

# Genetic Algorithm based Predictive model for COVID 19 – Theoretical Model based on an evolutionary approach

Srikanth Rangarajan (✉ [srangar@binghamton.edu](mailto:srangar@binghamton.edu))

Binghamton University

Srikanth Poranki

United Indian Health Services

Bahgat Sammakia

Binghamton University

---

## Research Article

**Keywords:** Covid-19, Genetic Algorithm, Evolutionary model, Artificial Intelligence, Second wave

**Posted Date:** May 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-107714/v3>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

***Genetic Algorithm based Predictive model for COVID 19 – Theoretical Model based on an evolutionary approach***

**Srikanth Rangarajan**

Assistant Research Professor,  
Department of Mechanical engineering  
State University of New York  
Binghamton NY 13905

**Srikanth Poranki**

Director, Performance improvement at United Health Services  
Binghamton, NY 13905

**Bahgat Sammakia**

Distinguished SUNY Professor  
Department of Mechanical engineering  
State University of New York  
Binghamton NY 13905

**Abstract**

In this manuscript we propose a novel *theoretical method* that models the evolution, spread and transmission of COVID 19 pandemic. The proposed model is inspired partly from the evolutionary based state of the art genetic algorithm. The rate of virus evolution, spread and transmission of the COVID 19 and its associated recovery and death rate are modeled using the principle inspired from evolutionary algorithm. ***Furthermore, the interaction within a community and interaction outside the community is modeled. The constraint with respect to interaction has been implemented by a machine learning type algorithm and becomes the unique part of our study.*** Using this model, the maximum healthcare threshold is fixed as a constraint. Our evolutionary based model distinguishes between individuals in the population depending on the severity of their symptoms/infection based on the fitness value of the individuals. There is a need to differentiate between virus infected diagnosed (Self isolated) and virus infected non-diagnosed (Highly interacting) sub populations/group. In this study the model results does not compare the number outcomes with any actual real time data based curves. However, the results from the model demonstrates that a strict lockdown, social-distancing measures in conjunction with more number of testing and contact tracing is required to flatten the ongoing COVID-19 pandemic curve. A reproductive number of 2.4 during the initial spread of virus is predicted from the model for the randomly considered population. The proposed model has the potential to be further fine-tuned and matched accurately against real time data.

Keywords : Covid-19, Genetic Algorithm, Evolutionary model, Artificial Intelligence, Second wave

Correspondence to [srangar@binghamton.edu]

## 1. Introduction

As the second wave of the COVID 19 pandemic is happening around the globe, the spread of COVID 19 has been the topic of discussion in today's life. As many researchers are globally working to understand the nature and the mechanism of spread, the healthcare systems are always in need for an accurate model to forecast the "COVID curve". A "COVID curve" is defined as the curve that represents the daily new cases with respect to time or the cumulative deaths with respect to time. Many predictive models based on curve fitting and stochastic are being produced with exact scales to forecast (extrapolate) the COVID 19 curve. These predictions have insights on how quickly the COVID 19 curve will grow and the extrapolated consequences. Forecast model play a vital role in understanding the healthcare demands/needs[1] , including how many intensive care unit beds, ventilator units and labors will be required to respond effectively. However, the authors believe that as the first and foremost objective of a predictive/forecast model is to estimate the relative effect of increased human interaction and increased social intervention measures [2], [3] on the COVID curve (cumulative death/daily new cases vs time curve). Simplified models with curve fitting approaches may provide less valid forecasts with real time accurately scaled data because they cannot extrapolate, account for the exact scale and extent of human interactions..

The SEIR (susceptible, exposed, infectious, removed) model is a commonly cited model. Lin et al. extended a SEIR model that considered various risks and the cumulative number of cases [4]. Stochastic transmission models have also been developed and considered [5]. The results from the model are stated with substantial uncertainty intervals. From the ongoing trends of the first, second wave and interpretation of data, it is very evident that epidemics will not follow similar paths in all places globally, even when geographically specific factors such as age distribution are considered. The rational way to go about distinguishing the geographies would be by population density, information on how busy the place is and information on age distribution. A detailed review of some of the existing models are listed by [6], [7]. The Standardized Infection Ratio (SIR) is the metric employed by the National Healthcare Safety Network (NHSN) to track health care associated infections (HAIs). The SIR is calculated by taking the ratio of the number of observed infections by the number of predicted infections. The proposed model in this manuscript future intends to help calculate the SIR with an objective of obtaining SIR close to 1. There are numerous modified SEIR (Susceptible - Exposed - Infectious - Recovered) models available in the CDC website [8]. List of different accepted predictive models are listed in the CDC website [8]. Christopher [9] proposed a theoretical model to understand the effect of social interventions on the spread of covid 19. The model proposed in the manuscript presently is not a data based model as it just tries to demonstrate the methodology involving evolutionary approach. There is no curve fitting approach incorporated in the method. The parameters varied in the population are similar to most of the models listed in [8]. The novelty in the method proposed in this manuscript is brought about in the way the interaction and mobility is modeled based on evolutionary principle. Furthermore, the complex mechanism of human interaction and sudden social intervention measures could be captured by employing evolutionary principle which the authors have attempted in this manuscript..

In this manuscript, the authors propose a new evolutionary based model for the COVID-19 pandemic with a predefined population that derives the methodology partly inspired from the state of the art Genetic Algorithm. Though, the genetic algorithm [10] is understood as an optimization methodology, parts of the methodology has been adopted to mimic the fitness of a population community that is assumed initially where in the crossover is replaced by interaction (spread and transmission) within the community. Genetic algorithms (GAs) are biological inspired optimization algorithms that derive it's methodology from the process of evolution. The GAs are widely used as a method to solve highly nonlinear optimization problems that demand global optimal solutions. GAs always involve parents and offspring as a part of the sub population. Any individual in a population is modelled to have a particular fitness. Only individual with high fitness values survive each generation (Survival of the fittest). This theme of the genetic algorithm has inspired the authors to model the spread of virus and survival of the individuals in a population community. De Jong[10] cautioned against understanding GAs as only optimization tools. The author stated that considering GAs as only an optimization tool is not the right approach. Furthermore, the author strongly suggested to perceive GAs as a tool to simulate natural process.

