Optimizing the SEIR Model Based on Genetic Algorithms by Incorporating Human Self-Isolation Willingness Factors

Jingbo Hou*
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, 43600, Malaysia
**Corresponding Author*

Abstract: Epidemics have consistently impacted human life, and the recent global outbreak of COVID-19 has profoundly affected both health and the economy. This underscores the urgency of optimizing infectious disease dynamics models and the necessity of conducting precise predictions of epidemic trends. Addressing the theoretical limitations of traditional SEIR models in capturing the dynamic nature of individual preventive behaviors, this study innovatively incorporates human self-isolation willingness as a variable into the model architecture. We focus on improving the genetic algorithm-based SEIR (GA-SEIR) model to construct an enhanced SEIR model grounded in genetic algorithms. Integrating epidemiology with computational intelligence, this study develops a multi-objective optimization model to achieve more precise predictions of epidemic trends. This supports targeted prevention and control decisions, thereby mitigating the impact of epidemics on human life. While numerous studies have extended the SEIR framework, existing research predominantly focuses on biological and physical transmission processes, with limited exploration of the dynamic factors of human individual behavior. The uniqueness of this study lies in incorporating human self-isolation willingness into the GA-SEIR model, known for its higher predictive accuracy, to further refine and optimize it. Adopting a cross-disciplinary perspective on the complex system coupling mechanism between "human behavior and disease transmission," this study employs a hybrid approach combining genetic algorithm (GA) optimization with multi-source data-driven modeling. It presents an enhanced GA-SEIR model that integrates human self-isolation factors, aiming to improve the model's predictive accuracy for epidemic spread and enhance the precision of control policies during public health emergencies.

Keywords: Genetic Algorithm (GA); SEIR Model; Human Self-Isolation Willingness; Epidemic Forecasting; COVID-19; Infectious Disease Modeling

1. Introduction

1.1 Research Background

The definition of a pandemic is "an epidemic occurring worldwide, over a very wide area, crossing international boundaries, and usually affecting a large number of people"[1]. This definition indicates that the vast majority of epidemics can be characterized and further described by measuring the transmissibility and severity of the disease. Previously, the World Health Organization (WHO) listed nineteen (19) pandemic epidemics[2]. On March 11, 2020, the WHO declared the novel coronavirus (2019-nCoV) as the twentieth addition to this list [3]. By April 25, 2020, confirmed cases of the novel coronavirus reached 2,810,325, with 193,825 confirmed deaths, affecting 213 countries, regions, or territories. The immense negative impact of infectious diseases has created an urgent need for robust and accurate predictive models, enabling humanity to better address the challenges posed by such outbreaks [4].

In recent years, infectious disease models such as the SEIR model have played a pivotal role in predicting the COVID-19 pandemic and evaluating prevention and control strategies. The traditional SEIR model reveals the fundamental patterns of disease transmission by categorizing the population into susceptible (S), exposed (E), infected (I), and recovered (R) states. However, existing models often assume fixed parameters, overlooking the dynamic impact of human behavior on the transmission process. For instance, studies indicate that public self-

isolation behaviors during the COVID-19 pandemic significantly altered viral transmission pathways, yet existing models remain inadequate in quantifying such behaviors. Furthermore, traditional parameter optimization methods (e.g., least squares) often converge to local optima when confronting complex nonlinear relationships, leading to predictive biases. Therefore, there is an urgent need to integrate behavioral science with intelligent algorithms to construct more realistic transmission models, thereby enhancing the accuracy of pandemic forecasting and the effectiveness of containment strategies.

