Biosafety and Health 6 (2024) 225–234

Contents lists available at ScienceDirect

Biosafety and Health

journal homepage: www.elsevier.com/locate/bsheal

Original Research

Automated robot and artificial intelligence-powered wastewater surveillance for proactive mpox outbreak prediction

Guanyong Ou [a,1], Yuxuan Tang [b,1], Jiexiang Liu [a,c], Yabin Hao [d], Zhi Chen [b], Ting Huang [a], Shaxi Li [a], Shiyu Niu [a], Yun Peng [a], Jiaqi Feng [a], Hongwei Tu [e], Yang Yang [a,*], Han Zhang [b,*], Yingxia Liu [a,*]

[a] *Shenzhen Key Laboratory of Pathogen and Immunity, National Clinical Research Center for Infectious Disease, State Key Discipline of Infectious Disease, the Third People's Hospital of Shenzhen, Second Hospital Affiliated to Southern University of Science and Technology, Shenzhen 518112, China*
[b] *International Collaborative Laboratory of 2D, Materials for Optoelectronics Science and Technology of Ministry of Education, Institute of Microscale Optoelectronics, College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China*
[c] *Shenzhen Clinical Medical College, Guangzhou University of Chinese Medicine, Shenzhen 518116, China*
[d] *Shenzhen Metasensing Tech Limited Company, Shenzhen 518000, China*
[e] *Guangdong Provincial Centre for Diseases Control and Prevention, Guangzhou 511430, China*

ARTICLE INFO

Article history:
Received 18 March 2024
Revised 8 July 2024
Accepted 8 July 2024
Available online 10 July 2024

Keywords:
Automated online sewage virus enrichment robot
Artificial intelligence (AI) model
Early warning system
Mpox
Monkeypox virus (MPXV)

ABSTRACT

In the wake of the largest-ever recorded outbreak of mpox in terms of magnitude and geographical spread in human history since May 2022, we innovatively developed an automated online sewage virus enrichment and concentration robot for disease tracking. Coupled with an artificial intelligence (AI) model, our research aims to estimate mpox cases based on the concentration of the monkeypox virus (MPXV) in wastewater. Our research has revealed a compelling link between the levels of MPXV in wastewater and the number of clinically confirmed mpox infections, a finding that is reinforced by the ability of our AI prediction model to forecast cases with remarkable precision, capturing 87 % of the data's variability. However, it is worth noting that this high precision in predictions may be related to the relatively high frequency of data acquisition and the relatively non-mobile isolated environment of the hospital itself. In conclusion, this study represents a significant step forward in our ability to track and respond to mpox outbreaks. It has the potential to revolutionize public health surveillance by utilizing innovative technologies for disease surveillance and prediction.

1. Introduction

Human monkeypox (mpox), a zoonotic disease of infectious etiology, is precipitated by the monkeypox virus (MPXV) [1,2]. This pathogen is a large, enveloped virus, endowed with a double-stranded deoxyribonucleic acid (DNA) genome of approximately 200 kilo base pairs in length. It is taxonomically situated within the *Orthopoxvirus* genus, under the expansive umbrella of the *Poxviridae* family [3]. The initial identification of this virus was in 1958 in Denmark, where an infection was detected in monkeys imported from Singapore, subsequently triggering an outbreak among experimental macaques [4]. The definitive animal reservoir remains enigmatic, but the prevailing hypothesis suggests that MPXV is perpetuated within small mammalian species indigenous to West and Central Africa [5,6]. The virus is endemic to these regions, with sporadic human infections and restrained transmission [7]. Isolated instances of mpox cases have been reported beyond these endemic zones, predominantly associated with travel, exposure in healthcare settings, or contact with imported rodents [8].

As a disease with significant implications for human health, mpox was first identified in Central Africa (Democratic Republic of the Congo) in 1970 [9,10]. Despite its etymological roots in monkeys, it is conjectured that the virus initially infected rodents and small mammals [11]. Currently, the primary pathway for human-to-human transmission is believed to be through direct skin-to-skin contact [12]. Most of the cases reported so far have predominantly involved men who are sexually active with other men, often referred to as men who have sex with men (MSM) [13]. There is mounting evidence that suggests sexual activity, especially when it involves skin-to-skin contact or

---

* Corresponding authors: Shenzhen Key Laboratory of Pathogen and Immunity, National Clinical Research Center for Infectious Disease, State Key Discipline of Infectious Disease, the Third People's Hospital of Shenzhen, Second Hospital Affiliated to Southern University of Science and Technology, No. 29, Bulan Road, Longgang District, Shenzhen 518112, China (Y. Liu, Y. Yang); International Collaborative Laboratory of 2D, Materials for Optoelectronics Science and Technology of Ministry of Education, Institute of Microscale Optoelectronics, College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China (H. Zhang).

E-mail addresses: young@mail.sustech.edu.cn (Y. Yang), hzhang@szu.edu.cn (H. Zhang), szsy_lyx@szsy.sustech.edu.cn (Y. Liu).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bsheal.2024.07.002

HIGHLIGHTS

Scientific question

The accurate and early prediction of monkeypox outbreaks is essential for effective public health response. We aim to develop an integrated system using automated robotics and AI-powered wastewater surveillance to enhance the proactive detection and management of monkeypox outbreaks.

Evidence before this study

The largest-ever recorded outbreak of monkeypox since May 2022 highlighted the need for innovative surveillance methods. Traditional disease tracking methods have limitations, especially in early detection and prediction of outbreaks. There was a lack of integration between automated virus enrichment methods and AI models for predicting disease outbreaks based on environmental samples.

New findings

This study introduced an automated online sewage virus enrichment and concentration robot coupled with an AI model, demonstrating a strong correlation between the concentration of monkeypox virus (MPXV) in wastewater and clinically confirmed cases. The AI model predicted monkeypox cases with high precision, capturing 87% of the data's variability.

Significance of the study

This research represents a significant advancement in public health surveillance for monkeypox, offering a novel, efficient, and proactive approach to predicting outbreaks. The integration of automated virus enrichment robots with AI models for analyzing wastewater samples could revolutionize the tracking and management of infectious diseases, potentially leading to earlier interventions and better outbreak control.

the sharing of bodily fluids, serves as a likely conduit for transmission. The virus gains entry into the human system via skin lesions, the respiratory tract, or mucous membranes [14,15]. In a host, a typical MPXV infection peaks within 2–4 weeks, potentially exhibiting symptoms such as fever, nausea, and the development of papules on the face, groin, or other regions of the body [16,17]. Two distinct clades of MPXV are recognized: clades I (the Congo Basin clades) and clades II (the West African clades) [18–20]. In May 2022, a hitherto unprecedented large-scale outbreak of human mpox was instigated by the clades IIb of MPXV, which rapidly disseminated across multiple nations [21]. In response, the World Health Organization (WHO) declared this outbreak as an international public health emergency on July 23, 2022 [22].

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has underscored the value of wastewater-based epidemiology (WBE) in providing rapid and reliable data on community-level viral transmission and mutation patterns [23–25]. This approach offers distinct advantages over traditional clinical testing methodologies, particularly in scenarios where societal stigma or discrimination associated with specific infections (such as AIDS patients or mpox patients) may deter at-risk populations from seeking medical attention [26,27]. Through the analysis of anonymous pooled samples, WBE allows for the visualization of infection dynamics within a community without compromising individual privacy. Empirical studies have substantiated the efficacy of WBE in monitoring the spread of various viruses, including Poliovirus and SARS-CoV-2 [28–31]. This approach enables the tracking of emergent variants and provides an accurate depiction of the infection landscape at the population level [24,32,33]. Analogously, the detection of MPXV DNA in domestic wastewater is posited as a valuable early-warning tool for tracking the spread of mpox within communities, serving as a beneficial adjunct to conventional clinical surveillance. MPXV DNA fragments have been identified in a wide array of biological materials, including feces, burst blisters, desquamated rash, swabs, sputum, semen, and urine, all of which can potentially enter the domestic sewage system [34–36]. Moreover, MPXV can infect and circulate among household pets and small mammals residing in sewer systems, providing a milieu for potential viral mutation and subsequent zoonotic transmission to humans [37,38]. Consequently, the wastewater-based monitoring of MPXV in sewage systems could emerge as a significant public health resource, aiding in the timely identification and containment of mpox outbreaks.