This motivates the work in this manuscript to adopt evolutionary approach to model spread of virus. Henceforth, we use a common notation, Evolution Programs (EP) for Genetic algorithm based models[11]. Genetic algorithms are a class of evolutionary programs[11]. The evolutionary model is a probabilistic algorithm which preserves a population of individuals for generations based on their fitness as it becomes the key to survival. Each individual is identified by a fitness value. Next, a new generation in the population is formed by preserving the fit individuals in the current generation. Randomly selected individuals from the initial population/sub population undergo modification by means of “genetic” operators (Operators that cause change in individual fitness) to form new solutions. A chosen sub population from a population of individuals undergoes some transformations which effects the fitness value of the individuals [11], and during this evolution process from one generation to another (in the units of time) the individuals strive for survival. The crossover and mutation has inspired the transmission, spread of virus and evolution of virus in a population community respectively. The step by step methodology is explained in Fig 1. Reiterating the fact that the idea of evolution programming is not new[12]–[14]. Many different versions of evolutionary systems exist. However, in this manuscript we discuss a novel evolutionary approach to model the spread of COVID 19 in a community. The most challenging part of the evolutionary based modelling of the pandemic spread is the implementation of the constraints. The violation of constraints can be associated with greater penalties. In the algorithm proposed in the current manuscript the penalty is posed naturally by constraining the interaction as more interaction leads to more spread of the infection. This is a great advantage of EPs as the constraint naturally grows or flattens the COVID curve.

Davis [15] believes that GAs are the most suitable algorithms to model many real world problems. Devis et al [16] stated that GAs are powerful tools that can be suitable to model the adaptation of the species to changing environment as the underlying principle of the GAs are the survival of the fittest. From the literature survey quoted, it is evident that understanding any evolutionary algorithm as just an optimization algorithm is not advised. This manuscript develops a model based on evolutionary principle to understand the spread of COVID-19 in a population. In this proposed algorithm to model the spread of COVID 19 we do not use parents and offspring, rather we have infected and non-infected people. The infected population could show symptoms and not show symptoms (Incubation). Furthermore, our model accounts for a distinction between infected diagnosed and infected non-diagnosed individuals, owing to the reason that the latter continue to interact in a population community and spread the infection more and the former are self-quarantined or isolated. Additionally, in this model, we consider the probability rate of the individuals in the recovered sub population becoming prone to the infection again.

## 2 Methodology

The flowchart of the methodology is shown in detail In Fig 1. The first foremost and the most sensitive input to the model is the initial population. The population can be distinguished in terms of the fitness value. Fitness is defined as the decimal equivalent of the binary string. For each individual the binary string length is fixed as 8 in this study. Once the population is generated, the process of interaction, spread and transmission inside and outside of population is simulated by modifying the binary string by choosing the most appropriate transmission site and interaction probability.

The maximum fitness for a binary string of length 8 is

$$11111111 = 255$$

The minimum fitness for a binary string of length 8 is

$$00000000 = 0$$

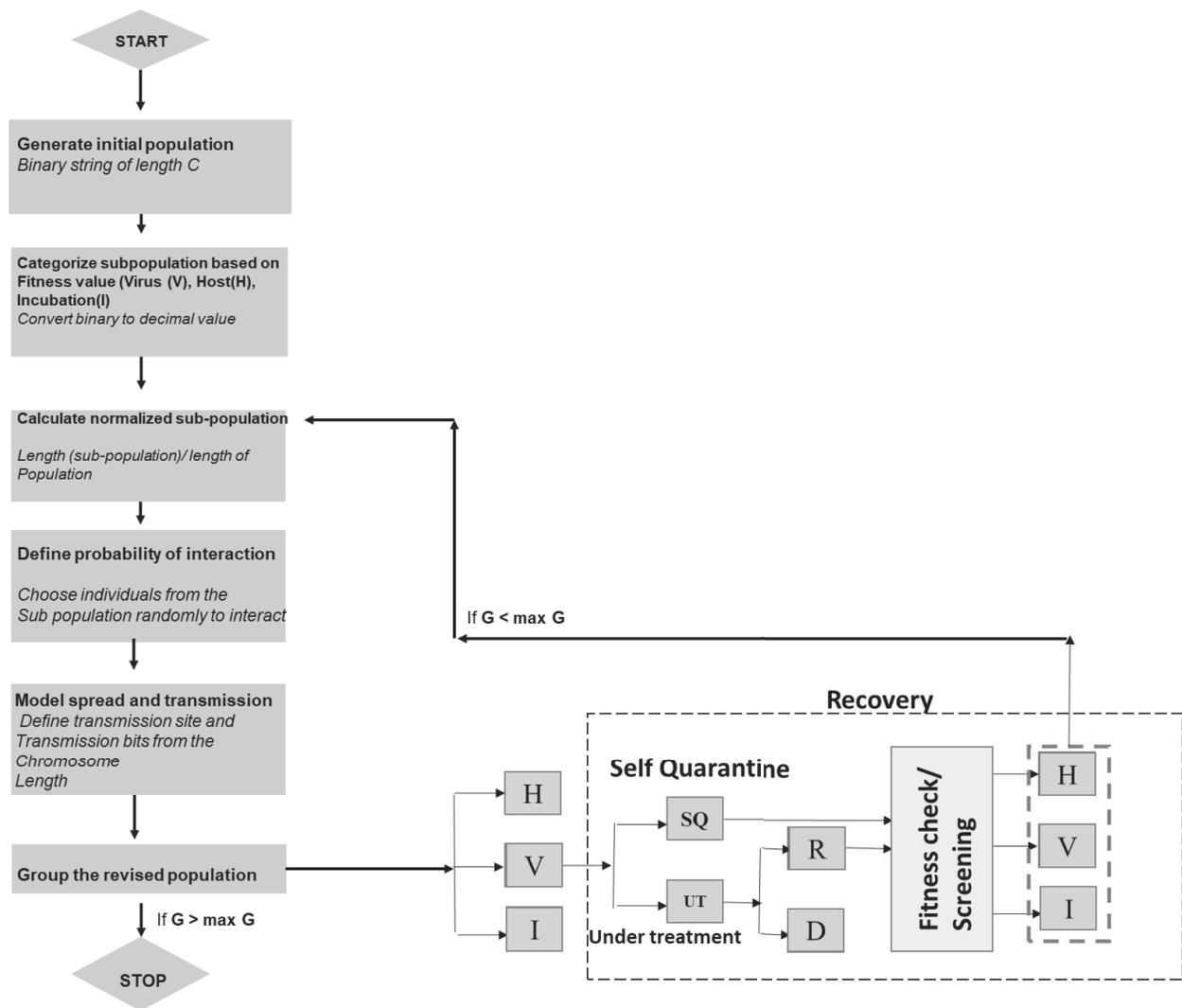


FIGURE 1: FLOWCHART OF THE METHODOLOGY

Any individual with a fitness value is 0.4 - 1 times the maximum fitness is qualified as a host population. Individual with fitness less than 0.1 times the maximum fitness is qualified as a part of the virus population. Any individual in between the fitness values of these two sub population become a part of incubation sub population. These bounds can be changed or accurately modeled when we have the exact information on the age wise population data. The age of the individual could be directly correlated to the fitness value of the individual.