As novel infectious diseases like COVID-19 ravage the globe, public health crises continue to escalate, exerting profound impacts on society, the economy, and individual health. In epidemic control efforts, the SEIR (Susceptible, Exposed, Infected, Removed) model-a classic framework for infectious disease transmission dynamics-has been widely applied to predict disease spread trends and formulate intervention strategies. However, traditional SEIR models primarily focus on population transmissibility, overlooking the crucial role of individual behavior in epidemics-particularly the key factor of population isolation willingness. Moreover, many extension studies based on SEIR models concentrate on biological and physical transmission processes, with limited research delving into the dynamic factors of human individual behavior. During actual epidemics, whether individuals are willing to adopt isolation measures directly impacts the transmission speed and scale of the outbreak. Therefore, incorporating population isolation willingness as a new infection model can more accurately reflect real-world disease transmission dynamics. The genetic algorithm (GA) proposed by Professor John Holland is well-suited for complex optimization problems, effectively handling nonlinear and multimodal issues. The introduction of GA for parameter optimization significantly enhances the accuracy of traditional SEIR models. This study combines GA optimization of the SEIR model with the incorporation of human behavioral factors (self-isolation willingness), thereby expanding the theoretical framework of infectious disease transmission modeling. This innovative perspective not only provides new research avenues for infectious disease modeling but also reveals the impact of individual behavior on epidemic spread, offering complementary insights for future epidemic prevention and control theories.

1.2 Research Review

In recent years, with the global outbreak of the COVID-19 pandemic, the importance of optimizing traditional SEIR models has been reaffirmed. Among numerous attempts, many studies have found that genetic algorithms can effectively improve SEIR models. One such study employed two enhanced SEIR models: one considering infectiousness during the incubation period and another accounting for both incubation period infectiousness and isolation measures. These models were combined with genetic algorithms (GA) for disease prediction. The study found that the improved SEIR model combined with genetic algorithms (GA) could better predict the entire epidemic period [5]. The effectiveness of genetic algorithms in optimizing the SEIR model was also well demonstrated in another study. The research team applied a Logarithmic Growth Regression (LGR) model to COVID-19 confirmed case data from Thailand's disease control department. They optimized model parameters using both genetic algorithms (GA) and the Gauss-Newton algorithm (GNA) to enhance prediction accuracy. Results indicated that compared to the traditional SIR model, the LGR-GA/GNA model demonstrated superior goodness-of-fit and lower root mean square error (RMSE), confirming its stronger predictive capability [6].

Currently, an increasing number of studies are attempting to optimize the SEIR model from multiple perspectives. Some research suggests that traditional infectious disease studies typically emphasize the biological aspects of pathogens, but must also consider the social and ecological processes influencing disease emergence and transmission [7], as well as the crucial role of human behavior in disease spread [8]. Findings from E.P. Fenichel et al. demonstrate that incorporating adaptive human behavior significantly alters epidemic prediction trajectories, with implications for parameter estimation, interpretation, and the formulation of social distancing policies [9]. The literature "A review and agenda for integrated disease models including social and behavioral factors" repeatedly emphasizes that social and behavioral factors are crucial for the emergence, transmission, and control of human diseases,

serving as key determinants of epidemic processes, duration, and outcomes. It further contends that better integrated modeling incorporating social and behavioral dynamics will enhance prediction accuracy [10]. "Modeling the interplay between human behavior and the spread of infectious diseases" similarly acknowledges the positive impact of social and behavioral factors on improving model prediction accuracy [11]. The paper "Integrating socio-psychological factors in the SEIR model optimized by a genetic algorithm for COVID-19 trend analysis" highlights and demonstrates that incorporating socio-psychological factors enhances the realism of infectious disease predictions. It further underscores the predictive accuracy of the GA-SEIR model through comparative analysis [12].

The Sigmoid function (S-function) finds extensive application in psychology, physiology, and biology for characterizing human behavioral psychology. Featuring an S-shaped curve, the Sigmoid function smoothly describes a system's transition from one state to another rather than abrupt changes. This makes it highly suitable for modeling gradual phenomena in sociology and psychology-such as behavioral shifts, attitude transformations, and decision-making processes-accurately reflecting the psychological and behavioral changes observed in real-world human interactions. [13]. The Sigmoid function concisely expresses complex social dynamics models (e.g., media diffusion, social influence, opinion dynamics). Its properties can mathematically model social psychological theories, aiding the development of interpretable predictive models through machine learning and mathematical modeling methods [14]. The properties of the Sigmoid function-such as capturing threshold effects, hierarchical responses, and nonlinear dynamics-align closely with the characteristics of social factors influencing ethical decision-making. By leveraging these features, the model provides a reasonable and practical mathematical framework for incorporating social influence [15]. Therefore, this study will employ the Sigmoid function to introduce the factor of human willingness to self-isolate.

