Journal of
Information Systems Engineering
and Business Intelligence

Vol.10, No.2, June 2024
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Comparison of Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for Estimating the Susceptible-Exposed-Infected-Recovered (SEIR) Model Parameter Values

Aminatus Sa'adah [1)*], Ayomi Sasmito [2)], Asysta Amalia Pasaribu [3)]

[1)]*Informatics Engineering, Faculty of Infomatics, Institut Teknologi Telkom Purwokerto, Banyumas, Indonesia*
[1)]aminatus@ittelkom-pwt.ac.id

[2)]*Business Mathematics, School of STEM, Universitas Prasetiya Mulya, Tangerang, Indonesia*
[2)]ayomi.sasmito@pmbs.ac.id

[3)]*Statistic, School of Computer Science, Universitas Bina Nusantara, Jakarta, Indonesia*
[3)]asysta.amalia@binus.ac.id

Abstract

Background: The most commonly used mathematical model for analyzing disease spread is the Susceptible-Exposed-Infected-Recovered (SEIR) model. Moreover, the dynamics of the SEIR model depend on several factors, such as the parameter values.
Objective: This study aimed to compare two optimization methods, namely genetic algorithm (GA) and particle swarm optimization (PSO), in estimating the SEIR model parameter values, such as the infection, transition, recovery, and death rates.
Methods: GA and PSO algorithms were compared to estimate parameter values of the SEIR model. The fitness value was calculated from the error between the actual data of cumulative positive COVID-19 cases and the numerical data of cases from the solution of the SEIR COVID-19 model. Furthermore, the numerical solution of the COVID-19 model was calculated using the fourth-order Runge-Kutta algorithm (RK-4), while the actual data were obtained from the cumulative dataset of positive COVID-19 cases in the province of Jakarta, Indonesia. Two datasets were then used to compare the success of each algorithm, namely, Dataset 1, representing the initial interval for the spread of COVID-19, and Dataset 2, representing an interval where there was a high increase in COVID-19 cases.
Results: Four parameters were estimated, namely the infection rate, transition rate, recovery rate, and death rate, due to disease. In Dataset 1, the smallest error of GA method, namely 8.9%, occurred when the value of $mutrate = 0.5$, while the numerical error of PSO was 7.5%. In Dataset 2, the smallest error of GA method, namely 31.21%, occurred when $mutrate = 0.5$, while the numerical error of PSO was 3.46%.
Conclusion: Based on the parameter estimation results for Datasets 1 and 2, PSO had better fitting results than GA. This showed PSO was more robust to the provided datasets and could better adapt to the trends of the COVID-19 epidemic.

Keywords: Genetic algorithm, Particle swarm optimization, SEIR model, COVID-19, Parameter estimation.

Article history: Received 12 February 2024, first decision 17 May 2024, accepted 20 June 2024, available online 28 June 2024

I. INTRODUCTION

Mathematical models are valuable tools for studying and exploring the dynamics of real-world problems, such as the spread of infectious diseases. In this context, the most commonly used mathematical model for this purpose is the Susceptible-Exposed-Infected-Recovered (SEIR) model [1], [2], [3]. This model was first developed by Kermark and McKendrick in 1927, and it describes the spread of infectious diseases through four compartments. Moreover, the model has evolved to describe the dynamics of various diseases, including COVID-19 (Coronavirus Disease 2019) [4], dengue [5], cancer [6], diphtheria [7], and tuberculosis [8]. The SEIR COVID-19 model used in most disease spread modeling is a physiological model developed based on assumptions and simplifications of the spread in actual conditions.

Accurate data are important for obtaining model parameters that can effectively describe the dynamics of a disease [9]. These data can be used to estimate parameter values within the model using viable approaches like genetic algorithm (GA) [10], [11] and particle swarm optimization (PSO) [12][13]. GA is an optimization and search method based on genetics and natural selection principles [14], [15], [16]. Meanwhile, PSO is based on the behavior of flocks

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.10.2.290-301

of birds and schools of fish. It optimizes by finding the optimum solution based on the position and speed of the particles, influenced by the position of the group (swarm) [12]. Several studies have used both methods to estimate the parameters of epidemic mathematical model. GA has been specifically used to find optimal parameters, ensuring that the model accurately represents the dynamics of COVID-19 based on available data from different countries [10], [11], [17], [18], with PSO serving the same purpose [19][20]. PSO has also been used to estimate parameters in Lotka-Volterra model based on annual profits from commercial and rural banks [21]. According to [22], GA was used to estimate parameters in dengue model using weekly dengue case data in Indonesia.

Analyzing the dynamics of the model is crucial for developing effective disease control strategies. The dynamics depend on several factors, with one of the most essential being the value of the model parameters [23]. Parameter values that are specific to a location or certain time interval can provide a detailed understanding of disease patterns in a specific area or time. The dynamical analysis of the model in [24] and [25] has not been optimized in terms of parameter. In this context, the parameter values are obtained from assumptions or references from other studies, and not from precise fitting or parameter estimation. This lack of specificity can lead to less accurate disease control strategies when applied in a specific location [26]. Therefore, identifying the best algorithm for estimating the SEIR model parameter is crucial for improving analysis results.

The current study provided a comprehensive explanation of the procedural steps in using GA and PSO for parameter estimation of the SEIR COVID-19 model, as well as focused on comparing the result in different dataset. Exploring different datasets can facilitate deeper understanding of the health requirements of the local population. COVID-19 has varying behavior in different regions, and insights gained from this study can inform healthcare policies and strategies. Furthermore, access to detailed COVID-19 data from certain region improves the precision and reliability of results. Both GA and PSO were used to explain the procedural steps for estimating key parameters within the model. COVID-19 dataset was fitted to the model, which was solved numerically using Runge-Kutta, a popular and effective method for solving nonlinear ordinary differential equations. The four critical parameters estimated are infection rate (β), transition rate (α), recovery rate (ε), and mortality rate attributed to the disease (μ_1). The resulting parameter values were validated against the developed model to ensure accuracy and applicability.