## 2.1 Initial population

The initial population for the population community can be generated in two different ways

1. Random generation
2. Based on available data in a specific location

This manuscript tries to demonstrate the methodology in general and hence adopted random sampling of initial population. The population generated based on random samples are subdivided into groups based on the individual fitness value. The inputs required for generating the initial population are the population size and the length of chromosome. Each chromosome represents an individual [11] and the length of the chromosome implies the number of genes in each chromosome. The number of genes in the study is assumed to be 8. The chromosome consist of binary strings 0s and 1s generated randomly. The fitness value is the decimal equivalent of the chromosome binary string.

$$f(x) = y$$

Based on the fitness value of each chromosome, the population (P) is divided into three sub population

1. Host (H)
2. Incubation (I)
3. Virus (V)

The host population consist of people who are not infected. The virus population consists of people infected with virus. The incubation population is the class of sub population who are infected with virus but with less/no symptoms and potential transmitters of the virus.

Two more sub categories

4. "Recovered"
5. "Dead"/"expired"

These two categories are defined dynamically as the interaction begins.

In this study if the fitness value of an individual is greater than 30% of the maximum fitness, the individual becomes a part of the host sub population. If the fitness value is less than 30% of the lower bound of the host individual (least fitness among the host sub population) then the chromosome becomes a part of the virus sub population. *The lower bound of the fitness in the virus population is termed as "super spreader"* [17]. The chromosome whose fitness is intermediate between host and virus becomes a part of incubation sub population.

By definition the incubator sub population are the individuals who potentially infected by COVID 19 but have no symptoms, but they are still potential transmitters/carriers of the virus.

The population generated based on available data can be subdivided not groups based on age. The population above the age 65 can be regarded as the population with weak immunity. The fitness of the individuals are a direct function of the age. After the initial population is generated, the whole of the population is subject to a random mutation. Mutation is defined as the process of changing 0s to 1s and vice versa. The inputs that are required to perform the mutation are the probability of mutation. The total number of bits/genes that can undergo mutation can be determined from the mutation probability. Once the mutation is completed, the sub populations are formed. *The individuals in the subpopulation namely host, virus and incubation are identified by the fitness value. Furthermore, the dynamic subpopulation recovered and dead are also identified by the fitness value. There is no recommendation on what fitness value defines the limit of each of the sub population. However, as a part of proposing this theoretical model a reasonable*

*assumption on the model parameters are made The individuals part of the virus sub population with very less fitness values constitute the super spreader[17]*

## **2.2 Interaction within a population community**

The interaction of individuals within a population community is modeled through the mechanism of transmission. A pictorial representation of the transmission mechanism is shown in figure 2. This transmission is also discussed in [17] The direction of interaction between various sub populations is as shown in figure 3.

The individuals from the sub population that interact are chosen randomly. This mimics the real scenario where in every individual is equally susceptible. As seen from figure 3 the host population is always a receiver. The virus population is always a transmitter. However, the incubation population is both a transmitter and receiver. As the interaction begins, two new sub populations emerge as stated earlier. They are the recovered sub group and the dead/expired sub group. The research published recently makes it evident that the recovered population is still susceptible and potential transmitters. This algorithm models the recovered sub group in two different ways with a set of assume

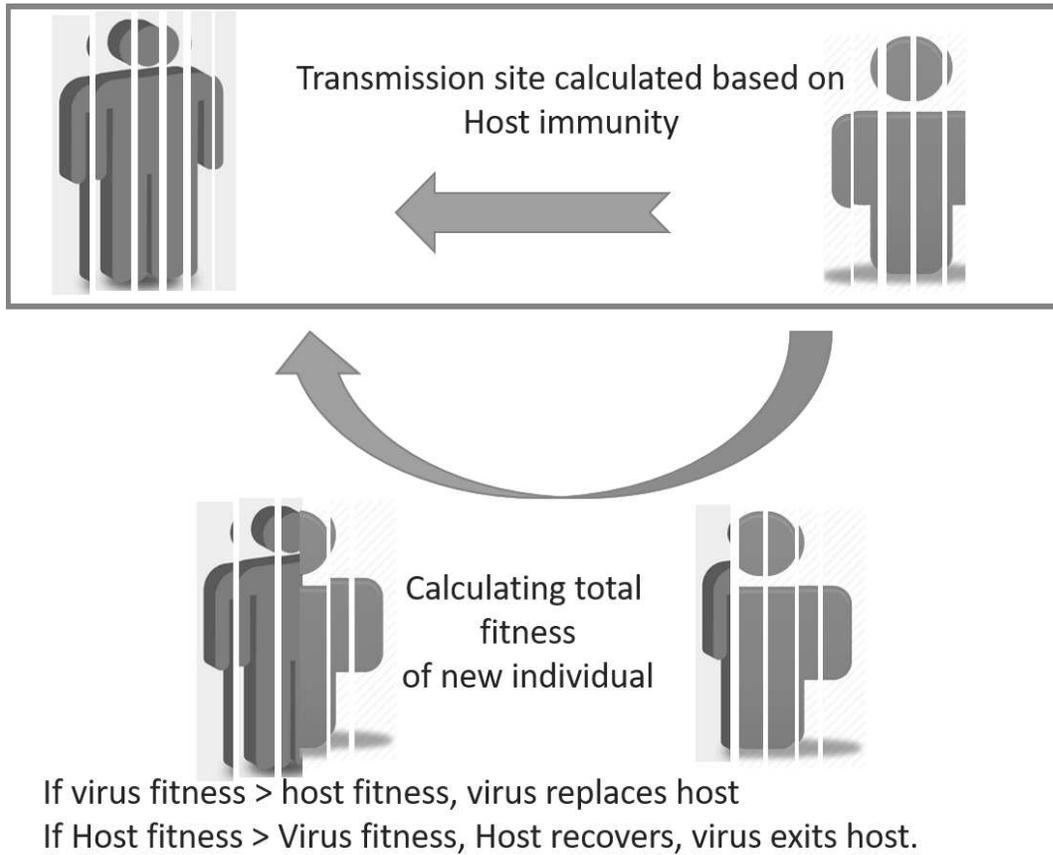
1. The population recovers, develops anti bodies and become a part of host
2. The population recovers, develops anti bodies and become a part of the incubation
3. None of the individuals become totally immune to the virus.