In summary, this study will utilize COVID-19 data to optimize the GA-SEIR model by incorporating the self-isolation willingness factor through genetic algorithm parameter optimization, aiming to enhance the accuracy of epidemic predictions.

1.3 Research Hypotheses and Objectives

The study will first employ the Sigmoid function to characterize the psychological response of individuals' willingness to self-isolate based on the severity of the epidemic. This will be incorporated into the SEIR model for dynamic adjustment of contact rates, thereby constructing an improved SEIR model that integrates human self-isolation preferences to enhance predictive accuracy. Concurrently, genetic algorithms will be utilized for parameter optimization to further refine the model's forecasting precision. Finally, the model's predictive accuracy and policy simulation capabilities will be validated using historical epidemic data.

The research may encounter several challenges. First is the adaptability and robustness of genetic algorithm optimization. Given the nonlinearity and uncertainty in epidemic transmission processes, designing an appropriate genetic algorithm framework to avoid getting stuck in local optima requires careful consideration. Additionally, data-related issues may arise. Incomplete or noisy epidemic data could impact both model accuracy and the effectiveness of genetic algorithm optimization.

This study will enhance the classical SEIR model by incorporating the factor of "human self-isolation willingness," thereby increasing its realism and adaptability. Concurrently, genetic algorithm optimization of model parameters will improve predictive accuracy in scenarios where human behavior significantly influences epidemic transmission.

2. Research Methods

2.1 SEIR model

In the classic SEIR model, the population is divided into four categories:

S(t): Susceptible individuals

E(t): Exposed individuals (infected but not yet infectious)

I(t): Infectious individuals (capable of transmitting the disease)

R(t): Removed individuals (including recovered or deceased)

The fundamental differential equations are:

$$\frac{dS}{dt} = -\beta \cdot \frac{SI}{N},$$

$$\frac{dE}{dt} = \beta \cdot \frac{SI}{N} - \sigma E,$$

$$\frac{dI}{dt}=\sigma E-\gamma I,$$
$$\frac{dR}{dt}=\gamma I \quad (1)$$

Where:

β: Contact rate (the probability of a susceptible individual contracting the infection after contact with an infected person);

σ: Conversion rate during incubation (the rate at which individuals in the incubation state become infected, approximately 1/average incubation period);

γ: Recovery rate (the rate at which infected individuals recover and are removed from the system, approximately 1/average infectious period);

N = S + E + I + R: Population size.

2.2 Factors Influencing Human Self-Isolation Willingness

In traditional SEIR models, β is assumed to be constant. However, in reality, public self-isolation behaviors dynamically alter the actual contact rate. To address this, this study introduces the isolation willingness function q(t), formulated as a Sigmoid function:

$$q(t)=\frac{1}{1+e^{-k\cdot(I(t)-\theta)}} \quad (2)$$

Where:

$q(t)$ Average willingness to self-isolate among the population at time t, ranging from 0 (completely unwilling to isolate) to 1 (completely willing to isolate)

I(t): Proportion or number of infected individuals.

k. Behavioral sensitivity parameter reflecting the public's reaction speed to changes in infection numbers. Higher values indicate more "intense" public response and steeper changes.

θ: Behavioral threshold indicating when infection numbers reach a certain level, the population begins to significantly increase its willingness to isolate.

Based on this, the contact rate is adjusted:

$$\beta(t)=\beta_0\cdot\left(1-q(t)\right) \quad (3)$$

Where:

β_0 : Baseline contact rate (value without protective measures)

$\beta(t)$: Dynamic contact rate, influenced by isolation preferences

Then substitute the modified contact rate $\beta(t)$ into the model:

$$\frac{dS}{dt}=-\beta(t)\cdot\frac{SI}{N},$$
$$\frac{dE}{dt}=\beta(t)\cdot\frac{SI}{N}-\sigma E,$$
$$\frac{dI}{dt}=\sigma E-\gamma I,$$
$$\frac{dR}{dt}=\gamma I. \quad (4)$$

2.3 GA Optimization of SEIR Models

Since the key parameters β_0 , σ, γ, k, and θ in SEIR models are difficult to obtain directly from statistical data, and complex nonlinear coupling relationships exist among these parameters, traditional least squares or gradient descent methods are prone to local optima. This study employs a Genetic Algorithm (GA) for global optimization of model parameters to reduce prediction errors between model outputs and actual observational data, thereby enhancing forecasting accuracy.

