with MR can be facilitated through the development of advanced smart devices and controllers, enabling a truly immersive experience.

In conclusion, this study proposed a novel data-driven digital twin model in the metaverse for personalized and precision viral surveillance of emerging infectious diseases. The proposed approach was successfully demonstrated using COVID-19 as a case for viral surveillance. This new approach also has significant implications for extending 4P medicine into metaverse healthcare in the future.

DATA AVAILABILITY

Aggregated data that supports the findings of this study may be available upon request from the corresponding author [H.H.C.].

Individual data supporting the study's findings are not publicly available due to participant privacy.

CODE AVAILABILITY

All statistical analyses were conducted using SAS software (Version 9.4; SAS

Institute Inc. 2012). The code will be shared upon reasonable request by the first author [T.Y.L.].

ACKNOWLEDGEMENTS

The authors thank the financial support by Ministry of Science and Technology, Taiwan (MOST 109-2327-B-002-009; 111-2321-B-002-017; 111-2811-M-002 -061 -MY2; 111-2118-M-002 -004 -MY2; 111-2118-M-038-002; 113-2811-M-002-048).

AUTHOR CONTRIBUTIONS

TYL and HHC conceptualized and design the study. TYL, MFY and HHC were responsible for drafting of the manuscript. TYL, MFY and CYH were responsible for data management and statistical analysis. LSC and YPY were in charge of the data collection and management. YPY and HHC interpreted results and revised the manuscript. All authors agreed the findings and provided input on the revision of the manuscript.

COMPETING INTERESTS

All authors declare no competing interests.

ETHICS STATEMENTS

Ethical approval

The study was approved by the Institutional Review Board (TMU-JIRB: N202007018).

References

1. Garavand, Ali, and Nasim Aslani. "Metaverse phenomenon and its impact on health: A scoping review." Informatics in Medicine Unlocked 32 (2022): 101029.
2. Chengoden, Rajeswari, et al. "Metaverse for healthcare: a survey on potential applications, challenges and future directions." IEEE Access 11 (2023): 12765-12795.
3. Wang, Ge, et al. "Development of metaverse for intelligent healthcare." Nature Machine Intelligence 4.11 (2022): 922-929.
4. Arensmeyer, Jan, et al. "A System for Mixed-Reality Holographic Overlays of Real-Time Rendered 3D-Reconstructed Imaging Using a Video Pass-through Head-Mounted Display—A Pathway to Future Navigation in Chest Wall Surgery." Journal of Clinical Medicine 13.7 (2024): 2080.
5. Tsai, Tsung-Ying, et al. "Merging virtual and physical experiences: extended realities in cardiovascular medicine." European Heart Journal 44.35 (2023): 3311-3322.
6. Rad, Arian Arjomandi, et al. "Virtual and augmented reality in cardiac surgery." Brazilian journal of cardiovascular surgery 37.01 (2022): 123-127.
7. Song, Yeong-Tae, and Jiachen Qin. "Metaverse and personal healthcare." Procedia Computer Science 210 (2022): 189-197.
8. Venkatesh, Kaushik P., Gabriel Brito, and Maged N. Kamel Boulos. "Health

digital twins in life science and health care innovation." Annual Review of Pharmacology and Toxicology 64.1 (2024): 159-170.

9. Tang, Chenyu, et al. "A roadmap for the development of human body digital twins." Nature Reviews Electrical Engineering 1.3 (2024): 199-207.
10. Elkefi, Safa, and Onur Asan. "Digital twins for managing health care systems: rapid literature review." Journal of medical Internet research 24.8 (2022): e37641.
11. Milne-Ives, Madison, et al. "Life course digital twins–intelligent monitoring for early and continuous intervention and prevention (LifeTIME): Proposal for a retrospective cohort study." JMIR research protocols 11.5 (2022): e35738.
12. Vallée, Alexandre. "Envisioning the Future of Personalized Medicine: Role and Realities of Digital Twins." Journal of Medical Internet Research 26 (2024): e50204.
13. Katsoulakis, Evangelia, et al. "Digital twins for health: a scoping review." NPJ Digital Medicine 7.1 (2024): 77.
14. Chia, Po Ying, et al. "Virological and serological kinetics of SARS-CoV-2 Delta variant vaccine breakthrough infections: a multicentre cohort study." Clinical microbiology and infection 28.4 (2022): 612-e1.
15. Weinreich, David M., et al. "REGEN-COV antibody combination and outcomes in outpatients with Covid-19." New England Journal of Medicine 385.23 (2021): e81.

16. He, Xi et al. “Temporal dynamics in viral shedding and transmissibility of COVID-19.” Nature medicine vol. 26,5 (2020): 672-675. doi:10.1038/s41591-020-0869-5
17. Hay, James A., et al. "Estimating epidemiologic dynamics from cross-sectional viral load distributions." Science 373.6552 (2021): eabh0635.
18. Stephenson, Neal. Snow crash. Penguin UK, 1994.
19. Kayakoku, Hakan. "History and development of virtual worlds and metaverse." Metaverse: Technologies, opportunities and threats. Singapore: Springer Nature Singapore, 2023. 19-30.
20. Lin, Ting-Yu, et al. "Kinetics of Viral Shedding for Outbreak Surveillance of Emerging Infectious Diseases: Modeling Approach to SARS-CoV-2 Alpha and Omicron Infection." JMIR Public Health and Surveillance 10 (2024): e54861.
21. Wang, Hsiao-Chi, et al. "Community-based digital contact tracing of emerging infectious diseases: design and implementation study with empirical COVID-19 cases." Journal of Medical Internet Research 25 (2023): e47219.
22. Yeh, Yen-Po, et al. "Quantifying the effects of viral load on multistate COVID-19 infection and the progression of the Alpha and Omicron VOCs: a Bayesian competing Markov exponential regression model." Stochastic Environmental Research and Risk Assessment 38.2 (2024): 621-636.
23. Zhang, Kang, et al. "Concepts and applications of digital twins in healthcare and

medicine." Patterns 5.8 (2024).

Supplementary Files

Multimedia Appendixes

Untitled.
URL: http://asset.jmir.pub/assets/c0a6bf0996a52895f97847d3a82dfc37.docx