II. METHODS

This section described the development of the SEIR mathematical model and parameter estimation using GA and PSO based on the cumulative cases of COVID-19 in DKI Jakarta. Four parameters of the model were estimated, namely infection rate (β), transition rate (α), recovery rate (ε), and death rate (μ_1). The values of other parameters are provided in Table 2. Meanwhile, the solution for the model was obtained using the fourth-order Runge-Kutta algorithm (RK-4) with the help of the ode45 solver in MATLAB. The stability and convergence of RK-4 algorithm are discussed in [27]. A summary of the study flow is provided in Fig. 1.

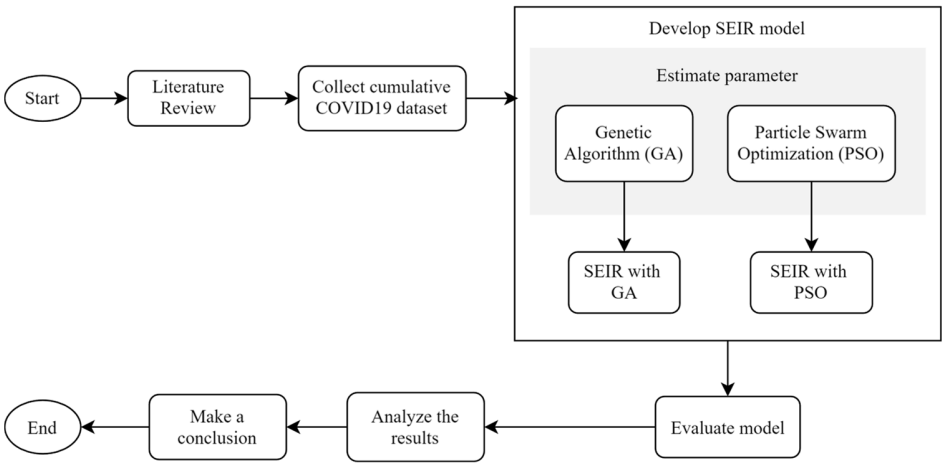


Fig. 1 Research method diagram

A. Dataset

This study used two datasets, representing the cumulative number of positive COVID-19 cases. The data were publicly accessible on the website https://corona.jakarta.go.id/id, which provided daily updates on COVID-19 cases in Indonesia, including the number of recoveries, deaths, and exposures. Two different datasets were used, namely

Dataset 1, covering cumulative cases from April 1st to August 31st, 2020, and Dataset 2, spanning from March 23rd to July 31st, 2021. Dataset 1 represents the initial outbreak of the first variant of the coronavirus, while Dataset 2 represents the period with the highest peak of the second variant. Brief descriptions of the datasets are presented in Tables 1 and 2. Cumulative data of positive cases were chosen over active case data as cumulative data tended to have less volatility, facilitating the identification of patterns.

TABLE 1
DATASET 1

Date	The number of cumulative COVID-19 cases (People)
01/04/2020	816
02/04/2020	909
03/04/2020	990
04/04/2020	1071
05/04/2020	1151
⋮	⋮
31/08/2020	40309

TABLE 2
DATASET 2

Date	The number of cumulative COVID-19 cases (People)
23/03/2021	372871
24/03/2021	373761
25/03/2021	375487
26/03/2021	376868
27/03/2021	378222
⋮	⋮
31/07/2021	814635

B. The SEIR Mathematical Model

The model categorized human population into four compartments, namely susceptible (S), exposed (E), infected (I), and recovered (R) [8]. The susceptible compartment increased at the recruitment rate Λ and decreased at the natural death rate μS. Susceptible individuals became exposed at an infection rate $\beta SI/N$. The exposed compartment transitioned to the infected compartment at a recovery rate εI. The exposed and recovered compartments decreased due to natural death at μR and μE, rates respectively. Similarly, the infected individuals could die at a rate of $\mu_1 I$. The transmission diagram of the model is presented in Fig. 2. Based on this explanation, mathematical model was formulated as follows:

$$\begin{aligned}
\frac{dS}{dt} &= \Lambda - \frac{\beta SI}{N} - \mu S, \\
\frac{dE}{dt} &= \frac{\beta SI}{N} - \alpha E - \mu E, \\
\frac{dI}{dt} &= \alpha E - \varepsilon I - \mu_1 I, \\
\frac{dR}{dt} &= \varepsilon I - \mu R.
\end{aligned} \tag{1}$$

where $S(t), E(t), I(t), R(t) \geq 0$, for each $t \geq 0$. Compartment descriptions and parameters are presented in Tables 3 and 4.

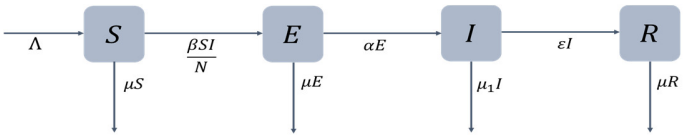


Fig. 2 Transmission diagram of the SEIR model

TABLE 3
MODEL COMPARTMENTS DESCRIPTION

Notation	Description	Unit
$S(t)$	The number of susceptible individuals at time t	People
$E(t)$	The number of exposed individuals at time t	People
$I(t)$	The number of infected individuals at time t	People
$R(t)$	The number of recovered individuals at time t	People

TABLE 4
MODEL PARAMETERS DESCRIPTION

Notation	Description	Value	Unit
Λ	Recruitment rate	$\frac{N}{365\times74.8}$	People×day^{-1}
β	Infection rate	Estimated	Day^{-1}
α	Transition rate	Estimated	Day^{-1}
ε	Recovery rate	Estimated	Day^{-1}
μ	Natural death rate	$\frac{1}{365\times74.8}$	Day^{-1}
μ_1	Death rate due to disease	Estimated	Day^{-1}

C. Genetic Algorithm

GA was first introduced by Haupt and Haupt in 1975 [28]. It comprises four key components, namely chromosome population (individual), parent selection based on fitness value, crossover to produce offspring and random mutation. Pair selection entails choosing two chromosomes from a collection of prospective parent individuals to produce two new individuals (offspring). Crossing over is the formation of one or two new individuals (offspring) from the selected parent individuals. Finally, mutation entails changing one or several individuals in the population with new individuals [29]. The pseudocode of GA for estimating the SEIR model parameters is presented in Table 5.