The Genetic Algorithm is a heuristic optimization method based on "natural selection and genetic mechanisms," exhibiting strong global search capabilities and robustness. Real-coded GA is employed to represent parameter individuals, ensuring precision in continuous parameter optimization. For the fitness function design, the Log-MSE (Logarithmic Mean Squared Error) metric is adopted to measure the discrepancy between model predictions and actual epidemic data, with its core computational logic illustrated in Figure 1.

```
def fitness(params, init_vals, t_range, N, real_I):  2 usages
    result = run_seir(params, init_vals, t_range, N)
    I_model = result[:, 2]
    error = np.mean((np.log1p(I_model) - np.log1p(real_I)) ** 2)
    return error
```

Figure 1. Implementation Logic of the Fitness Function in the GA-SEIR Model

Here, I_model denotes the sequence of predicted infected individuals obtained via the SEIR differential equations (solved using odeint), while real_I represents the actual infected cases published by the World Health Organization (WHO). Applying a logarithmic transformation (np.log1p) effectively mitigates the impact of extreme values during the epidemic peak phase, thereby enhancing the smoothness and robustness of the error metric.

In the algorithm design, the initial population size is set to 100. with each individual defined by five parameters within the following ranges: $\beta_0 \in [0.05, 1.0]$, $\sigma \in [0.05, 0.5]$, $\gamma \in [0.01, 0.5]$, $k \in [1, 100]$, and $\theta \in [0.001, 0.1]$. The primary workflow of the GA comprises six stages: initialization, fitness evaluation, selection, crossover, mutation, and elite retention. To prevent the loss of optimal solutions during evolution, the algorithm employs an elite retention strategy. As shown in Figure 2, the top 20% of individuals with the highest fitness in each generation are directly retained for the next generation, thereby ensuring the continuity of optimal solutions.

```
elite_count = max(1, pop_size // 5)
elites = [ind for ind, score in scored_pop[:elite_count]]
```

Figure 2. Implementation of Elite Retention Strategies in GA-SEIR Models

To prevent premature convergence of the algorithm, this study introduces an adaptive mutation mechanism. As shown in Figure 3, this mechanism maintains strong perturbations during the early stages to enhance global search capabilities while gradually reducing the mutation rate in later stages to promote precise convergence, thereby achieving a balance between exploration and convergence.

```
mutation_strength = 0.5 * (1 - gen / generations)
```

Figure 3. Implementation of the Adaptive Variation Mechanism in the GA-SEIR Model

To further enhance fitting accuracy and computational efficiency, this study introduces a local search algorithm for parameter fine-tuning following genetic algorithm optimization. Specifically, the Nelder-Mead simplex method is employed to perform secondary optimization on the optimal parameters generated by the genetic algorithm, ensuring rapid convergence within the optimal region. This hybrid optimization strategy combines the global search capability of genetic algorithms with the fine-tuning characteristics of local algorithms, effectively improving the model's fitting accuracy and stability. The algorithm is illustrated in Figure 4.

```
res = minimize(
    lambda x: fitness(x, init_vals, t_range, N, real_I),
    params_ga,
    method='Nelder-Mead',
    options={'maxiter':500, 'xatol':1e-6, 'fatol':1e-6}
)
```

Figure 4. Implementation of the Local Search Algorithm (Nelder–Mead) in the GA-SEIR Model

In summary, this study combines genetic algorithms with a local fine-tuning strategy to construct a parameter optimization framework with adaptive search capabilities. This approach enhances the fitting accuracy and stability of the SEIR model, improves its generalization ability and interpretability under complex epidemic data, and provides an efficient and feasible solution for parameter estimation in epidemiological modeling.