During the initial period the curve of the virus infected persons are of prime interest. It is evident that the curves from the financial times website that for some of the countries are flattening early and some flatten later which can be purely attributed to the level of social intervention.

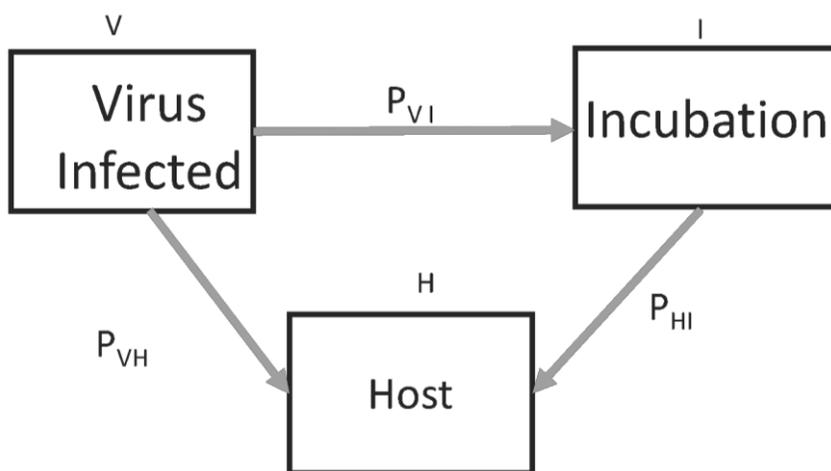
During the process of transmission, the transmission site is defined randomly. The transmission site is not defined/ controlled by any function as the level of interaction is not controlled in reality. Therefore the inputs required to model the transmission are the probability of transmission and the transmission site.

The total probability of interaction = 1

Total probability = Probability of host interaction with virus + Probability of virus interaction with host + Probability of incubation interaction with host + Probability of incubation interaction with virus.



**FIGURE 2: PICTORIAL REPRESENTATION OF VIRUS TRANSMISSION BETWEEN HOST (BLUE) AND THE VIRUS (RED)**



**FIGURE 3: DIRECTION OF INTERACTION (TRANSMITTER AND RECEIVER) ADOPTED IN THIS ALGORITHM**

To make the algorithm be relevant to any population community the sub population size is normalised.

$$h = H/P$$

$$v = V/P$$

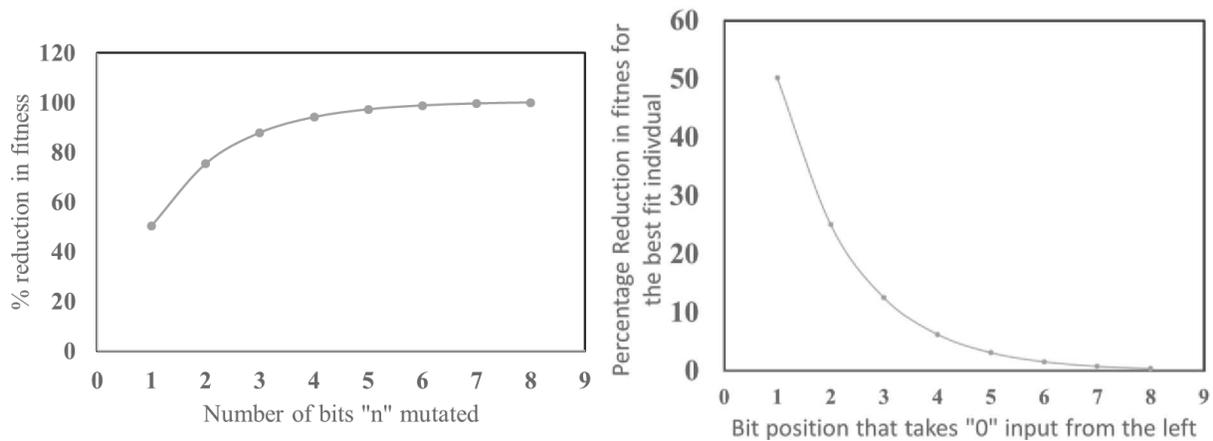
$$I = I/P$$

For example let us consider,

Individual 1 = a1 a2 a3 a4 a5 a6 a7 a8

Individual 2 = b1 b2 b3 b4 b5 b6 b7 b8

The bit positions that contribute most to the fitness of the individual are to the extreme left (a1, a2, a3 and b1, b2, b3). Henceforth the individual with maximum fitness has value 1 occupied in the bit location a1 to a4 and b1 to b4. For example if we assume that individual 2 is low in fitness, while individual 1 is high in fitness and since the transmission is one way as indicated in Fig 3, the transmission site and the number of transmission bits play a major role in determining the fitness of the individual 1. If the number of bits that can be changed is 4, then total possible fitness for individual 1 becomes a function of the exact position of the bit that can undergoes change. Figure 4 shows the effect of bit position and the number of bits that are mutated to the reduction in fitness value.



**FIGURE 4: (a) Reduction in fitness value with the number of bits that is changed/mutated (b)Reduction in fitness value with bit position that is mutated**

Henceforth, the level of interaction can be constrained by having a low value of number of bits “n”.