Algorithm 1
Genetic Algorithm

Initialize parameters:
$npop, nvar, ngen, xrate, mutrate, a, b$

Generate initial population:
  for each chromosome in population:
    initialize genes with $N(0,1)$
    transform genes to interval $[a, b]$
  Repeat for ngen generations:
    Calculate the SEIR model using Runge-Kutta method
    Select parents:
      $n\_endure = round(npop * xrate)$
      Create rank order
      Calculate cumulative probability matrix
    Evaluate fitness value for each chromosome:
      for each chromosome in population:
        Calculate RMSE using actual and numerical data
    Create pairs of parent individuals:
      for each pair of parents:
        Apply linear combination crossover to create two children
    Apply elitism by preserving best chromosome
    Apply mutation to selected chromosomes:
      for each chromosome with mutation:
        Replace selected genes with random values
    Create new population by combining elitism and mutated chromosomes
    Select chromosome with smallest fitness value
    Check termination condition (reach $ngen$ *iterations)*
    when termination condition is met:
      Display the solution
  End loop

The first step in using GA for parameter estimation is the initialization of parameters for both GA and the SEIR model. The parameter values for GA include population size ($npop$), number of genes/individuals/chromosomes ($nvar$), number of generations or iterations ($ngen$), probability of crossing over ($xrate$), and probability of mutation ($mutrate$). Meanwhile, the parameter values include $\Lambda, \mu_1, S(0), E(0), I(0), R(0)$. A random population of chromosomes was generated using normal distribution with a mean of zero and a variance of one, $N(0,1)$, order

$(npop \times nvar)$. These values were subsequently transformed into lower and upper bound intervals (a, b) using the formula:

$$pop_{new} = a + (b - a) \times rand(0,1) \quad (2)$$

The chromosomes formed at this stage served as the initial population.

The second step was parent selection, where several chromosomes in the population were selected to become subpopulations of parent individuals. The number of parent individuals that survived to become a subpopulation was determined by the formula:

$$n\_endure = round(npop \times xrate). \quad (3)$$

Create a rank order from $nbertahan$ to one and count $cumulative\ rank = \frac{n\_endure \times (nbertahan+1)}{2}$. Create an ordered cumulative probability matrix $(nbertahan \times 1)$ with the k^{th}-element as:

$$prob_{cum}(k) = \frac{\sum_{x=1}^{k} rank(x)}{cumulative\ rank}, \quad (4)$$

where $rank(x) = n\_endure + 1 - x$. The value $\frac{rank(x)}{cumulative\ rank}$ states the probability of the ranked individual being selected as a parent.

The fitness value (objective function) for each chromosome in the population was subsequently evaluated. In the SEIR model (1), the cumulative positive cases of COVID-19 are the sum of infected, recovered, and deceased individuals. Therefore, new compartments were defined, namely death (D) and cumulative (K), where

$$K(t) = I(t) + R(t) + D(t) \quad (5)$$

$$\frac{dD}{dt} = \mu_1 I \quad (6)$$

$$\frac{dK}{dt} = \frac{dI}{dt} + \frac{dR}{dt} + \frac{dD}{dt} = \alpha E - \mu R \quad (7)$$

The fitness value for this parameter estimation was calculated using Mean Absolute Percentage Error (MAPE) in the following equation:

$$e = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^* - y_i}{y_i^*}\right|, \quad (8)$$

$$y(t) = K(t), \quad (9)$$

where y_i represents the cumulative actual data of positive COVID-19 cases on day-i and y_i^* represents the numerical data of the cumulative ode45 positive cases of COVID-19 on day-i . The numerical value of $K(t)$ was calculated using RK-4 method with the ode45 solver package in MATLAB software.

The third step is cross over, which entails the formation of one or two new individuals (offspring) from the selected parent individuals. It started with the selection of several pairs from the subpopulation of parent individuals. Crossing over was conducted on these pairs to produce new offspring. These offspring were subsequently added to the chromosome population. The cross over method used in this study was a linear combination, requiring a random number γ with $0 < \gamma < 1$. Two offspring can be obtained using the linear combination of parent individuals with the following formula:

$$c = \gamma a + (1 - \gamma)b, \quad (10)$$

$$d = (1 - \gamma)a + \gamma b. \quad (11)$$

The fourth step is elitism, which aims to separate the chromosome with the best fitness value, ensuring it does not experience mutations. The fifth step is mutation, where several chromosomes were randomly mutated by replacing selected genes with random values. This stage explored the solution space to obtain the most optimal solution. Six different mutation rates were used in this study, namely 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5 [5]. Subsequently, seven experiments were conducted for each value of $mutrate$ and the parameter model value with the smallest error was selected for each $mutrate$. Seven experiments were conducted for each $mutrate$, and the parameter model value with the smallest error was selected for each rate. The final step is the formation of a new population by combining the chromosomes resulting from elitism and those from the mutation process. The chromosome with the smallest fitness value was chosen, and the iteration continued until the stopping criteria was met, namely maximum iteration.

D. Particle Swarm Optimization

PSO explanation is based on [30] and [31]. The algorithm is a computational technique that addresses optimization issues using a meta-heuristic approach. Inspired by the collective behaviour of natural phenomena such as flocks of

birds or schools of fish, this method was first developed by Kennedy and Eberhart in 1995. In PSO, each particle or agent represents the coordinates of different points within the solution space during each iteration. The algorithm can thoroughly explore the entire functional landscape and identify both local and global solutions. The specific outcome, whether local or global, depends on the design of the solution and the available computational resources. In the current study, PSO was used to estimate parameter values in the model for the spread of COVID-19 based on cumulative data of positive COVID-19 cases in DKI Jakarta. The optimal solution is the one that best fits the system of differential equations (the SEIR model) to the COVID-19 dataset and the given initial values.