3. Research Findings

The data used in this study originates from the "WHO COVID-19 Global Daily Data" released by the World Health Organization (WHO). This dataset systematically records daily COVID-19 epidemic statistics for countries and regions worldwide from early 2020 to the present. This study extracted and utilized the epidemic data for China from this dataset. The GA-SEIR model, incorporating human self-isolation willingness factors, was applied to predict the COVID-19 epidemic trends in China.

3.1 Comparison of Model Prediction Performance

To validate the predictive accuracy and applicability of the GA-SEIR model incorporating human self-isolation willingness, this study compared the model's projections with actual epidemic data from China. Figure 4 illustrates the fit between the model's predicted curve and real-world data, where blue points represent smoothed actual infection counts and the orange line indicates the predictions from the improved GA-SEIR model.

As shown in Figure 5, the model predictions align closely with the actual observed data in terms of overall trends. The model effectively captures the rapid rise phase, plateau phase, and

subsequent slow growth phase of the epidemic, accurately reflecting the dynamic evolution of transmission. During the early epidemic phase, model predictions nearly overlapped with actual data, indicating high precision in parameter initialization and transmission rate fitting. At the peak and plateau phases, the model curve's amplitude closely matched actual data fluctuations, demonstrating that incorporating self-isolation willingness effectively simulates the public's behavioral suppression of epidemic spread.

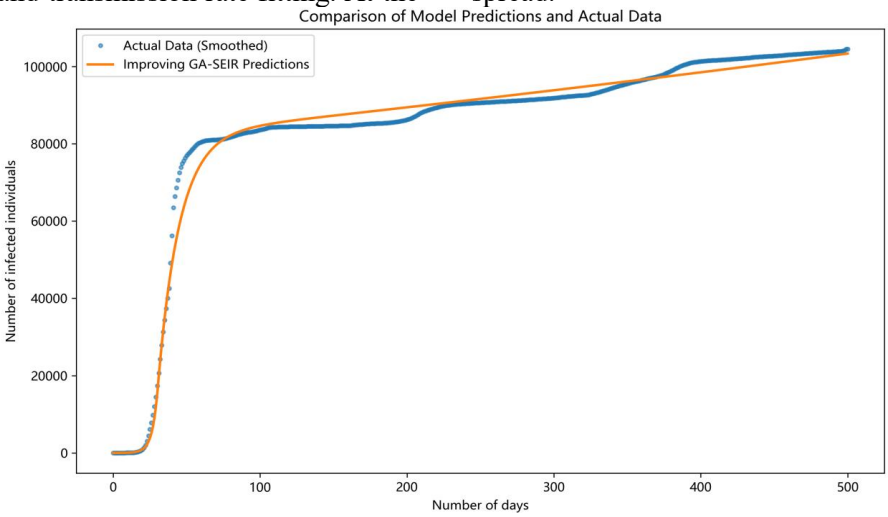

Figure 5. Comparison Chart of Actual vs. Predicted Data

Quantitative results show the improved GA-SEIR model outperforms traditional SEIR and GA-optimized SEIR models in predictive performance, with key fitting metrics being:

Mean Absolute Percentage Error (MAPE) = 6.94%, Coefficient of Determination (R^2) = 0.986.

An R^2 close to 1 indicates strong explanatory power, effectively reflecting actual epidemic trends; The low MAPE value indicates minimal prediction error, demonstrating the model's high stability and accuracy across different time periods.

In summary, the improved GA-SEIR model exhibits outstanding performance in both overall fitting effectiveness and capturing local trend variations. This confirms that incorporating self-isolation willingness factors and genetic algorithm optimization significantly enhances the model's precision, robustness, and interpretability.

3.2 Analysis of Model Error and Residual Distribution

To validate the stability and unbiasedness of the improved GA-SEIR model's prediction results, this study conducted a statistical analysis of the model residuals on a logarithmic scale. The residual is defined as the difference between the actual number of infections and the model's predicted value, expressed in logarithmic form as:

$$Residual_{log}(t)=\ln(1+I_{real}(t))-\ln(1+I_{model}(t)) \quad (5)$$

Figure 6 displays the distribution of logarithmic residuals over time for the improved GA-SEIR model. The figure reveals that residual values predominantly cluster around zero with random fluctuations, showing no discernible upward or downward trend. This indicates the absence of systematic bias throughout the fitting period. The low dispersion of residuals suggests high predictive stability across different time intervals.