In the present study, we innovatively developed an automated online sewage virus enrichment and concentration robot for disease tracking, coupled with an artificial intelligence (AI) model, to estimate the mpox cases based on the MPXV concentration in wastewater. Our research observed a significant correlation between the concentration of MPXV in wastewater and the clinically reported cases of mpox, further underscored by the performance of the AI predictive model. This study highlights the potential of applying online sewage virus enrichment and concentration robot for MPXV surveillance in wastewater, marking a significant advancement in disease outbreak tracking and response, with the potential to revolutionize public health surveillance.

2. Materials and methods

2.1. Sample collection and concentration

The collection and enrichment of wastewater samples were carried out through the collaboration between the Shenzhen Third People's Hospital and Shenzhen Metasensing Technology Co., Ltd. An online sewage virus enrichment and concentration robot was developed as part of this collaboration. This robot has two operating modes: field mode and laboratory mode. In field mode, the robot can automatically collect sewage samples and perform subsequent experimental operations according to the set program. In laboratory mode, laboratory personnel place the collected sewage samples on the robot's sample rack for the robot to perform subsequent experimental operations automatically. In this study, the polymerase chain reaction (PCR) results of six different samples were first compared between the robot's laboratory mode and full manual experimental operations to test the performance of the machine. Then the field mode was carried out at the hospital's wastewater treatment plant for sample collection and processing. The wastewater sample collection site was the inlet of the hospital's wastewater treatment system, where the water pump automatically pumped water based on scheduled intervals before the sewage passed through the grid in time scale of three months. Pumping commenced at 9:00 am daily and recurred every 6 h. Simultaneously, clinically confirmed data on patients infected with mpox in the corresponding time period were collected.

The enrichment and concentration method employed was based on the aluminum hydroxide adsorption-precipitation technique outlined in the industry standard WS/T 799-2022 published by the National Health Commission of China, both for the robot program and the manual experiments. This method had previously been utilized for the enrichment and concentration of the SARS-CoV-2 in wastewater [29]. In essence, it involved the use of positively charged aluminum hydroxide particles generated through the hydrolysis of aluminum chloride to adsorb negatively charged viral particles in the sewage. Subsequently, centrifugation was employed to separate the viral particles enveloped by aluminum hydroxide. The virus was then released from the aluminum hydroxide colloid through the chelating effect of ethylenediaminetetraacetic acid (EDTA) disodium, resulting in the concentrated virus solution.

2.2. Nucleic acid extraction and detection

The nucleic acid extraction procedure utilized the nucleic acid extraction kit from Guangzhou Hybribio Technology Co., Ltd (C221111A). The extraction steps involved adding 300 μL of the sample into the sample slot containing lysis buffer in the extraction kit. Subsequently, the extraction kit was placed into a fully automated nucleic acid extraction instrument. Following the instructions, the instrument was set for a 10-minute magnetic bead lysis, a 1-minute first wash, a 1-minute second wash, and a 5-minute elution. Throughout this process, magnetic rods with magnetic properties automatically transferred total nucleic acid substances. Finally, 60 μL of the sample was obtained in the elution slot. The nucleic acid detection was performed using the fluorescence quantitative PCR method. The PCR reagents utilized included custom-designed primers (Table S1) from Sangon Biotech (Shanghai) Co., Ltd and TransScript® Probe One-Step qRT-PCR SuperMix (Code#AQ221-01) from TransGen Biotech Co., Ltd.

2.3. Machine learning models

In the selection of machine learning models, the problem of low accuracy or instability of predictions from a single model was taken into consideration. Therefore, the stacking ensemble learning was identified as the appropriate choice for this study, as it has multiple advantages that make it popular in practical applications, and it was applied in a coronavirus disease 2019 (COVID-19) wastewater prediction case in Brazil with forecasting error less than 7 % [39]. Stacking allows for combining multiple different base learners and leveraging their respective strengths to enhance the performance and robustness of the model. In addition, it can reduce the risk of overfitting as it better balances the predictions from each base model, minimizing the impact of incorrect decisions from a single model on the overall model.

The algorithmic software utilized in this study is Waikato Environment for Knowledge Analysis (WEKA). Prior to regression testing, our data underwent preprocessing. The processing method involved using the Standardize filter for normalization. Its principle lies in applying linear transformations to each numeric attribute, converting attribute values into a form with a mean of 0 and a standard deviation of 1. Specifically, it employs the following formula for transforming each attribute.

2.3.1. Data preparation

Initially, the training dataset was divided into multiple subsets, each used to train a base learner. Subsequently, each base learner underwent individual training. In this scenario, the models selected were support vector machine regression (SVMR), random forest (RF), and light gradient-boosting machine (LightGBM) as base learners, representing different types of learning algorithms.

2.3.2. Training of base learners

SVMR is a powerful model for regression tasks, as it attempts to find a hyperplane to minimize the error between actual values and predictions. On the other hand, RF is an ensemble learning method comprising multiple decision trees, where each tree is trained on a random subset of the dataset and their results are combined to provide robust predictions. Additionally, LightGBM is a gradient-boosting framework known for its strong performance in gradient-boosting trees; it trains multiple decision trees using gradient-boosting to enhance predictive performance gradually. Together, these models offer varied and complementary approaches for achieving accurate and reliable predictions in regression tasks.

2.3.3. Prediction by base learners

After training, the base learners were employed to predict both the training set and potential test sets, generating predictions for each base learner.

2.3.4. Training of meta learners

Subsequently, the predictions from base learners were used as inputs to train a meta learner. Additive regression was chosen as the meta learner. The task of this meta learner was to learn how to combine predictions from various base learners to generate the final integrated prediction.

2.3.5. Integrated prediction

Once the meta learner was trained, it was utilized to generate the ultimate integrated predictions. These integrated prediction results are typically employed in real-world applications for decision-making and forecasting purposes.

2.3.6. Model evaluation

Utilizing mean absolute percentage error (MAPE), R-squared (R^2) Score, root mean square error (RMSE) and mean absolute error (MAE) to assess the predictive accuracy of the model.

2.4. Statistical analysis

Statistical analyses were performed in Microsoft Office Excel (2021), SPSS 25.0 and R v4.0.2. Normality tests were conducted to ascertain the parametric or non-parametric nature of the data. A statistical significance level of $P < 0.05$ was considered for all tests.

3. Results

3.1. The automated online sewage concentration method

In this study, we employed an automated online sewage concentration method based on the aluminum hydroxide adsorption precipitation method. This automated enrichment technique involved deploying a robotic system at the sewage inlet of the hospital's wastewater treatment plant (Fig. S1). The sewage was pumped into a sewage basin, where a robot, using a robotic arm, operated a pipetting gun and a centrifuge tube. The sewage from the basin was transferred to a 50 mL centrifuge tube containing aluminum chloride coagulant. Subsequently, the robotic system transferred the centrifuge tube to a pH adjustment module. Using a digital pH meter and a precision syringe pump, hydrochloric acid and sodium hydroxide were added to adjust the pH of the solution to 6.0. The robotic arm then agitated the solution with the pipetting gun for a specific duration before transferring the centrifuge tube to a position-controlled centrifuge machine equipped with temperature control functionality. The technical roadmap of this automated equipment is depicted in Fig. 1.