A swarm is defined as follows:

$$Swarm = \{x_1, x_2, \dots, x_i | x_i \in (I_1, I_2, \dots, I_n)\} \quad (12)$$

where x_i are the particles that make up the swarm in the interval range limited by $I_1, I_2, \dots, I_n$. These particles move to find solution parameters that minimize the objective function. The objective function in PSO is the same as in GA, using equation (8). The objective function was first calculated for each particle, and subsequently compared across all particles to identify the best solution, becoming the swarm's guide. This process repeats until the stopping criteria are met.

Next, each particle moves toward the best solution with speed v_i. The velocity vector for each particle was calculated using the following equation (13):

$$v_i(t+1) = c_0 v_i(t) + c_1 r_i^1 \left(x_i^{pbest} - x_i(t)\right) + c_2 r_i^2 \left(x^{gbest} - x_i(t)\right). \quad (14)$$

The velocity vector comprises three distinct components, namely an inertial component denoted as c_0, responsible for preserving the particle's forward motion; a learning component denoted as c_1, incorporating a random term r_i^1 to guide the particle toward its personal best position x_i^{pbest}; and a global learning component denoted as c_2, incorporating a random term r_i^2 to guide the particle toward the best solution discovered by the whole group, denoted as x^{gbest}. The velocity vector dynamically adjusted the collective velocity of the swarm. Once the speed vector has been computed, the subsequent procedure is to update the position of each particle, denoted as x_i, based on the corresponding speeds. This update was conducted using the following formula:

$$x_i(t+1) = x_i(t) + v_i(t+1). \quad (15)$$

Algorithm 2
Particle Swarm Optimization (PSO)

Input: $SwarmSize, MaxIt, d, c_1, c_2, ub, lb$
Output: x^{gbest}, F
Random initialization of population
for each t $(t \in MaxIt)$ *do,*
  for each i $\left(i \in N_{pop}\right)$ *do,*
    for each j $(j \in d)$ *do,*
      velocity update by Eq (14)
      position update by Eq (15)
    end
    fitness evaluation $\left(F(x_i) = J(x_i)\right)$
    If $F\left(x_i(t+1)\right) < F\left(x_i^{pbest}(t)\right)$ *then,*
      $x_i^{pbest}(t) = x_i(t+1)$
    end
    If $F\left(x_i^{pbest}(t)\right) < F(x^{gbest})$ *then,*
      $x^{gbest} = x_i^{pbest}(t)$
    end
  end
end

E. Evaluation

MAPE was used to evaluate the parameter value estimation results of the SEIR model from GA and PSO algorithms. In the model (1), the cumulative positive cases of COVID-19 are the sum of infected, recovered, and deceased individuals. Therefore, new compartments were defined, namely death (D) and cumulative (K), where

$$K(t) = I(t) + R(t) + D(t) \quad (10)$$

$$\frac{dD}{dt} = \mu_1 I \tag{11}$$

$$\frac{dK}{dt} = \frac{dI}{dt} + \frac{dR}{dt} + \frac{dD}{dt} = \alpha E - \mu R \tag{12}$$

The fitness value for this parameter estimation was calculated using MAPE, as shown in the following equation:

$$e = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i^* - y_i}{y_i^*} \right| \times 100\%, \tag{13}$$

$$y(t) = K(t), \tag{14}$$

where y_i represents the cumulative actual data of positive cases of COVID-19 on day-i and y_i^* represents the numerical data of the cumulative ode45 positive cases of COVID-19 on day-i . The numerical value of $K(t)$ was calculated using RK-4 method with the ode45 solver package in MATLAB software, using parameters estimated by GA and PSO. The stability and convergence of RK-4 algorithm are detailed in [30]. Both algorithms can be compared to determine which is more effective for parameter estimation under different conditions or data quality.

III. Results

In this section, GA and PSO were compared to estimate the parameter values in the COVID-19 model ($\beta, \alpha, \varepsilon,$ and μ_1) for cases in DKI Jakarta, Indonesia. First, GA and PSO were implemented on Dataset 1, with the best estimated parameter values given in Table 5. Based on Table 5, the smallest error for GA method, namely 8.90%, occurred when the value of $mutrate = 0.1$, while the numerical error of PSO was 7.5%. The best parameter values for the estimation results from GA and PSO are presented in Table 6. In Dataset 1, the estimation value of infection rate (β) was 0.0425, transition rate (α) was 0.1123, recovery rate (ε) was 0.0089, and death rate (μ_1) was 0.0123. Furthermore, the model and Dataset 1 were validated using the model in system (1). Figure 3a shows the comparison between the model and Dataset 1. This graphic explains the general trend of under-predicting cases and over-predicting casualties over time. During the observation time interval, PSO provided better-fitting results than GA. For the first 120 days of observations, PSO performed excellently, and even in the last 30 days of observation, PSO still outperformed GA.

TABLE 5
Best Parameter Value for Dataset 1 using Particle Swarm Optimization (PSO) and Genetic Algorithm (GA)

Algorithm	β	α	ε	μ_1	Error
GA (mutrate = 0.05)	0.3447	0.5366	0.0571	0.2706	10.11%
GA (mutrate = 0.1)	0.1908	0.5028	0.0268	0.1431	8.90%
GA (mutrate = 0.2)	0.3490	0.9482	0.0397	0.2936	10.74%
GA (mutrate = 0.3)	0.3690	0.3157	0.0686	0.2742	9.88%
GA (mutrate = 0.4)	0.4857	0.3573	0.0822	0.3766	10.31%
GA (mutrate = 0.5)	0.4042	0.6919	0.0351	0.3444	9.75%
PSO	0.0425	0.1123	0.0089	0.0123	7.50%

TABLE 6
Estimation Parameter Value for Dataset 1

Parameter	Value (GA)	Value (PSO)
β	0.1908	0.0425
α	0.5028	0.1123
ε	0.0268	0.0089
μ_1	0.1431	0.0123