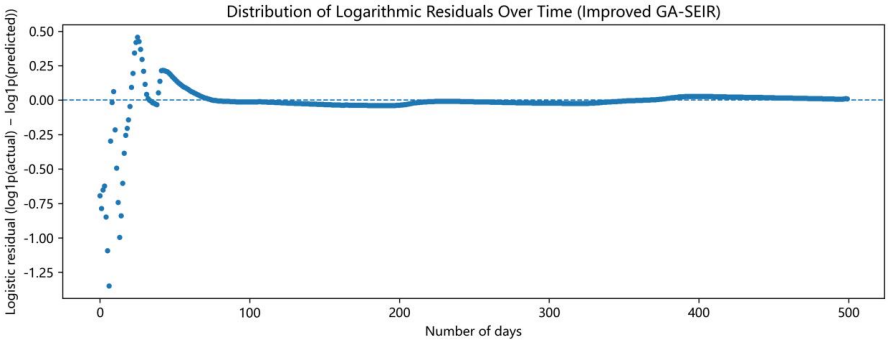

Figure 6. Logarithmic Residuals Scattered over Time.

Figure 7 shows the probability distribution histogram of residuals. It can be observed that the residual distribution exhibits an approximately unimodal structure, primarily concentrated within the interval [-0.25, 0.25], displaying an approximate normal distribution pattern. This indicates that the primary source of error in the model stems from random noise rather than structural inadequacies in the model. Furthermore, the logarithmic transformation significantly mitigates the impact of extreme values during the epidemic peak on overall error, enabling the model to exhibit consistent fitting performance both during low-infection phases and peak periods.

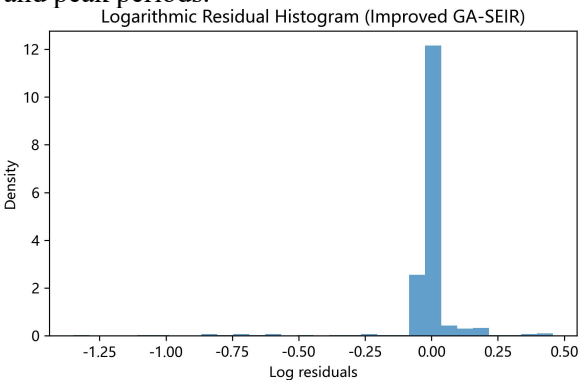

Figure 7. Logarithmic Residual Histogram

Based on the numerical results of the model, the mean of the log residuals is approximately $\mu \approx 0.00$, with a standard deviation of $\sigma \approx 0.18$. This indicates that the overall error fluctuation of the model is relatively small, and the prediction results are statistically close to unbiased. In other words, the discrepancy between the model's predicted values and the actual values is primarily caused by random disturbances rather than systematic biases inherent to the model mechanism itself.

In summary, the residual analysis demonstrates that the improved GA-SEIR model exhibits excellent randomness, stability, and unbiasedness. These findings further validate the model's reliability and generalization capability when applied to real-world epidemic data, providing robust support for enhancing the accuracy of its predictions.

4. Discussion

This study innovatively incorporates human self-isolation willingness into the traditional SEIR model and employs genetic algorithms (GA) for parameter optimization, aiming to enhance the model's predictive capability regarding epidemic transmission trends. Results demonstrate that the improved GA-SEIR model accurately captures the dynamic evolution of epidemic spread. This highlights the significance of incorporating human behavioral and psychological factors into epidemiological modeling, particularly when public perception and response to risk directly influence disease transmission rates, leading to markedly improved predictive accuracy.