Following centrifugation, the samples in the centrifuge tube separated into a supernatant and a flocculent precipitate enclosing viral particles. The robotic system removed the centrifuge tube from the centrifuge machine, opened the lid, removed the supernatant, and added powdered disodium EDTA to the flocculent precipitate for thorough mixing. The mixture was subsequently transferred to a metal bath heating module and heated for 10 min. Finally, the concentrated samples were placed into a refrigerated storage module, awaiting collection by laboratory personnel for subsequent nucleic acid extraction and PCR testing.

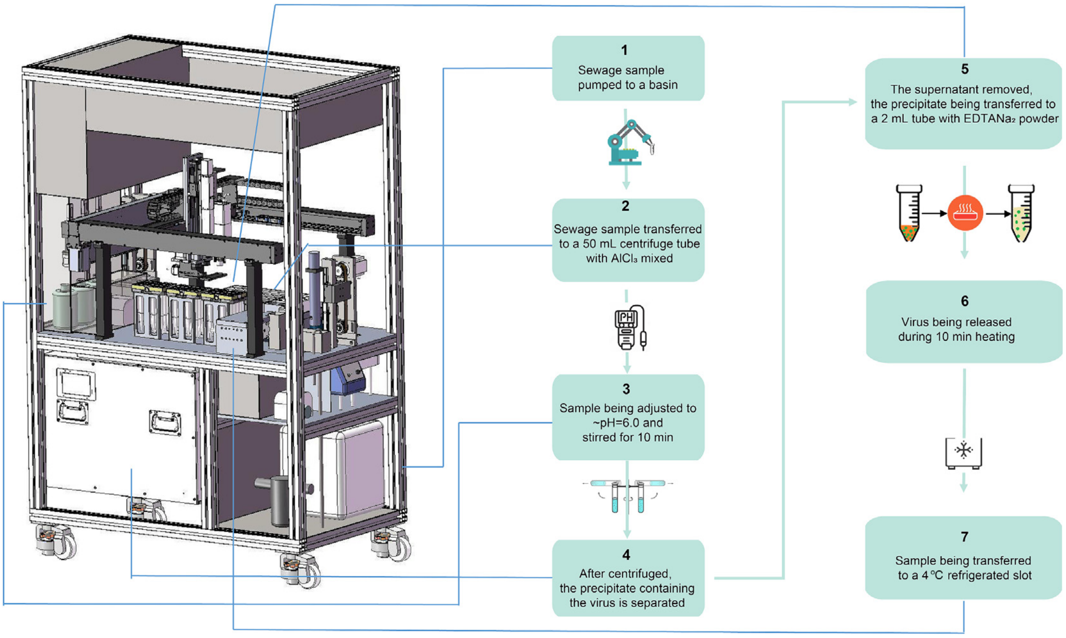

Fig. 1. The technique scheme of the online automated robot.

Table 1
PCR detection results of MPXV in sewage samples from six different sampling points using automated and manual experimental methods.

Sampling point	PCR results *via* automated concentration method						PCR results *via* manual concentration method					
	1	2	3	Average	CV (%)	R_r (%)	1	2	3	Average	CV (%)	R_r (%)
1	38.49	39.55	39.27	39.10	1.1	25.3	39.14	39.39	39.02	39.18	0.4	24.1
2	39.18	38.95	39.75	39.29	0.9	22.4	38.78	39.31	38.88	38.99	0.6	27.2
3	39.56	37.87	39.20	38.88	1.9	29.2	38.29	39.55	38.88	38.91	1.3	28.7
4	38.86	38.72	39.44	39.01	0.8	26.9	38.65	39.08	39.49	39.07	0.8	25.8
5	38.71	39.35	38.34	38.80	1.1	30.7	39.28	39.32	39.12	39.24	0.2	23.2
6	38.85	39.54	38.92	39.10	0.8	25.3	38.61	39.12	39.25	38.99	0.7	27.1

Abbreviations: R_r, recovery rate; CV, coefficient of variation; PCR, polymerase chain reaction.

Throughout the process, the robotic system autonomously handled tasks such as opening and closing centrifuge tube lids and verified the accuracy of lid closures, mimicking human experiments entirely. The refrigerated storage module had the capacity to store eight samples. In this study, we collected samples four times a day, and laboratory personnel collected the samples within 48 h for PCR testing in the laboratory.

Prior to commencing the online monitoring of MPXV in the hospital sewage station, the performance of this automated concentration method was rigorously evaluated. Extensive testing was conducted using blank samples to ensure the seamless operation of the equipment over a continuous 7-day period without any malfunctions. Subsequently, sewage samples, sourced from areas where no cases of mpox had been reported in the past 7 days, were marked as “blank sewage samples” and subjected to quantitative addition of MPXV using this automated system.

In parallel, manual experiments were conducted synchronously during the experimental process for comparative analysis against the automated method. A total of six blank sewage samples were selected from different sampling points with similar wastewater treatment plant (WWTP) facilities as the hospital, then mpox *pseudovirus* was artificially added to each sample at a concentration of 6,000 copies/mL. Each sample was divided into parallel comparative portions, with one portion being processed through the machine's laboratory mode and the other through manual experimental operations. The samples were then processed for virus concentration, purification and subsequent detection experiments. Deviations observed in the three experiments were meticulously compared, alongside a detailed analysis of discrepancies between the results obtained through the automated method and manual experimentation.

The outcomes, as illustrated in Table 1, demonstrated a remarkable consistency between the results obtained through the automated method and those derived from manual experiments. In fact, the automated approach exhibited a slightly superior level of repeatability in comparison to manual operations, underscoring its efficacy and precision in virus concentration processes. Simultaneously, the recovery rates of MPXV in sewage samples were calculated using both the automated method and manual experiments. The recovery rate R_r was calculated according to Equation (1).

$$R_r = \frac{N_{aft}}{N_{org}} = \frac{C_{aft}\hat{A}\cdot V_{aft}}{30,000\,\text{copies}} = \frac{C_{(PCRct)}\hat{A}\cdot 1.2\,\text{mL}}{30,000\,\text{copies}} \tag{1}$$

N_{aft} refers to the virus copy numbers recovered after concentration and the N_{org} refers to the virus copy numbers in the original water samples. C_{aft} is the virus concentration of the enrichment solution while the V_{aft} is the volume of it which is 1.2 mL both in manual and automated methods. $C_{(PCRct)}$ is the PCR sample concentration determined based on the standard curve function generated through PCR testing