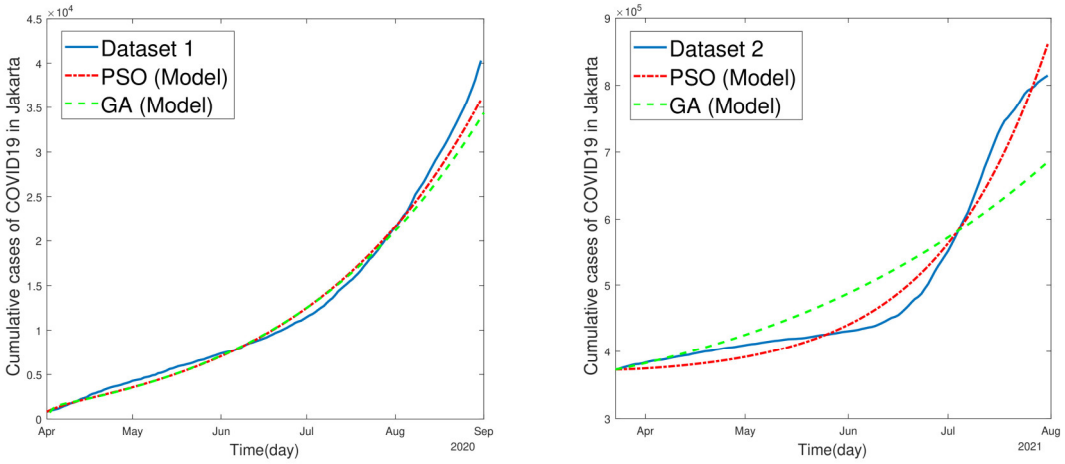
Second, GA and PSO were implemented on Dataset 2, with the best estimated parameter values presented in Table 7. Based on Table 7, the smallest error for GA method, namely 31.21%, occurred when the value of $mutrate = 0.5$, while the numerical error of PSO was 3.46%. The best parameter values for the estimation results from GA and PSO are presented in Table 8. In Dataset 2, the estimation value of infection rate (β) was 0.4987, transition rate (α) was 0.0132, recovery rate (ε) was 0.0541, and death rate (μ_1) was 0.0535. Furthermore, the model and Dataset 2 were validated using the SEIR model in system (1). Figure 3b shows the comparison between the model and Dataset 2. This graphic explains the general trend of under-predicting cases and over-predicting casualties over time. In Figure 3b, PSO produced superior fitting results than GA, despite neither algorithm being able to achieve a fit almost identical to Dataset 2.

TABLE 7
BEST PARAMETER VALUE FOR DATASET 2 USING PARTICLE SWARM OPTIMIZATION (PSO) AND GENETIC ALGORITHM (GA)

Algorithm	β	α	ε	μ_1	Error
GA (mutrate = 0.05)	0.3215	0.2008	0.1123	0.1672	31.79%
GA (mutrate = 0.1)	0.3534	0.2125	0.0967	0.2112	31.92%
GA (mutrate = 0.2)	0.1868	0.0623	0.0485	0.0764	31.58%
GA (mutrate = 0.3)	0.2980	0.1960	0.1359	0.1232	31.77%
GA (mutrate = 0.4)	0.1630	0.1052	0.0567	0.0747	31.27%
GA (mutrate = 0.5)	0.1929	0.1209	0.0485	0.1095	31.21%
PSO	0.4987	0.0132	0.0541	0.0535	3.46%

TABLE 8
ESTIMATION PARAMETER VALUE FOR DATASET 2

Parameter	Value (GA)	Value (PSO)
β	0.1929	0.4987
α	0.1209	0.0132
ε	0.0485	0.0541
μ_1	0.1095	0.0535



(a) (b)

Fig. 3 The validation of (a) Dataset 1 and (b) Dataset 2 with model using PSO and GA

IV. DISCUSSION

Optimized parameters can provide an accurate picture of the rate of disease spread in a population. In Dataset 1, representing the initial interval for the spread of COVID-19, the parameters estimated using PSO method provided smaller numerical error values than GA. However, the difference in error values needed further investigation. The second parameter results from PSO and GA accurately followed the dataset trend. Even at certain early time points, the validation graphs for both methods had the same value. This differed from the results obtained by [10], using weighted fitness by including the element $ln(t)$ in the objective function. The current study showed that the model and data agreed well over the last 30 days, likely because the fitness function weighed more heavily on recent data by design. In Dataset 2, where there was a high increase in cases, the numerical error results from PSO method provided much better fitting results than GA. This showed PSO was more robust to the dataset and could better adapt to the trends of the COVID-19 epidemic.

The SEIR COVID-19 model used in most disease spread modeling is not typically based on data but a physiological model built on assumptions and simplifications of actual disease spread conditions. When no dataset is available to obtain parameter values, model analysis can still be carried out using assumed parameter values sourced from medical literature. However, when there is a dataset related to the number of infected individuals, the model parameter values can be estimated based on that dataset. The estimated parameter values depend on the dataset used, and each dataset (varying by time or location) can produce different estimated parameter values. This approach provides a specific picture regarding the rate of disease spread in the location and time from which the dataset originates.

The parameter values estimated from the two datasets were subsequently interpreted. Dataset 1 represented the period of the initial outbreak of the first coronavirus variant, while Dataset 2 represented the period of the highest peak of the second variant. The COVID-19 infection rate from Dataset 1 was 0.0425, while 0.4987 was obtained for Dataset 2. Also, the COVID-19 recovery rate value from Dataset 1 was 0.0089, while 0.0541 was obtained for Dataset 2. This showed the spread rate of the second coronavirus variant was eleven times higher compared to the first coronavirus variant. However, the fast spread was offset by a recovery rate for the second variant that was approximately six times faster compared to the first variant. This showed that while the second variant spread, it also had a higher chance of recovery. Conversely, the first variant spread more slowly but had a lower chance of recovery. This information was crucial for informing policies related to treatment types and disease control strategies.

A similar estimation of COVID-19 model parameters using GA was carried out by [10], using datasets from early March to May 2020 in six cities, namely Washington DC, GA, Michigan, New York, PA, and VA. The current study produced estimated recovery rate values of 0.18, 0.13, 0.10, 0.10, 0.11, and 0.20, respectively. Averaging these values produced a recovery rate of 0.13, varying significantly from the estimation results of this study for a similar time interval (Dataset 1), namely 0.0089. However, compared to the results from [17], which also used a dataset of COVID-19 cases in Indonesia for the same interval as Dataset 1, the estimated recovery rate was 0.0083. This value corresponded with the results of the current study, showing COVID-19 spread differently in each location. The variance could be influenced by several factors, such as weather, genetics, and social behavior.