### 2.3 Interaction from outside population

The interaction of the sub population with the outside population community is modeled through one way mutation. The one way mutation is the process of changing 1’s to 0’s. This implies that the interaction with the outside population does deteriorate one’s fitness or does no change, but it does not improve the fitness value of the individual. This mutation probability can be varied based on specific population community. For example when modeling the situation at busy cities with world’s busiest airports, the mutation probability can be set to a high value. For towns and non-busy cities, this mutation probability is kept low. The mobility data published in Unacast [18] website clearly indicates that the reduction in mobility within and outside of the population community has a great influence on the COVID 19 curve. It is stated that World

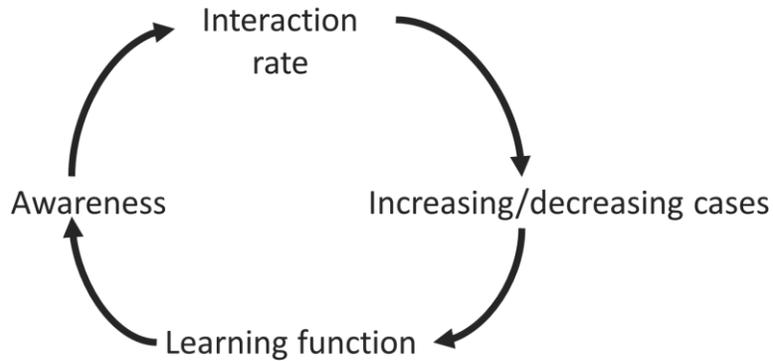
Health Organization and the CDC declares that social distancing is currently the most effective way to steadily flatten the curve of COVID-19 spread [8]. Furthermore, Migration patterns of people from each city and state have been captured utilizing the available GPS data. The data indicates that the reduction in migration rate has contributed to the flattening of the curve in each population community.

#### **2.4 Modeling recovery**

The recovered population is a very dynamic population that gets formed/grouped in the course of interaction. The recovered population is a sub group formed from the virus sub population. In reality not all of the virus infected population (infected with virus and show symptoms) get hospitalized. Among the percentage of the hospitalized population, some recover and some die. The population that gets hospitalized is chosen based on the information of initial fitness of the individuals. The individual of the hospitalized population either recovers or expires. The recovered population could still transmit and be susceptible. The inputs required to model recovery are probability of recovery. The recovery is modeled via one way mutation by changing 0's to 1's. This means that recovery is a process of improving an individual fitness value. Probability of recovery is kept to a small value as the recovery is gradual and slow. The flow chart of the modeling approach is as shown in Figure 4.

#### **2.5 Constraints**

The constraints on interaction is brought by the awareness social distancing, self-isolation, hand washing. The interaction rate is controlled by changing the rate of interaction operator. The healthcare threshold is also a potential constraints in knowing the number of hospital beds available in a population community. The awareness of the ongoing situation serves as an important constraint to any population community to flatten the curve. The objective function for the algorithm from the virus perspective is to maximize the fitness of the virus sub population. The maximum value for the normalized fitness is equal to 1. The problem is solved both as a constrained as well as on an unconstrained approach. The algorithm proposed in the manuscript can be used to analyze and forecast the curve only by knowing an approximate value of the parameters. Like the modified SIER models proposed in the CDC website [8], the model proposed in the manuscript also do not make specific assumptions about which interventions have been implemented or will remain in place. The constraint with respect to interaction and transmission is implemented by a probability function. The probability of interaction reduces with the increased awareness about the disease among the population. The model self learns from the increasing cases and awareness and damps the interaction probability. This self learning artificial intelligent based learning becomes the uniqueness of the proposed approach.



## 2.6 Reproduction number $R_0$

The reproduction number is defined as the number of non-infected people possibly infected by one infected person. This number is crucial to understand the intensity of spread in a population community. In this model the value of  $R_0$  is calculated as a function of time.

## 3. Results and discussion

The following results and discussion pertains to following parameters in the algorithm. The model results does not compare the number outcomes with any actual real time data based curves

- Population size = 190000
- Population density = Varied with models/simulations
- Type of location = Busy/Moderately busy/ Village
- Interaction/ Mobility level[18] = Varied with simulation
- Healthcare threshold = Assumed to be 30-40% of the total population
- $R_0$  = the reproduction number is calculated based on the numbers (The number of people can get infected from one infected person).

The normalised population is defined as the ratio of the number of respective sub population to the total number on the population community. The maximum value of the normalised population is 1 and the minimum value is 0.

### 3.1 Effect of social intervention

Figure 5 shows the normalised population with and without social intervention.

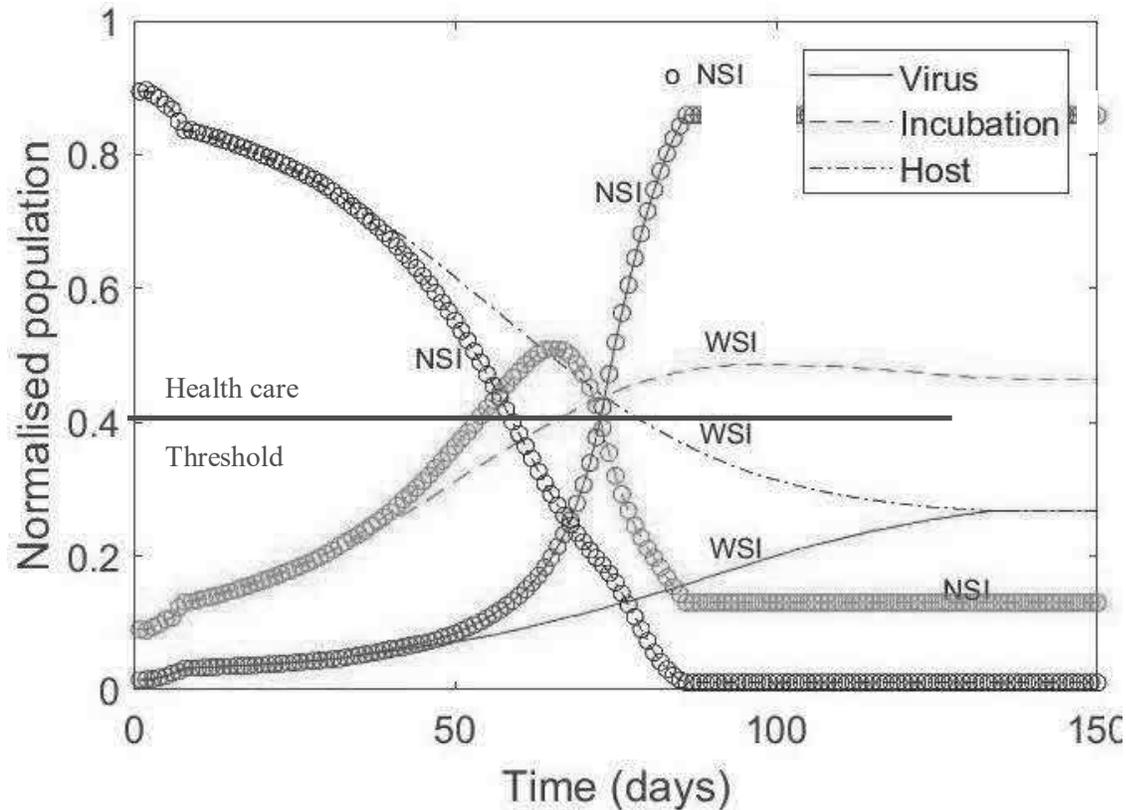


FIGURE 5: FIGURE SHOWING THE EFFECT OF SOCIAL INTERVENTION ON THE NORMALIZED SUB POPULATION

The rate of increase in the infected numbers and the associated decrease/increase in other sub population is shown in Fig 6. NSI indicates “No Social Intervention”, WSI indicates “With Social Intervention”. This clearly indicates that the social intervention measures are seen to have a substantial impact on the growth of the virus infected population number. It is also evident that the social intervention measures can flatten the curve of the host population and the incubation population. The intensity of the social intervention measures can be increased by introducing much higher decay rate for interaction operator. The plot assumes that the social intervention measures are introduced around the day 50.