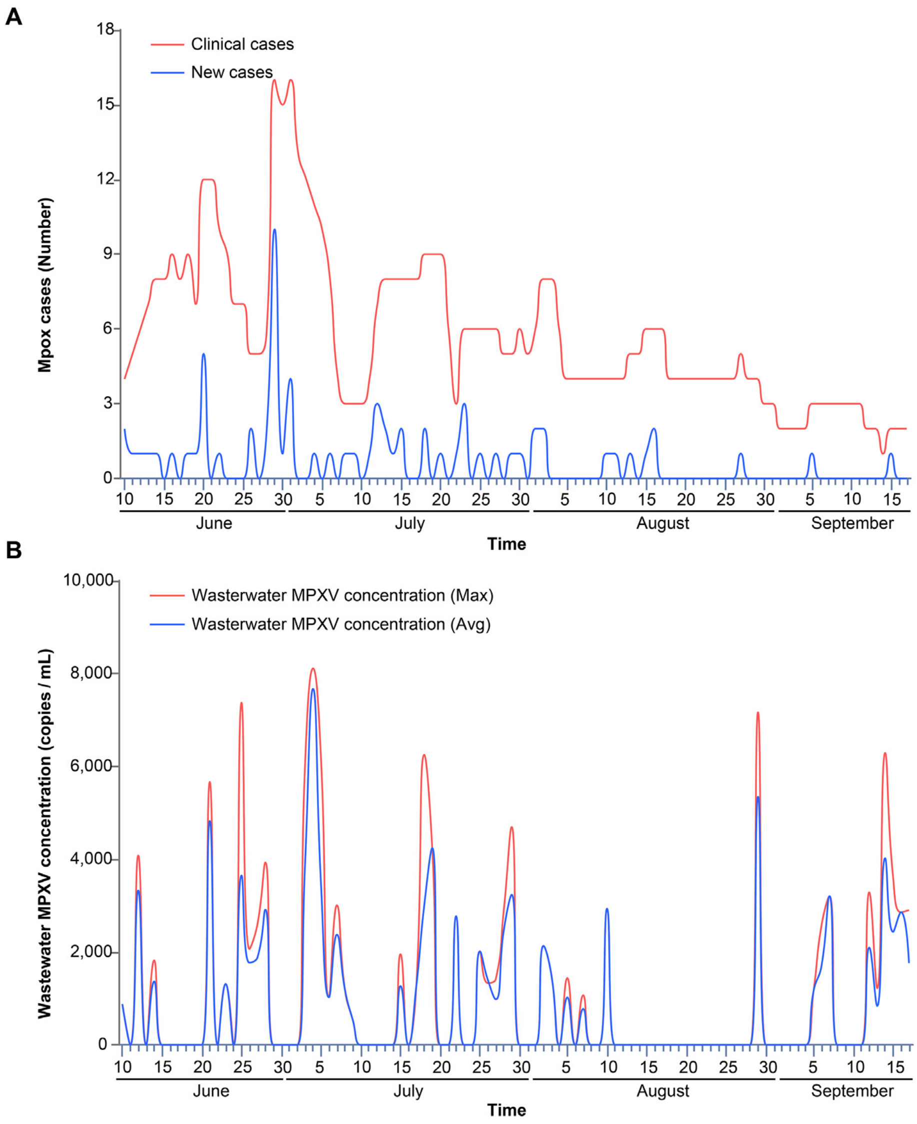

Fig. 2. Daily surveillance dynamic data of mpox cases and MPXV. A) Daily dynamics in hospital admissions for patients with mpox. B) Daily dynamic variation of MPXV concentration in wastewater. Abbreviations: MPXV, monkeypox virus; Max, maximum; Avg, average.

with quantitative *pseudovirus*. It can be converted into sample concentration. In this study, the PCR sample concentration refers to the concentration before nucleic acid extraction, thus $C_{(PCRct)}$ encompasses the nucleic acid extraction recovery rate. The results of 6 sampling points indicated a close similarity in the recovery rates of MPXV between the two methods, ranging from 22 % to 30 %.

3.2. MPXV in hospital wastewater and its correlation with reported cases

From June 10 to September 17, 2023, we conducted a comprehensive collection of wastewater samples from the sewage treatment system at the Shenzhen Third People's Hospital. Initiated at 9:00 am daily, samples were systematically gathered every six hours using an online sewage virus enrichment and concentration robot. At the start of our study on June 10, 2023, the hospital recorded 4 clinically confirmed cases and 2 new cases of mpox on that day. All confirmed and new cases of mpox patients were quarantined in the hospital's isolation ward area. Throughout the study period, we observed a maximum of 16 daily inpatient cases of mpox (recorded on June 29 and July 1, 2023), with a total of 10 days exceeding 10 inpatient cases. Additionally, the highest number of new daily cases reached 10 on June 29, 2023 (Fig. 2A).

During the whole wastewater sample collection process, which spanned 100 days, a positive signal for the MPXV in the wastewater was detected on 41 days. Out of the 283 wastewater samples collected, 153 tested positive for the presence of the MPXV, resulting in a positive detection rate of 54.06 % (Fig. 2B). Notably, during the 25-day period from August 11 to September 1, only one day (August 29) registered a positive signal for the MPXV in the wastewater, with a concentration of approximately 2,161 copies/mL. Concurrently, within this same period (August 11 to September 1), a total of six new mpox cases were reported, with a maximum of two new cases per day.

In an endeavor to explore the association between the concentration of MPXV in wastewater and clinical mpox cases, we initially aligned the data of mpox patients to correspond with the days a positive signal for the MPXV was detected in wastewater samples (Fig. 3). Our findings revealed that out of 41 days where a positive signal for the MPXV was detected in wastewater, new cases of mpox were observed on 19 of these days.

For a more comprehensive investigation of the correlation between the concentration of MPXV in wastewater and the reported clinical

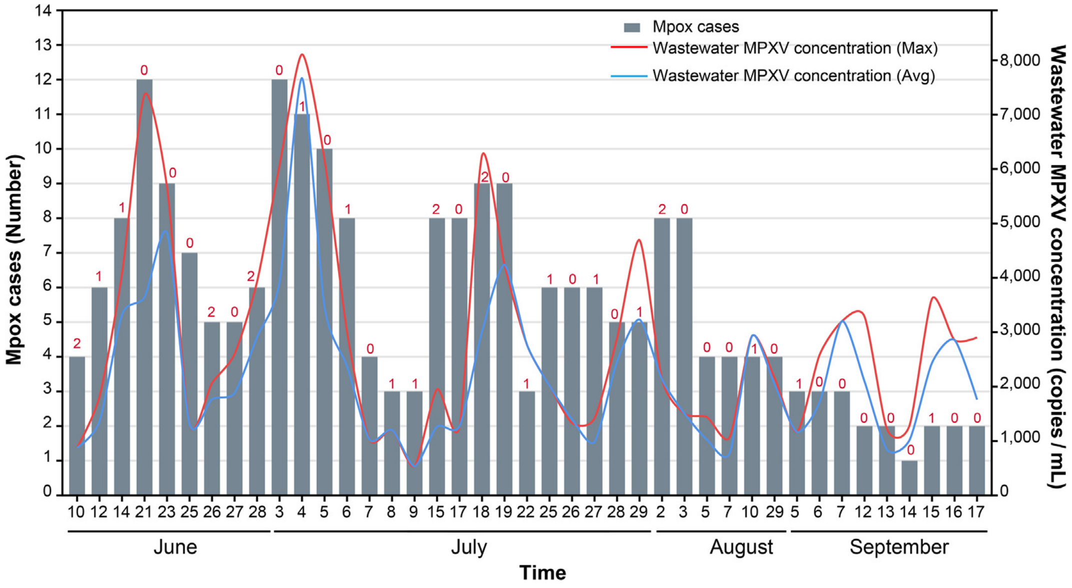

Fig. 3. The co-distribution of MPXV concentrations in wastewater and confirmed clinical cases of mpox patients. The histogram displays the frequency of mpox cases, while the red line represents the maximum wastewater MPXV concentration, and the blue line signifies the average wastewater MPXV concentration. The red numeric data indicate the daily count of new mpox cases. Abbreviations: MPXV, monkeypox virus; Max, maximum; Avg, average.