In the GA scheme, as presented in Tables 5 and 7, there was no correlation between the $mutrate$ value and the resulting error value. In terms of computation time, PSO required less time than GA. In the GA scheme, the algorithm needed to be run several times for different $mutrate$ values or varying $npop, nvar,$ and $maximum\ iteration$ values to achieve the best results. For instance, selecting five different $mutrate$ and running each seven times, necessitated running the algorithm 35 times to achieve the best result. Meanwhile, with PSO, it was crucial to vary the $npop, nvar,$ or $maximum\ iteration$ values. In the current study, PSO provided better results than GA. According to [17], the parameters of COVID-19 SEIR model were estimated using GA algorithm. The study also used a dataset from the spread of COVID-19 in Indonesia during the initial period of cases, from April 15 to August 24, 2020. The lowest error value obtained was 13.17% when the $mutrate = 0.125$. At the same time interval, represented by Dataset 1, better results were obtained, with a smallest error of 8.9% when $mutrate = 0.1$. The best $mutrate$ value obtained was smaller compared to that in [17], with PSO algorithm producing the smallest error, around 7.50%.

The challenge of the current study was that both PSO and GA failed to estimate the model parameter values in Dataset 2 accurately. For datasets with exponential data trends, like in Dataset 1, PSO provided excellent estimation results with relatively minimal errors. On the other hand, GA required many trials to find the mutation rate with the least error. The study compared GA and PSO algorithms to estimate the SEIR model parameter values at different time intervals. Future studies could expand this by including data on the spread of COVID-19 from several regions or by using other estimation methods.

The model had been widely applied, especially for COVID-19 cases, in various countries like China [11], [32], Bangladesh [33], Sri Lanka [34], the United States [35], France [36], and England [37]. While each country had unique characteristics like culture, population and certain events, the model generally provided satisfactory estimates. This was also confirmed by the current study, focusing on datasets from Indonesia, especially in Jakarta. To gain a deeper understanding of the model results, it was crucial to consider specific local characteristics. For example, social distancing and cloth face coverings effectively reduced the spread of the virus [35], overcrowded population increased susceptibility rates [33], and the current study examined the impact of the first variant's outbreak and the peak of the second variant of COVID19 in Jakarta. This approach allowed for comparisons of epidemics and the extent of prevention and control in various locations at different times [11].

Studies have also explored various methods to improve the SEIR model, such as using cluster analysis [32], hybrid modeling [38], generic algorithms [10], [11], multiple regression [33], and machine learning [39]. Improvements could be made by applying Artificial Intelligence and Deep Learning to improve prediction results [40], [41]. Besides the algorithmic perspective, future studies could explore integrating online news data [42], [43] to gain insights from a phenomenon point of view and incorporating sentiment analysis [44]. Another valuable approach was to analyze data from a system dynamic perspective [45].

V. CONCLUSIONS

In conclusion, GA and PSO were used to estimate parameter values in COVID-19 model using a dataset of cumulative positive COVID-19 cases. The fitness value was calculated from the error between the actual data of cumulative positive COVID-19 cases and the numerical solution of the SEIR COVID-19 model using MAPE formula. Based on estimation results, PSO provided better fitting results than GA, with smaller error and minimum

computational time. Furthermore, PSO fit the model better when the dataset had exponential trends without fluctuations or periodicals. Using the parameter values obtained from the cumulative COVID-19 case dataset, numerical simulation could be performed to study the dynamics of the disease spread in a population, facilitating policy-making for disease control. For future studies, incorporating a dataset representing the susceptible and exposed compartments could help obtain parameter values that better describe the spread of the disease.

Author Contributions: *Aminatus Sa'adah*: Conceptualization, Methodology, Formal analysis, Project administration. *Ayomi Sasmito*: Resources, Formal analysis, Writing – Original Draft, *Asysta Amalia Pasaribu*: data curation, Writing - Review & Editing, Visualization.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The data is open publicly and can be accessed on the website https://corona.jakarta.go.id/id.

Institutional Review Board Statement: Not applicable.

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

ORCID:
Aminatus Sa'adah: https://orcid.org/0000-0002-3469-7745
Ayomi Sasmito: -
Asysta Amalia Pasaribu: -

REFERENCES

[1] F. A. Muqtadiroh *et al.*, "Fuzzy Unsupervised Approaches to Analyze Covid-19 Spread for School Reopening Decision Making," *IECON Proc. (Industrial Electron. Conf.*, vol. 2021-Octob, pp. 1–7, 2021, doi: 10.1109/IECON48115.2021.9589699.

[2] A. A. Suwantika, I. Dhamanti, Y. Suharto, F. D. Purba, and R. Abdulah, "The cost-effectiveness of social distancing measures for mitigating the COVID-19 pandemic in a highly-populated country : A case study in Indonesia," *Travel Med. Infect. Dis.*, vol. 45, no. December 2021, p. 102245, 2022, doi: 10.1016/j.tmaid.2021.102245.

[3] H. Amir, S. Sudarman, A. Asfar, and A. S. Batara, "Covid19 Pandemic: Management and Global Response," *J. Kesehat. Lingkung.*, vol. 12, no. 1si, p. 121, 2020, doi: 10.20473/jkl.v12i1si.2020.121-128.

[4] D. Aldila, M. Z. Ndii, N. Anggriani, H. Tasman, and B. D. Handari, "Impact of social awareness , case detection , and hospital capacity on dengue eradication in Jakarta : A mathematical model approach," *Alexandria Eng. J.*, vol. 64, pp. 691–707, 2023, doi: 10.1016/j.aej.2022.11.032.