### 3.2 Effect of hospitalization and recovery without interventions

Figure 6 shows the normalised population with and without recovery (No intervention)

Fig 6 compares four interesting cases with respect to the intervention measures and community wide testing that can lead to hospitalization.

- NSI - No Social Intervention
- WR - With Recovery
- WSI - With Social Intervention
- NR - No Recovery

The four curves shown in the figure corresponds to the virus sub population. From the results shown in the model, it is evident that combined social distancing and community wide testing is essential for the early curve flattening scenario. Reasonable amount of testing without social intervention measures

doesn't lead to the early curve flattening scenario. However, with social intervention measures and moderate testing/hospitalization can still be very effective.

### 3.4 Effect of interaction

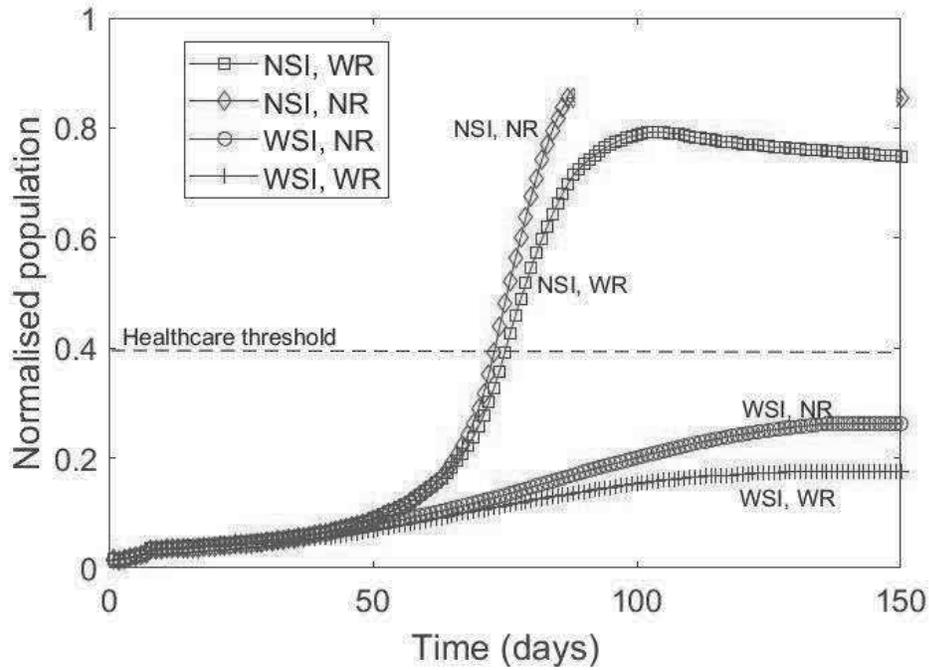
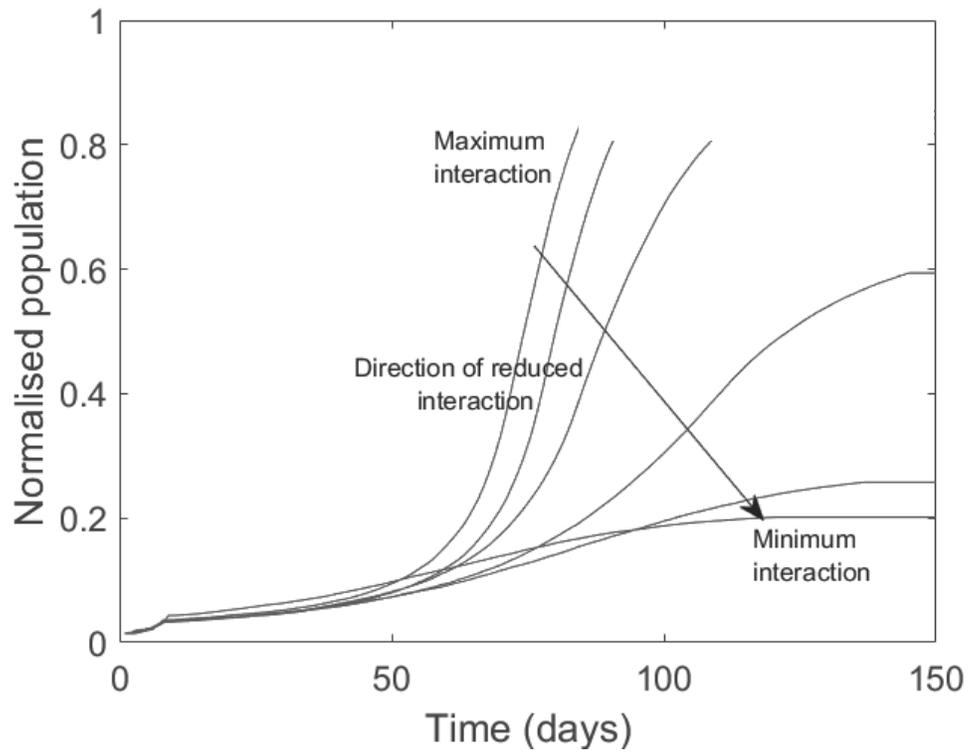


FIGURE 6: FIGURE SHOWING THE EFFECT OF HOSPITALIZATION AND SOCIAL INTERVENTION ON THE NORMALIZED VIRUS POPULATION



**FIGURE 7: FIGURE SHOWING THE EFFECT OF INTERACTION ON THE NORMALIZED VIRUS POPULATION**

The Effect of interaction within a population community on the growth of virus infected number is demonstrated in Fig 7. The output from the model suggest that minimum the interaction, earlier the curve flattens. The results from Shamam model [19] based on metapopulation SIER model indicates the same trend. The authors concluded that individual living in highly populated neighborhood was infected more.