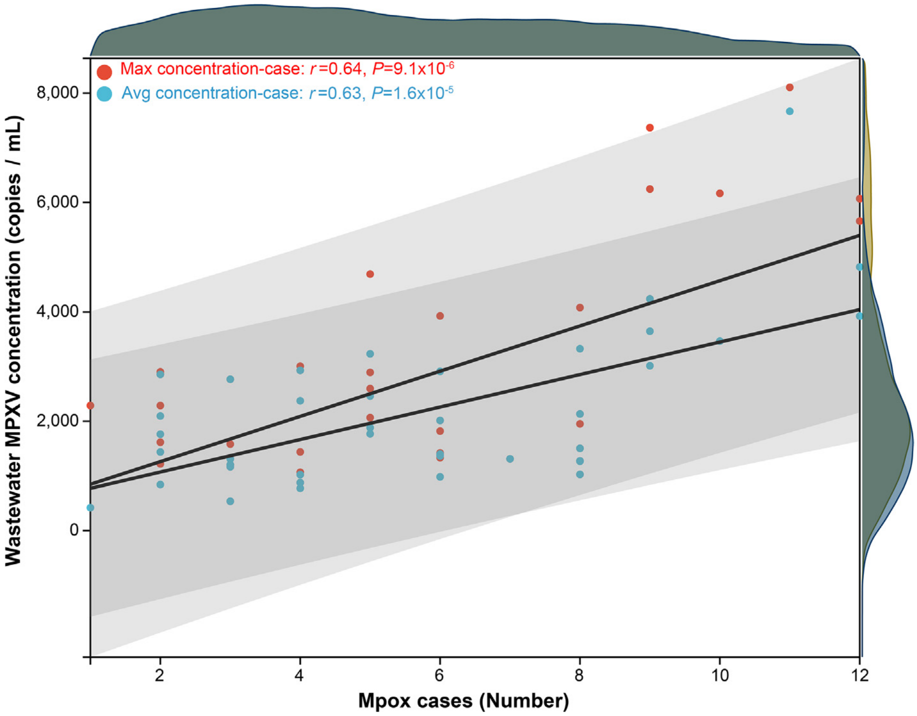

Fig. 4. Correlation between wastewater MPXV concentration and clinical mpox cases. Abbreviations: MPXV, monkeypox virus; Max, maximum; Avg, average.

cases, we employed Pearson's correlation analysis, in addition to considering the normal distribution of the data. A significant positive correlation was discerned between the concentration of MPXV observed in hospital wastewater samples and the presence of clinically reported cases (Fig. 4). Specifically, the peak concentration of MPXV in wastewater bore a correlation coefficient of $r = 0.82$, $P < 0.001$ with the number of mpox cases, while the mean concentration of MPXV in wastewater exhibited a correlation coefficient of $r = 0.73$, $P < 0.001$ with the number of mpox cases, denoting a notably strong correlation.

3.3. Forecasting case numbers using sewage virus concentration via machine learning

In order to accurately forecast the number of mpox infections through the information obtained from viruses in wastewater, we constructed a predictive model using machine learning models. Arash Zamyadi's research team has collected WBE data from 108 sampling points in five regions: Scotland, Catalonia, Ohio, the Netherlands, and Switzerland and utilized a machine learning model, specifically the Random Forest algorithm, for epidemic prediction [40]. We collected data from the influent water quality at hospital sewage treatment stations, virus concentrations in sewage, and the number of confirmed cases on the same day, totaling 101 sets of such data (of which viruses in sewage were detected for 42 days, yielding 42 valid datasets). Subsequently, 70 % of the data were used as the training set, and 30 % as the test set. We employed a stacking fusion model to predict the number of confirmed cases. In this study, three hydrological and water quality parameters were utilized, namely chemical oxygen demand (COD), pH of the water, and flow rate, along with the concentration data of MPXV in sewage, as input variables to predict the number of confirmed disease cases. The accuracy of predictions and model performance were evaluated mainly according to the MAPE value. The framework of stacking models was shown in Fig. 5.

MAPE is a metric employed to assess the performance of regression models, especially when measuring the relative error between model predictions and actual values is necessary. It calculates the average absolute percentage error, representing the extent of the model's error in percentage terms.

The formula for calculating MAPE is as follows (2):

$$\mathrm{MAPE} = \frac{1}{n}\sum_{1}^{n}\frac{|(Forecast_i - Actual_i)|}{Actual_i} \tag{2}$$

This approach allowed us to comprehensively assess the model's predictive accuracy in the context of disease cases, integrating various water quality parameters along with virus concentration data for a more robust evaluation.

The performance metrics of the final integrated model are as follows: MAPE = 16.94 %, R^2 score = 0.87, RMSE = 1.11, MAE = 0.67.

The predicted number of confirmed cases and the actual confirmed cases are compared in the graph shown in Fig. 6.

These metrics demonstrate the accuracy and reliability of the integrated model in predicting the number of confirmed disease cases. The MAPE value of 16.94 % indicates that, on average, the model's predictions deviate by approximately 16.94 % from the actual values, showcasing the model's robustness in forecasting disease cases. The high R^2 value (0.87) signifies that a significant proportion of the variance in the data is captured by the model. Additionally, the relative low values of MAE and RMSE indicate that the model's predictions are close to the actual values.

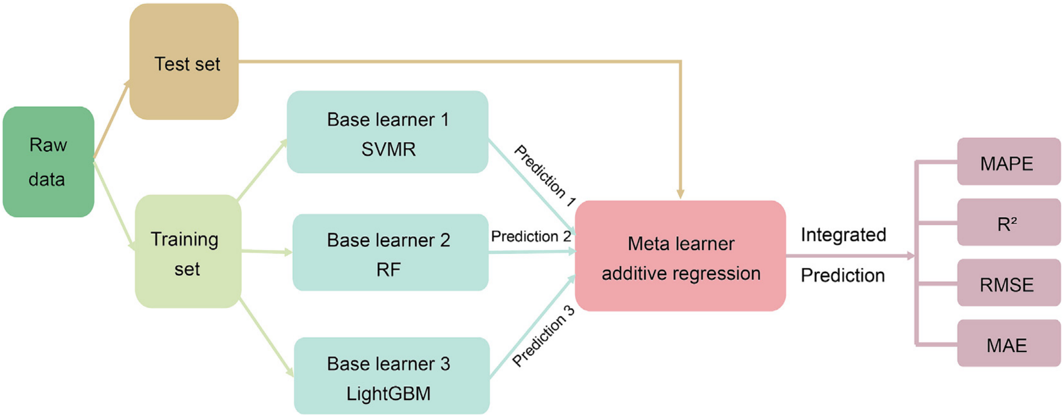

Fig. 5. The framework of stacking models. Abbreviations: SVMR, support vector machine regression; RF, random forest; LightGBM, light gradient-boosting machine; MAPE, mean absolute percentage error; R^2, R-squared score; RMSE, root mean square error; MAE, mean absolute error.

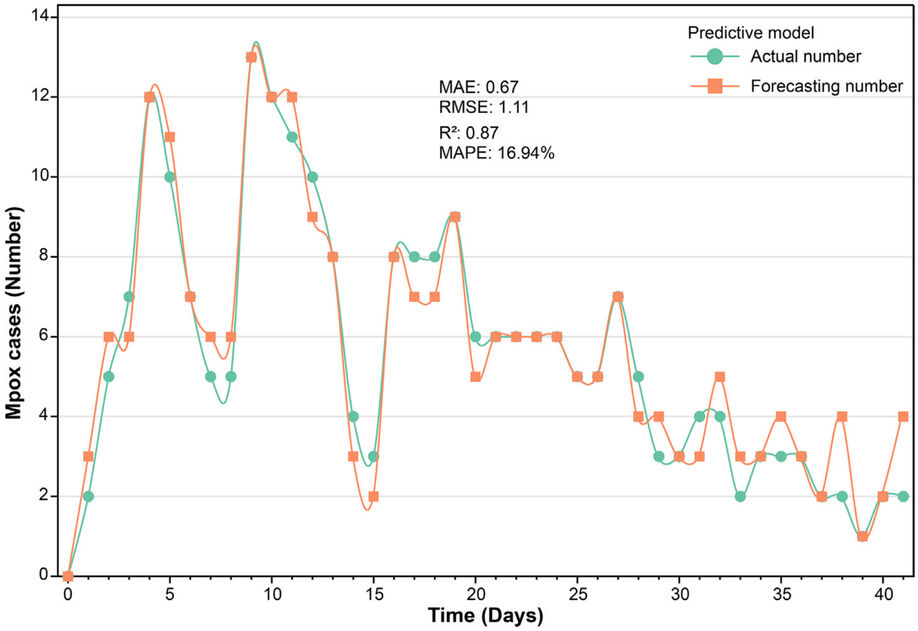

Fig. 6. Comparison of forecasting and actual number of confirmed cases of mpox based on a machine learning model. The *x*-axis represents the timeline of the study period in days, while the *y*-axis indicates the number of confirmed mpox cases. Abbreviations: MAPE, mean absolute percentage error; R^2, R-squared score; RMSE, root mean square error; MAE, mean absolute error.