[5] X. Li, L. Cai, M. Murshed, and J. Wang, "Dynamical analysis of an age-structured dengue model with asymptomatic infection," *J. Math. Anal. Appl.*, vol. 524, no. 2, p. 127127, 2023, doi: 10.1016/j.jmaa.2023.127127.

[6] A. Sa'adah, D. A. Kamil, and G. E. Setyowisnu, "Modeling the viral dynamics of SARS-CoV-2 infection on tumor-immune system treated by chemotherapy," in *AIP Conference Proceedings*, 2022, p. 020004. doi: 10.1063/5.0091002.

[7] M. Ghani, I. Qutsiati, U. Fadillah, W. Triyayuda, and M. Afifah, "A fractional SEIQR model on diphtheria disease," *Model. Earth Syst. Environ.*, vol. 9, no. 2, pp. 2199–2219, 2023, doi: 10.1007/s40808-022-01615-z.

[8] K. Das, B. S. N. Murthy, S. A. Samad, and M. H. A. Biswas, "Mathematical transmission analysis of SEIR tuberculosis disease model," *Sensors Int.*, vol. 2, no. April, p. 100120, 2021, doi: 10.1016/j.sintl.2021.100120.

[9] M. A. Abdoon, R. Saadeh, M. Berir, F. EL Guma, and M. ali, "Analysis, modeling and simulation of a fractional-order influenza model," *Alexandria Eng. J.*, vol. 74, pp. 231–240, 2023, doi: 10.1016/j.aej.2023.05.011.

[10] P. Yarsky, "Using a genetic algorithm to fit parameters of a COVID-19 SEIR model for US states," *Math. Comput. Simul.*, vol. 185, pp. 687–695, Jul. 2021, doi: 10.1016/j.matcom.2021.01.022.

[11] Z. Qiu *et al.*, "Application of genetic algorithm combined with improved SEIR model in predicting the epidemic trend of COVID-19, China," *Sci. Rep.*, vol. 12, no. 1, pp. 1–9, 2022, doi: 10.1038/s41598-022-12958-z.

[12] D. Akman, O. Akman, and E. Schaefer, "Parameter Estimation in Ordinary Differential Equations Modeling via Particle Swarm Optimization," *J. Appl. Math.*, vol. 2018, 2018, doi: 10.1155/2018/9160793.

[13] A. Eka, W. Widianto, K. A. Ms, and V. R. Tjahjono, "Penentuan Effective Reproduction Number COVID-19 dengan Metode Particle Swarm Optimization pada Enam Provinsi di Pulau Jawa," *J. Math. its Apl.*, vol. 20, no. 2, pp. 131–143, 2023, doi: 10.12962/limits.v20i2.8585.

[14] D. Rahmalia, T. Herlambang, and T. E. Saputro, "Fertilizer Production Planning Optimization Using Particle Swarm Optimization-Genetic Algorithm," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 5, no. 2, p. 120, Oct. 2019, doi: 10.20473/jisebi.5.2.120-130.

[15] I. Fadah, A. Elliyana, Y. A. Auliya, Y. Baihaqi, M. Haidar, and D. M. Sefira, "A Hybrid Genetic-Variable Neighborhood Algorithm for Optimization of Rice Seed Distribution Cost," *Math. Model. Eng. Probl.*, vol. 9, no. 1, pp. 36–42, Feb. 2022, doi: 10.18280/mmep.090105.

[16] D. Herawatie, E. Wuryanto, and F. Jie, "Course scheduling using Modified Genetic Algorithm in vocational education," *Int. J. Oper. Quant. Manag.*, vol. 24, no. 3, pp. 203–210, 2018.

[17] E. A. D. Kurniawan, F. Fatmawati, and A. Dianpermatasari, "Model Matematika SEAR dengan Memperhatikan Faktor Migrasi Terinfeksi untuk Kasus COVID-19 di Indonesia," *Limits J. Math. Its Appl.*, vol. 18, no. 2, p. 142, Nov. 2021, doi: 10.12962/limits.v18i2.7774.

[18] D. Okuonghae and A. Omame, "Analysis of a mathematical model for COVID-19 population dynamics in Lagos, Nigeria," *Chaos, Solitons & Fractals*, vol. 139, p. 110032, Oct. 2020, doi: 10.1016/j.chaos.2020.110032.

[19] B. Ma, J. Qi, Y. Wu, P. Wang, D. Li, and S. Liu, "Parameter estimation of the COVID-19 transmission model using an improved quantum-behaved particle swarm optimization algorithm," *Digit. Signal Process. A Rev. J.*, vol. 127, 2022, doi: 10.1016/j.dsp.2022.103577.

[20] S. He, Y. Peng, and K. Sun, "SEIR modeling of the COVID-19 and its dynamics," *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1667–1680, 2020, doi: 10.1007/s11071-020-05743-y.

[21] Windarto, Eridani, and U. D. Purwati, "A comparison of continuous genetic algorithm and particle swarm optimization in parameter estimation of Gompertz growth model," *AIP Conf. Proc.*, vol. 2084, 2019, doi: 10.1063/1.5094281.

[22] T. A. Prasetyo, R. Saragih, and D. Handayani, "Genetic algorithm to optimization mobility-based dengue mathematical model," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 4, pp. 4535–4546, 2023, doi: 10.11591/ijece.v13i4.pp4535-4546.

[23] J. M. Carcione, J. E. Santos, C. Bagaini, and J. Ba, "A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model," *Front. Public Heal.*, vol. 8, no. May, 2020, doi: 10.3389/fpubh.2020.00230.

[24] M. Ghani, I. Qutsiati, U. Fadillah, W. Triyayuda, and M. Afifah, "A fractional SEIQR model on diphtheria diseasee," *Model. Earth Syst. Environ.*, vol. 9, no. 2, pp. 2199–2219, 2023, doi: 10.1007/s40808-022-01615-z.

[25] K. Das, B. S. N. Murthy, S. A. Samad, and M. H. A. Biswas, "Mathematical transmission analysis of SEIR tuberculosis disease model," *Sensors Int.*, vol. 2, no. July, p. 100120, 2021, doi: 10.1016/j.sintl.2021.100120.