### 3.5 Effect of interaction with PPE

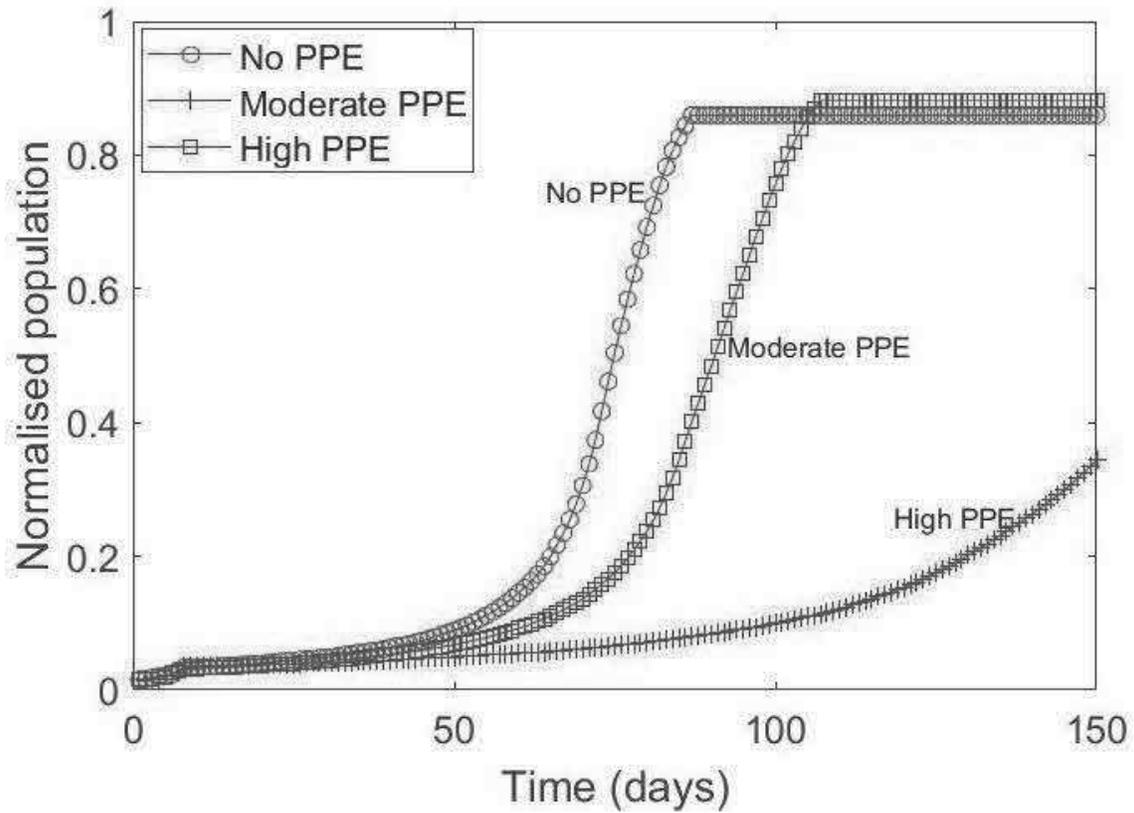


Figure 8: FIGURE SHOWING THE EFFECT OF INTERACTION WITH PPE on the normalized virus population

The effect of interaction with personal protective equipment's (PPE) such as face mask and hand gloves is shown in Figure 8. The result indicate that the use of PPE during interaction has considerable effect on the rate of growth of virus sub population and results in delayed rise in the curve as seen for the case of high PPE. High PPE pertains to a case where more number of people use PPE while interaction. The probability of transmission as a function of distance between two individuals is discussed in [17].

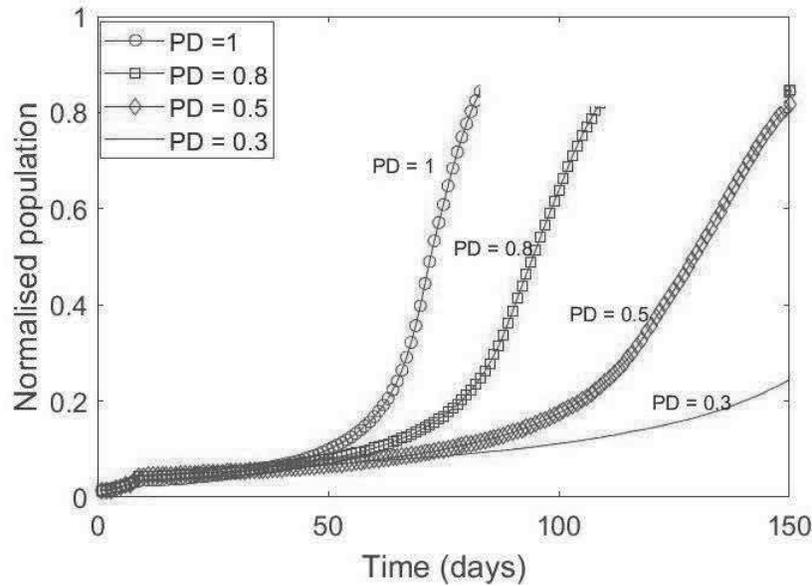


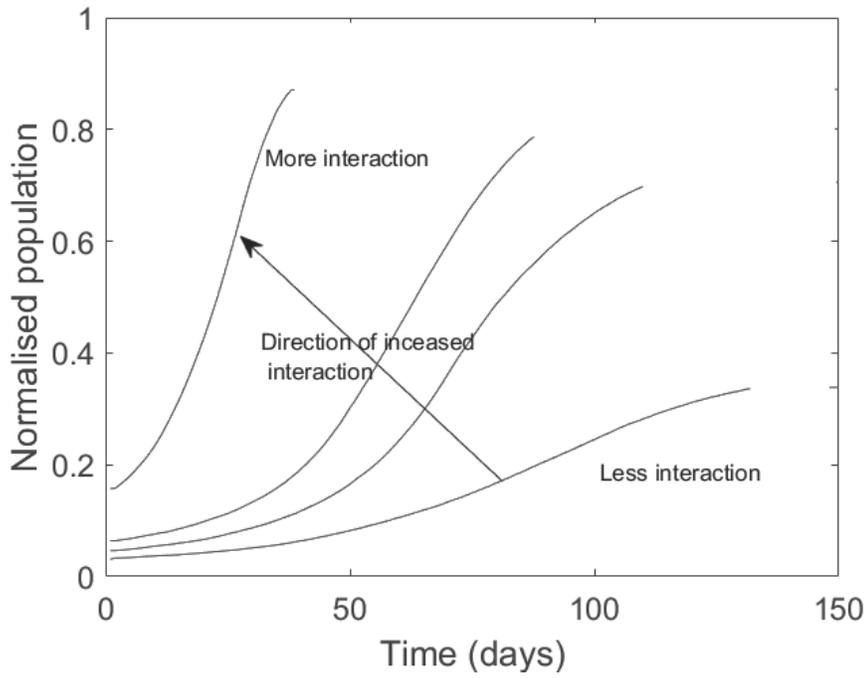
Figure 9: FIGURE SHOWING THE EFFECT OF POPULATION DENSITY on the normalized virus population