4. Discussion

Since May 2022, the largest-ever documented outbreak of mpox in terms of magnitude and geographical spread began in human history, both in terms of its scale and geographical reach [41]. This unprecedented epidemic, caused by the clade IIb MPXV, has swiftly traversed borders, infiltrating several countries that have historically remained untouched by this disease.

Here, we developed an online sewage virus enrichment and concentration robot and underscored a significant relationship between the concentration of MPXV in wastewater and the clinically reported cases of mpox. Furthermore, we constructed an AI model for precise inferring the number of confirmed cases through the concentration of viruses in sewage. Over the course of this 100-day study, we employed an online sewage virus enrichment and concentration robot to systematically collect wastewater samples every six hours. This high-frequency online sewage sample processing solution, as opposed to traditional low-frequency approaches, aids in capturing the detoxification periods of patients and obtaining effective signals. Within this duration, we identified the presence of the MPXV in the wastewater on 41 distinct days out of a total of 100 days. Furthermore, out of the 283 collected samples, 153 tested positive for the MPXV, translating to a positive detection rate of 54.06 %. In our quest to uncover the link between the concentration of MPXV in wastewater and the clinically reported mpox cases, we matched the data of mpox patients with the dates when a positive signal for the MPXV was identified in the wastewater samples. The findings pointed towards a significant positive correlation between the concentration of MPXV in the hospital's wastewater samples and the clinically reported cases. Specifically, the peak concentration of MPXV in wastewater correlated with the number of mpox cases, and the mean concentration of MPXV in wastewater correlated with the number of mpox cases, indicating a notably strong correlation. Our research results are in concurrence with studies from France and the United States [42,43], which have underscored a potent correlation between the concentration of MPXV in wastewater and the number of mpox cases. Wurtzer et al. [43] observed a strong association between the weekly number of mpox patients and the corresponding concentration of mpox virus in sewage. Likewise, Wolfe et al. [42] observed a positive relationship between the occurrence of MPXV DNA in solid waste and the reported instances of the disease at locations where positive samples were detected for more than 10 days. Contrarily, Sherchan et al. [36] did not find any correlation between the MPXV concentration in wastewater and the count of new mpox cases. These findings suggest that the concentration of MPXV in wastewater could potentially act as an indicator of the number of clinical cases in the hospital. However, given that our study was conducted in a single hospital, it may limit the broader applicability of our findings. Therefore, further research is needed to validate our findings in diverse settings and to explore the potential applications of this method for monitoring other types of viruses in wastewater.

McMahan et al. [44], conducted a WBE study in South Carolina to monitor SARS-CoV-2 transmission. Through the analysis of wastewater samples for SARS-CoV-2 ribonucleic acid (RNA) and the utilization of a susceptible-exposed-infectious-recovered (SEIR) model based on the RNA mass rate, they were able to predict the number of infected individuals. Upon comparing their model predictions with confirmed cases, they found that the rate of unreported COVID-19 cases was approximately 11 times that of confirmed cases, a figure consistent with an independent estimate of 15 infections for each confirmed case in South Carolina. This underscores the robustness of the SEIR model as a method for estimating the total number of infected individuals in a sewershed, thus providing an additional and valuable tool for informing policy decisions. Ando et al. [45], developed a highly sensitive method, the Efficient and Practical Virus Identification System with Enhanced Sensitivity for Membrane (EPISENS-M), and a mathematical model based on viral shedding dynamics, for tracking SARS-CoV-2 RNA concentrations in wastewater and estimating newly reported cases using the RNA data and recent clinical data. Their study successfully predicted the cumulative number of newly reported cases after 5 days with a high degree of precision, demonstrating that the EPISENS-M method combined with the mathematical model can be a powerful tool for predicting COVID-19 cases, particularly in the absence of intensive clinical surveillance. In contrast to COVID-19 testing, clinical testing for MPXV is limited to cases in which patients exhibit severe skin boils. Asymptomatic and mildly symptomatic cases often go unreported, as individuals may not seek medical attention due to the self-limiting nature of the disease. Furthermore, the development of severe boils after contracting the virus can be a gradual process. The absence of precise data on MPXV infection presents challenges in creating a prediction model or establishing a correlation between clinical cases and WBE data.

Fascinatingly, a study modeled MPXV excretion to investigate the likelihood of detecting MPXV in wastewater across various settings and detection limit ranges [32]. This theoretical research indicated the potential for monitoring MPXV in sewage. However, it also emphasized the importance of collecting data on infection kinetics and shedding in different human body fluids to establish a connection between sewage surveillance data and clinical mpox cases [32]. Our ensemble model based on AI-Powered wastewater MPXV surveillance data exhibits strong accuracy and reliability in predicting the number of mpox cases. The relatively low values of MAE and RMSE indicate a close correspondence between the model's predictions and the actual values, demonstrating its efficacy and practical utility in predicting disease case numbers. Additionally, the model's R^2 score of 0.87 implies that it can account for 0.87 of the variability in the data. This high score underscores the model's excellent fit to the data, capturing most of its variability, thereby further validating the model's effectiveness in predicting disease case numbers. However, we also note that the MAPE is 16.94 %, indicating that the model's predictions deviate from the actual values by an average of approximately 16.94 %. Although this error rate is not negligible, it is acceptable given the numerous variables our model incorporates, including water quality parameters and virus concentrations, and the potentially complex interrelationships among these variables. Overall, our model demonstrates high accuracy and robustness in predicting the number of disease cases.

However, we must also acknowledge some limitations of the model, such as potential prediction errors. Future research can further optimize the model by incorporating more input variables, adjusting model parameters to enhance the model's predictive accuracy. This AI-driven prediction approach opens up a new avenue of thinking. Hospitals are typically the first warning and important diffusion protection points for infectious diseases. Long-term monitoring of infectious disease viruses such as mpox in hospital sewage stations can accurately predict the number of infections, provided that the monitoring duration is sufficiently long and the sample database is rich. This is of significant importance for the construction of urban epidemic surveillance and early warning systems. The non-intrusive, non-complainant nature, and point-to-surface characteristics of sewage virus monitoring can enhance the efficiency and objectivity of epidemic data collection while reducing data acquisition costs. Furthermore, hospital sewage virus data can provide practical references for nosocomial infection monitoring.

In conclusion, our study highlights the promising potential of using automated online sewage virus enrichment and concentration robots for monitoring the presence of viruses in wastewater. The strong correlation between the concentration of MPXV in wastewater and clinically confirmed cases suggests that this method could potentially be utilized as an early warning system for outbreaks of mpox and possibly other viruses. Future research should focus on enhancing the sensitivity and specificity of this method and exploring its potential applications in different settings.

Acknowledgements

This work is supported by National Key Research and Development Program of China (2023YFC3041500), Shenzhen Medical Research Funding (D2301014), Shenzhen High-level Hospital Construction Fund (23250G1001, XKJS-CRGRK-005), Shenzhen Clinical Research Center for Emerging Infectious Diseases (No. LCYSSQ20220823091203007), The Science and Technology Innovation and Entrepreneurship Special Foundation of Shenzhen (JSGG20220226090203006).