[26] N. Nuraini, K. K. Sukandar, P. Hadisoemarto, H. Susanto, A. I. Hasan, and N. Sumarti, "Mathematical models for assessing vaccination scenarios in several provinces in Indonesia," *Infect. Dis. Model.*, vol. 6, pp. 1236–1258, 2021, doi: 10.1016/j.idm.2021.09.002.

[27] A. I. Abdel Karim, "The stability of the fourth order Runge-Kutta method for the solution of systems of differential equations," *Commun. ACM*, vol. 9, no. 2, pp. 113–116, 1966, doi: 10.1145/365170.365213.

[28] R. L. Haupt and S. E. Haupt, *Practical genetic algorithms*. John Wiley & Sons, 2004. doi: 10.1002/0471671746.

[29] Windarto, S. W. Indratno, N. Nuraini, and E. Soewono, "A comparison of binary and continuous genetic algorithm in parameter estimation of a logistic growth model," in *AIP conference proceedings*, 2014, pp. 139–142. doi: https://doi.org/10.1063/1.4866550.

[30] V. A. Navarro Valencia, Y. Díaz, J. M. Pascale, M. F. Boni, and J. E. Sanchez-Galan, "Using compartmental models and Particle Swarm Optimization to assess Dengue basic reproduction number R0 for the Republic of Panama in the 1999-2022 period," *Heliyon*, vol. 9, no. 4, p. e15424, 2023, doi: 10.1016/j.heliyon.2023.e15424.

[31] H. Gupta and O. P. Verma, "A novel hybrid Coyote–Particle Swarm Optimization Algorithm for three-dimensional constrained trajectory planning of Unmanned Aerial Vehicle," *Appl. Soft Comput.*, vol. 147, p. 110776, 2023, doi: 10.1016/j.asoc.2023.110776.

[32] M. Jiang *et al.*, "Analysis on the Development Trend of COVID-19 Outbreak in Beijing Based on the Cluster Analysis and SEIR Model," in *2020 Chinese Automation Congress (CAC)*, Shanghai, China, 2020. doi: 10.1109/CAC51589.2020.9327560.

[33] T. J. Roy, M. A. Mahmood, A. Mohanta, and D. Roy, "An Analytical Approach to Predict the COVID-19 Death Rate in Bangladesh Utilizing Multiple Regression and SEIR Model," in *2021 IEEE International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON)*, Dhaka, Bangladesh, 2021. doi: 10.1109/RAAICON54709.2021.9929470.

[34] W. P. T. M. Wickramaarachchi and S. S. N. Perera, "An SIER model to estimate optimal transmission rate and initial parameters of COVD-19 dynamic in Sri Lanka," *Alexandria Eng. J.*, vol. 60, no. 1, pp. 1557–1563, 2021, doi: 10.1016/j.aej.2020.11.010.

[35] J. Li *et al.*, "Do Stay at Home Orders and Cloth Face Coverings Control COVID-19 in New York City? Results from a SIER Model Based on Real-world Data," *Open Forum Infect. Dis.*, vol. 8, no. 2, 2021, doi: 10.1093/ofid/ofaa442.

[36] D. Efimov and R. Ushirobira, "On interval prediction of COVID-19 development in France based on a SEIR epidemic model," in *Proceedings of the IEEE Conference on Decision and Control*, 2020, pp. 3883–3888. doi: 10.1109/CDC42340.2020.9303953.

[37] S. Jiang, A. Al-Ataby, and F. Al-Naima, "COVID-19 Cases Estimation in the UK using Improved SEIR Models," in *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, IEEE, 2021, pp. 469–474. doi: 10.1109/DESE54285.2021.9719390.

[38] W. Zhao, Y. Sun, Y. Li, and W. Guan, "Prediction of COVID-19 Data Using Hybrid Modeling Approaches," *Front. Public Heal.*, vol. 10, no. July, pp. 1–13, 2022, doi: 10.3389/fpubh.2022.923978.

[39] W. Wu, "Computer intelligent prediction method of COVID- 19 based on improved SEIR model and machine learning," in *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*, IEEE, 2022, pp. 934–938. doi: 10.1109/ICPECA53709.2022.9719312.

[40] D. N. Vinod and S. R. S. Prabaharan, "COVID-19-The Role of Artificial Intelligence, Machine Learning, and Deep Learning: A Newfangled," *Arch. Comput. Methods Eng.*, vol. 30, no. 4, pp. 2667–2682, 2023, doi: 10.1007/s11831-023-09882-4.

[41] F. Saleem, A. S. A. M. Al-Ghamdi, M. O. Alassafi, and S. A. Alghamdi, "Machine Learning, Deep Learning, and Mathematical Models to Analyze Forecasting and Epidemiology of COVID-19: A Systematic Literature Review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 9, 2022, doi: 10.3390/ijerph19095099.

[42] D. Siahaan, I. K. Raharjana, and C. Fatichah, "User story extraction from natural language for requirements elicitation: Identify software-related information from online news," *Inf. Softw. Technol.*, vol. 158, p. 107195, Jun. 2023, doi: 10.1016/j.infsof.2023.107195.

[43] Y. Qian, X. Deng, Q. Ye, B. Ma, and H. Yuan, "On detecting business event from the headlines and leads of massive online news articles," *Inf. Process. Manag.*, vol. 56, no. 6, p. 102086, 2019, doi: 10.1016/j.ipm.2019.102086.

[44] S. Salsabila, S. M. P. Tyas, Y. Romadhona, and D. Purwitasari, "Aspect-based Sentiment and Correlation-based Emotion Detection on Tweets for Understanding Public Opinion of Covid-19," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 9, no. 1, pp. 84–94, Apr. 2023, doi: 10.20473/jisebi.9.1.84-94.

[45] F. Allahi, A. Fateh, R. Revetria, and R. Cianci, "The COVID-19 epidemic and evaluating the corresponding responses to crisis management in refugees: a system dynamic approach," *J. Humanit. Logist. Supply Chain Manag.*, vol. 11, no. 2, pp. 347–366, 2021, doi: 10.1108/JHLSCM-09-2020-0077.