The core mechanism of the improved model lies in the dynamic treatment of the contact rate. By employing a Sigmoid function to characterize the human self-isolation willingness factor and integrating it into the traditional SEIR model, the contact rate varies with the number of infected individuals, simulating the psychological and behavioral responses of the public at different stages of the epidemic. During the early epidemic phase, with fewer infections and low public vigilance, the contact rate remains elevated due to low self-isolation willingness. As infections increase and approach a behavioral threshold, heightened public awareness leads to a significant rise in isolation willingness, causing the actual contact rate to drop rapidly and forming an inflection point on the transmission curve. This mechanism reflects the nonlinear coupling between epidemic spread and human behavior while revealing the potential role of behavioral responses in controlling outbreak diffusion. Compared to traditional fixed-parameter models, the proposed dynamic contact rate model more realistically reproduces transmission processes under social behavioral interventions.

Furthermore, genetic algorithms (GA) play a crucial role in parameter optimization. GA, characterized by global search and adaptive learning, effectively avoids the problem of traditional optimization algorithms getting stuck in local optima. In this study, GA performed global optimization for key model parameters (such as β_0, σ, γ, k, θ), enabling the model to achieve optimal fitting in complex nonlinear spaces. Experimental results indicate that the optimized parameter combinations effectively reflect the dynamic trends observed in actual epidemics.

However, this study has certain limitations. First, the model's parameters for self-isolation willingness are primarily set based on changes in infection numbers, without fully considering multidimensional behavioral factors such as policy interventions, media discourse, and social trust. Second, the epidemic data used originates

from macro-level statistics, which may suffer from data delays and noise, potentially affecting the model's fine-grained fitting. Additionally, the genetic algorithm is sensitive to parameter ranges and population size during computation; improper settings may compromise convergence efficiency and the attainment of global optimal solutions.

Future research could enhance this model through the following avenues: First, integrating multi-source data such as social media sentiment and mobile travel trajectories to capture human behavioral shifts via multidimensional data-driven methods, thereby constructing a more refined behavioral response model; Second, exploring multi-objective optimization algorithms (e.g., NSGA-II) to simultaneously optimize prediction accuracy and computational efficiency. Third, validating the model's transferability and generalization capabilities across different regions and time periods to assess its applicability in diverse social structures and cultural contexts.

5. Conclusion

This study proposes an improved GA-SEIR model that incorporates human self-isolation willingness into the traditional SEIR framework and employs genetic algorithms (GA) for global optimization of key parameters. By integrating epidemiological dynamics with behavioral science, this model addresses the limitations of traditional SEIR models in capturing individual behavioral responses.

Empirical validation using COVID-19 data from China demonstrates that the improved GA-SEIR model exhibits highly consistent prediction curves with actual epidemic trends. Its Mean Absolute Percentage Error (MAPE) is 6.94%, and the coefficient of determination (R^2) reaches 0.986, indicating high accuracy and interpretability. Residual analysis indicates that model errors primarily stem from random fluctuations rather than structural biases, validating the model's stability and unbiasedness. This study employs a sigmoid function to characterize public self-isolation willingness, enabling dynamic adjustment of contact rates based on infection numbers. This approach more accurately reflects the nonlinear impact of sociopsychological feedback on epidemic transmission.

In summary, integrating human behavioral and psychological factors into infectious disease dynamics models significantly enhances prediction realism and decision-making guidance. The improved GA-SEIR model not only provides a more precise tool for epidemic forecasting but also offers a theoretical basis for developing behavior-response-based public health intervention strategies. Future research could further integrate multi-source behavioral data (e.g., population mobility, social media sentiment) and explore multi-objective optimization algorithms to achieve a better balance between prediction accuracy and computational efficiency.

6. Data Sources and Preprocessing

The data used in this study originates from the WHO COVID-19 Global Daily Data dataset published by the World Health Organization (WHO) (Source: https://data.who.int/dashboards/covid19/data).

This dataset systematically records daily confirmed, recovered, and deceased cases across all countries and regions worldwide from January 22, 2020, to the present. This paper selected epidemic data from China as the research subject for model validation and parameter optimization.

Given that the objective of this study is to validate the accuracy and stability of the improved GA-SEIR model rather than conduct long-term forecasting, a continuous time window since the outbreak onset was chosen to cover a complete epidemic transmission cycle. The cumulative number of confirmed cases was used in the model to approximate the number of infected individuals (I(t)).