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Author contributions

Guanyong Ou: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Yuxuan Tang:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Jiexiang Liu:** Methodology, Investigation. **Yabin Hao:** Formal analysis. **Zhi Chen:** Formal analysis. **Ting Huang:** Methodology. **Shaxi Li:** Methodology. **Shiyu Niu:** Methodology. **Yun Peng:** Methodology. **Jiaqi Feng:** Investigation. **Hongwei Tu:** Supervision. **Yang Yang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Han Zhang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Yingxia Liu:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bsheal.2024.07.002.

References

[1] Y. Xiang, A. White, Monkeypox virus emerges from the shadow of its more infamous cousin: Family biology matters, Emerg. Microbes Infect. 11 (1) (2022) 1768–1777, https://doi.org/10.1080/22221751.2022.2095309.

[2] A. Gessain, E. Nakoune, Y. Yazdanpanah, Monkeypox, N. Engl. J. Med. 387 (19) (2022) 1783–1793, https://doi.org/10.1056/NEJMra2208860.

[3] T.G. Senkevich, N. Yutin, Y.I. Wolf, E.V. Koonin, B. Moss, Ancient gene capture and recent gene loss shape the evolution of orthopoxvirus-host interaction genes, mBio 12(4) (2021) e0149521, 10.1128/mBio.01495-21.

[4] P.T. Mwanbal, K.F. Tshioko, A. Moudi, V. Mukinda, G.N. Mwema, D. Messinger, L. Okito, D. Barakymfyte, P. Malfait, R. Pebody, et al., Human monkeypox in kasai oriental, zaire (1996–1997), Euro. Surveill. 2 (5) (1997) 33–35, https://doi.org/10.2807/esm.02.05.00161-en.

[5] A. Nalca, A.W. Rimoin, S. Bavari, C.A. Whitehouse, Reemergence of monkeypox: Prevalence, diagnostics, and countermeasures, Clin. Infect. Dis. 41 (12) (2005) 1765–1771, https://doi.org/10.1086/498155.

[6] M. Curaudeau, C. Besombes, E. Nakouné, A. Fontanet, A. Gessain, A. Hassanin, Identifying the most probable mammal reservoir hosts for monkeypox virus based on ecological niche comparisons, Viruses 15 (3) (2023) 727, https://doi.org/10.3390/v15030727.

[7] A.W. Rimoin, P.M. Mulembakani, S.C. Johnston, J.O. Lloyd Smith, N.K. Kisalu, T.L. Kinkela, S. Blumberg, H.A. Thomassen, B.L. Pike, J.N. Fair, et al., Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the democratic republic of congo, Proc. Natl. Acad. Sci. U. S. A. 107 (37) (2010) 16262–16267, https://doi.org/10.1073/pnas.1005769107.

[8] E.M. Bunge, B. Hoet, L. Chen, F. Lienert, H. Weidenthaler, L.R. Baer, R. Steffen, The changing epidemiology of human monkeypox-a potential threat? A systematic review, PLoS. Negl. Trop. Dis. 16 (2) (2022) e0010141, https://doi.org/10.1371/journal.pntd.0010141.

[9] S.S. Marennikova, E.M. Seluhina, N.N. Mal'ceva, K.L. Cimiskjan, G.R. Macevic, Isolation and properties of the causal agent of a new variola-like disease (monkeypox) in man, Bull. World Health Organ. 46 (5) (1972) 599–611.

[10] J.P. Thornhill, S. Barkati, S. Walmsley, J. Rockstroh, A. Antinori, L.B. Harrison, R. Palich, A. Nori, I. Reeves, M.S. Habibi, et al., Monkeypox virus infection in humans across 16 countries - April-June 2022, N. Engl. J. Med. 387 (8) (2022) 679–691, https://doi.org/10.1056/NEJMoa2207323.

[11] Z. Jezek, M. Szczeniowski, K.M. Paluku, M. Mutombo, Human monkeypox: Clinical features of 282 patients, J. Infect. Dis. 156 (2) (1987) 293–298, https://doi.org/10.1093/infdis/156.2.293.

[12] M. Ahmed, H. Naseer, M. Arshad, A. Ahmad, Monkeypox in 2022: A new threat in developing, Ann. Med. Surg. (Lond.) 78 (2022) 103975, https://doi.org/10.1016/j.amsu.2022.103975.

[13] G. Srivastava, G. Srivastava, Human monkeypox disease, Clin. Dermatol. 40 (5) (2022) 604–612, https://doi.org/10.1016/j.clindermatol.2022.08.009.

[14] P.v. Magnus, E.K. Andersen, K.B. Petersen, A. Birch-Andersen, A pox-like disease in cynomolgus monkeys, APMIS: Acta Pathologica Microbiologica, Scandinavica 46 (2) (1959) 156–176, https://doi.org/10.1111/j.1699-0463.1959.tb00328.x.

[15] N. Cassir, F. Cardona, H. Tissot-Dupont, C. Bruel, B. Doudier, S. Lahouel, K. Bendamardji, C. Boschi, S. Aherfi, S. Edouard, et al., Observational cohort study of evolving epidemiologic, clinical, and virologic features of monkeypox in Southern France, Emerg. Infect. Dis. 28 (12) (2022) 2409–2415, https://doi.org/10.3201/eid2812.221440.

[16] R.A. Okyay, Another epidemic in the shadow of COVID-19 pandemic: A review of monkeypox, Eurasian J. Med. Oncol. 6 (2022) 95–99, https://doi.org/10.14744/ejmo.2022.2022.

[17] E. Petersen, A. Kantele, M. Koopmans, D. Asogun, A. Yinka-Ogunleye, C. Ihekweazu, A. Zumla, Human monkeypox: Epidemiologic and clinical characteristics, diagnosis, and prevention, Infect. Dis. Clin. North. Am. 33 (4) (2019) 1027–1043, https://doi.org/10.1016/j.idc.2019.03.001.

[18] C.M. Gigante, B. Korber, M.H. Seabolt, K. Wilkins, W. Davidson, A.K. Rao, H. Zhao, T.G. Smith, C.M. Hughes, F. Minhaj, et al., Multiple lineages of monkeypox virus detected in the United States, 2021–2022, Science 378 (6619) (2022) 560–565, https://doi.org/10.1126/science.add4153.

[19] A.M. Likos, S.A. Sammons, V.A. Olson, A.M. Frace, Y. Li, M. Olsen-Rasmussen, W. Davidson, R. Galloway, M.L. Khristova, M.G. Reynolds, et al., A tale of two clades: Monkeypox viruses, J. Gen. Virol. 86 (Pt 10) (2005) 2661–2672, https://doi.org/10.1099/vir.0.81215-0.

[20] A. Karagoz, H. Tombuloglu, M. Alsaeed, G. Tombuloglu, A.A. AlRubaish, A. Mahmoud, S. Smajlović, S. Ćordić, A.A. Rabaan, E. Alsuhaimi, Monkeypox (mpox) virus: Classification, origin, transmission, genome organization, antiviral drugs, and molecular diagnosis, J. Infect. Public Health 16 (4) (2023) 531–541, https://doi.org/10.1016/j.jiph.2023.02.003.

[21] N.L. Bragazzi, W.A. Woldegerima, S.A. Iyaniwura, Q. Han, X. Wang, A. Shausan, K. Badu, P. Okwen, C. Prescod, M. Westin, et al., Knowing the unknown: The underestimation of monkeypox cases. Insights and implications from an integrative review of the literature, Front. Microbiol. 13 (2022) 1011049, https://doi.org/10.3389/fmicb.2022.1011049.

[22] M. Kozlov, Monkeypox declared a global emergency: Will it help contain the outbreak?, Nature (2022), https://doiorg/10.1038/d41586-022-02054-7.

[23] K. Jahn, D. Dreifuss, I. Topolsky, A. Kull, P. Ganesanandamoorthy, X. Fernandez-Cassi, C. Bänziger, A.J. Devaux, E. Stachler, L. Caduff, et al., Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using cojac, Nat. Microbiol. 7 (8) (2022) 1151–1160, https://doi.org/10.1038/s41564-022-01185-x.

[24] S. Karthikeyan, J.I. Levy, P. De Hoff, G. Humphrey, A. Birmingham, K. Jepsen, S. Farmer, H.M. Tubb, T. Valles, C.E. Tribelhorn, et al., Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission, Nature 609 (7925) (2022) 101–108, https://doi.org/10.1038/s41586-022-05049-6.

[25] B. Nelson, What poo tells us: Wastewater surveillance comes of age amid COVID, monkeypox, and polio, BMJ 378 (2022) o1869, https://doi.org/10.1136/bmj.o1869.

[26] F. Wu, W.L. Lee, H. Chen, X. Gu, F. Chandra, F. Armas, A. Xiao, M. Leifels, S.F. Rhode, S. Wuertz, et al., Making waves: wastewater surveillance of SARS-CoV-2 in an endemic future, Water Res. 219 (2022) 118535, https://doi.org/10.1016/j.watres.2022.118535.

[27] M. Mousazadeh, R. Ashoori, B. Paital, I. Kabdaşlı, Z. Frontistis, M. Hashemi, M.A. Sandoval, S. Sherchan, K. Das, M.M. Emamjomeh, Wastewater based epidemiology perspective as a faster protocol for detecting coronavirus rna in human populations: A review with specific reference to SARS-CoV-2 virus, Pathogens 10 (8) (2021) 1008, https://doi.org/10.3390/pathogens10081008.

[28] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C. L. Huang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (7798) (2020) 270–273, https://doi.org/10.1038/s41586-020-2012-7.

[29] G. Ou, Y. Tang, S. Niu, L. Wu, S. Li, Y. Yang, J. Wang, Y. Peng, C. Huang, W. Hu, et al., Wastewater surveillance and an automated robot: Effectively tracking SARS-CoV-2 transmission in the post-epidemic era, Natl. Sci. Rev. 10 (6) (2023) nwad089, https://doi.org/10.1093/nsr/nwad089.

[30] A. Plaza-Garrido, M. Ampuero, A. Gaggero, C.A. Villamar-Ayala, Norovirus, hepatitis a and SARS-CoV-2 surveillance within chilean rural wastewater treatment plants based on different biological treatment typologies, Sci. Total. Environ. 863 (2022) 160685, https://doi.org/10.1016/j.scitotenv.2022.160685.

[31] A.C. Singer, J.R. Thompson, C.R.M. Filho, R. Street, X. Li, S. Castiglioni, K.V. Thomas, A world of wastewater-based epidemiology, Nat. Water 1 (5) (2023) 408–415, https://doi.org/10.1038/s44221-023-00083-8.

[32] W. Chen, K. Bibby, Model-based theoretical evaluation of the feasibility of using wastewater-based epidemiology to monitor monkeypox, Environ. Sci. Technol. Lett. 9 (9) (2022) 772–778, https://doi.org/10.1021/acs.estlett.2c00496.

[33] I. Gul, C. Liu, X. Yuan, Z. Du, S. Zhai, Z. Lei, Q. Chen, M.A. Raheem, Q. He, Q. Hu, et al., Current and perspective sensing methods for monkeypox virus, Bioengineering (basel) 9 (10) (2022) 571, https://doi.org/10.3390/bioengineering9100571.

[34] A. Peiró-Mestres, I. Fuertes, D. Camprubí-Ferrer, M. Marcos, A. Vilella, M. Navarro, L. Rodriguez-Elena, J. Riera, A. Català, M.J. Martínez, et al., Frequent detection of monkeypox virus DNA in saliva, semen, and other clinical samples from 12 patients, Barcelona, Spain, May to June 2022, Euro. Surveill. 27 (28) (2022) 2200503, https://doi.org/10.2807/1560-7917.Es.2022.27.28.2200503.

[35] C.K. Yinda, D.H. Morris, R.J. Fischer, S. Gallogly, Z.A. Weishampel, J.R. Port, T. Bushmaker, J.E. Schulz, K. Bibby, N. van Doremalen, et al., Stability of monkeypox virus in body fluids and wastewater, Emerg. Infect. Dis. 29 (10) (2023) 2065–2072, https://doi.org/10.3201/eid2910.230824.

[36] S.P. Sherchan, T. Solomon, O. Idris, D. Nwaubani, O. Thakali, Wastewater surveillance of mpox virus in baltimore, Sci. Total. Environ. 891 (2023) 164414, https://doi.org/10.1016/j.scitotenv.2023.164414.

[37] S. Seang, S. Burrel, E. Todesco, V. Leducq, G. Monsel, D. Le Pluart, C. Cordevant, V. Pourcher, R. Palich, Evidence of human-to-dog transmission of monkeypox virus, Lancet 400 (10353) (2022) 658–659, https://doi.org/10.1016/s0140-6736(22)01487-8.

[38] T.Z. Song, Y.T. Zheng, Monkeypox, wild animals, and potential big problem, Zool. Res. 43 (4) (2022) 612–614, https://doi.org/10.24272/j.issn.2095-8137.2022.217.

[39] M. Ribeiro, R.G. da Silva, V.C. Mariani, L.D.S. Coelho, Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for brazil, Chaos Solitons Fractals 135 (2020) 109853, https://doi.org/10.1016/j.chaos.2020.109853.

[40] L. Vaughan, M. Zhang, H. Gu, J.B. Rose, C.C. Naughton, G. Medema, V. Allan, A. Roiko, L. Blackall, A. Zamyadi, An exploration of challenges associated with machine learning for time series forecasting of COVID-19 community spread using wastewater-based epidemiological data, Sci. Total. Environ. 858 (2023) 159748, https://doi.org/10.1016/j.scitotenv.2022.159748.

[41] H. Laurenson-Schafer, N. Sklenovská, A. Hoxha, S.M. Kerr, P. Ndumbi, J. Fitzner, M. Almiron, L.A. de Sousa, S. Briand, O. Cenciarelli, et al., Description of the first global outbreak of mpox: An analysis of global surveillance data, Lancet Glob. Health 11 (7) (2023) e1012–e1023, https://doi.org/10.1016/s2214-109x(23)00198-5.

[42] M.K. Wolfe, A.T. Yu, D. Duong, M.S. Rane, B. Hughes, V. Chan-Herur, M. Donnelly, S. Chai, B.J. White, D.J. Vugia, et al., Use of wastewater for mpox outbreak surveillance in California, N. Engl. J. Med. 388 (6) (2023) 570–572, https://doi.org/10.1056/NEJMc2213882.

[43] S. Wurtzer, M. Levert, E. Dhenain, M. Boni, J.N. Tournier, N. Londinsky, A. Lefranc, O. Ferraris, L. Moulin, First detection of monkeypox virus genome in sewersheds in france: The potential of wastewater-based epidemiology for monitoring emerging disease, Environ. Sci. Technol. Lett. 9 (11) (2022) 991–996, https://doi.org/10.1021/acs.estlett.2c00693.

[44] C.S. McMahan, S. Self, L. Rennert, C. Kalbaugh, D. Kriebel, D. Graves, C. Colby, J. A. Deaver, S.C. Popat, T. Karanfil, et al., COVID-19 wastewater epidemiology: A model to estimate infected populations, Lancet Planet. Health 5 (12) (2021) e874–e881, https://doi.org/10.1016/s2542-5196(21)00230-8.

[45] H. Ando, M. Murakami, W. Ahmed, R. Iwamoto, S. Okabe, M. Kitajima, Wastewater-based prediction of COVID-19 cases using a highly sensitive SARS-CoV-2 rna detection method combined with mathematical modeling, Environ. Int. 173 (2023) 107743, https://doi.org/10.1016/j.envint.2023.107743.