Before model training, the data underwent the following preprocessing steps:

6.1 Time Series Organization and Cleaning

Daily records were sorted by reporting date to ensure continuity of the time series. Anomalous or missing data entries were imputed using forward filling to maintain data consistency.

6.2 Data Smoothing

To mitigate daily statistical fluctuations and reporting noise, a 3-day centered moving average smoothing was applied to the raw infection sequence, yielding a smoothed infection series. This method effectively suppressed the impact of outliers and stabilized the parameter search process of the genetic algorithm.

6.3 Missing Value Handling

Missing values at the beginning and end of the series were repaired using forward and backward imputation. The final smoothed dataset contains no missing records.

Following data cleaning, smoothing, and missing value handling, the resulting continuous time series data were used for global parameter fitting of the model. This ensured the optimized parameters fully captured the dynamic process of epidemic transmission. Simultaneously, the robustness of the algorithm was validated through multiple independent runs. Results demonstrated high consistency among parameters obtained across optimization iterations, confirming the stability and reliability of the model parameter estimates.

References

[1] Porta, Miquel S., et al., eds. A dictionary of epidemiology. Oxford University Press, 2014.

[2] WHO. COVID-19 Epidemic disease. https://www.who.int/emergencies/diseases/news.

[3] AJMC Staff. A Timeline of COVID-19 Developments in 2020. 2021. https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020.

[4] COVID-19 CORONAVIRUS PANDEMIC. https://www.worldometers.info/coronavirus/

[5] Qiu Z, Sun Y, He X, Wei J, Zhou R, Bai J, Du S. Application of genetic algorithm combined with improved SEIR model in predicting the epidemic trend of COVID-19, China. Sci Rep. 2022 May 26;12(1):8910. doi: 10.1038/s41598-022-12958-z. PMID: 35618751; PMCID: PMC9133826.

[6] Wongsathan, R. (2023). The logistic growth regression model with the genetic algorithm for predicting the third wave of the COVID-19 epidemic in Thailand. Asia-Pacific Journal of Science and Technology, 28(1), 1-16.

[7] Janes, C. R., Corbett, K. K., Jones, J. H., & Trostle, J. (2012). Emerging infectious diseases: the role of social sciences. The Lancet, 380(9857), 1884-1886.

[8] Du, E., Chen, E., Liu, J., & Zheng, C. (2021). How do social media and individual behaviors affect epidemic transmission and control?. Science of the Total Environment, 761, 144114.

[9] E.P. Fenichel, C. Castillo-Chavez, M.G. Ceddia, G. Chowell, P.A.G. Parra, G.J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, M. Springborn, L. Velazquez, & C. Villalobos, Adaptive human behavior in epidemiological models, Proc. Natl. Acad. Sci. U.S.A. 108 (15) 6306-6311, https://doi.org/10.1073/pnas.1011250108 (2011).

[10] Bedson, J., Skrip, L. A., Pedi, D., Abramowitz, S., Carter, S., Jalloh, M. F., ... & Althouse, B. M. (2021). A review and agenda for integrated disease models, including social and behavioural factors. Nature human behaviour, 5(7), 834-846.

[11] Manfredi, P., & D'Onofrio, A. (Eds.). (2013). Modeling the interplay between human behavior and the spread of infectious diseases. Springer Science & Business Media

[12] Wang, H., Wu, D., Luo, J., & Zhang, J. (2024). Integrating socio-psychological factors in the SEIR model optimized by a genetic algorithm for COVID-19 trend analysis. Scientific Reports, 14(1), 15684.

[13] Uncovering the Sigmoid Function Step by Step, by Paula Livingstone on Jan. 17, 2024, https://paulalivingstone.com/blog/post/uncovering-sigmoid-function-step-step/

[14] Okawa, Maya, and Tomoharu Iwata. "Predicting opinion dynamics via sociologically-informed neural networks." Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. 2022.

[15] Kleden, A. (2025). Modeling Moral Decision-Making: A Sigmoid-Based Framework for Conative and Inhibitory Forces.