### 3.6 Effect of population density

Figure 9 shows the number of cases trend for a high population density place and a low population density place. The results from the model makes it evident that the population density has a great impact on the rate of increase of the virus sub population. The population density (PD) of 1 indicates high population density and PD of 0.3 indicates low population density. The trend of the daily infected cases from the model is seen to be more correlated with the population density of a particular geographical location.

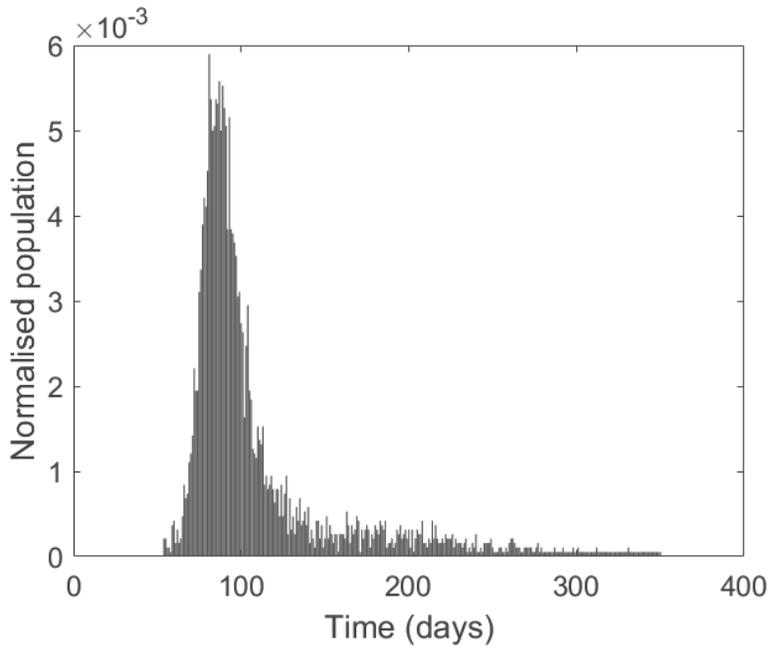
### 3.7 Effect of interaction from outside community

Figure 10 shows the effect of mutation probability on the growth of number of cases. The effect of mobility on the curve flattening is published by Unacast Though social distancing and hand washing within a population community could slow down the rate of spread of the virus, the interaction with population community outside is seen to have a major impact on the early rise of the curve, indirectly contributing to the rise of the curve at later stages. The results from the model makes it evident that more interaction of the population with outside community that is possibly infected with the virus can have adverse effect on the trend of the curve. Hence the borders should remain closed. The results from [18] also indicates that mobility has a direct impact on the COVID curve.

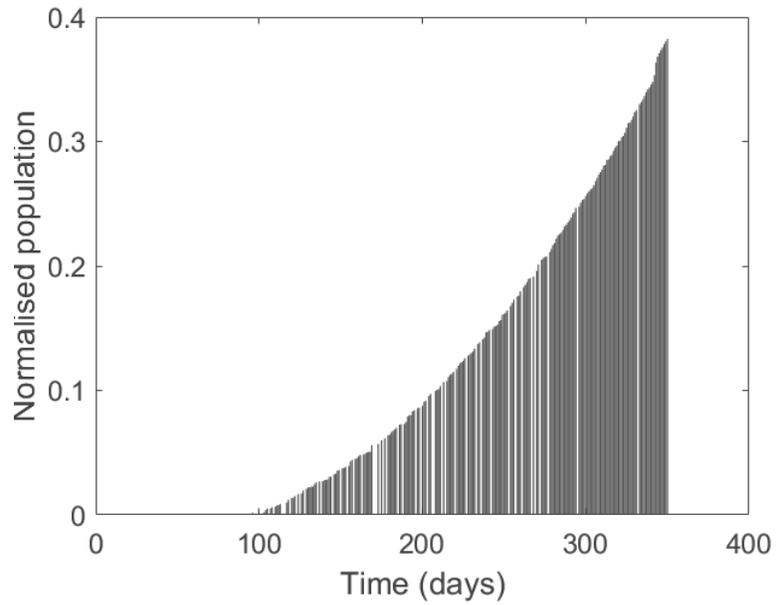


**FIGURE 10: FIGURE SHOWING THE EFFECT OF INTERACTION OUTSIDE THE POPULATION COMMUNITY**

**3.8 Recovery plot, death plot**



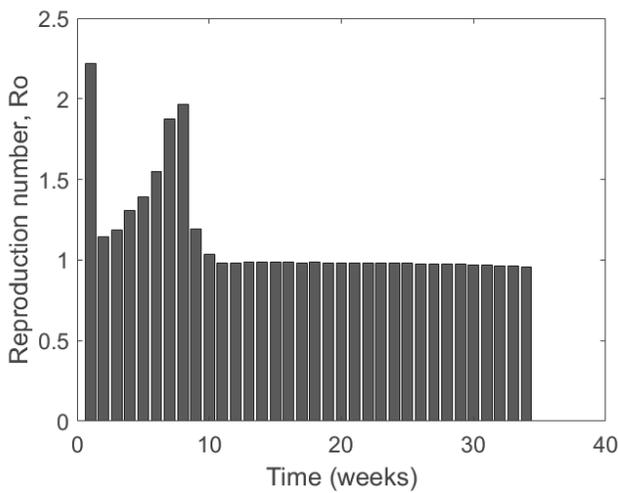
**FIGURE 11: FIGURE SHOWING THE TREND OF EXPIRED POPULATION DURING THE SPREAD OF COVID 19**



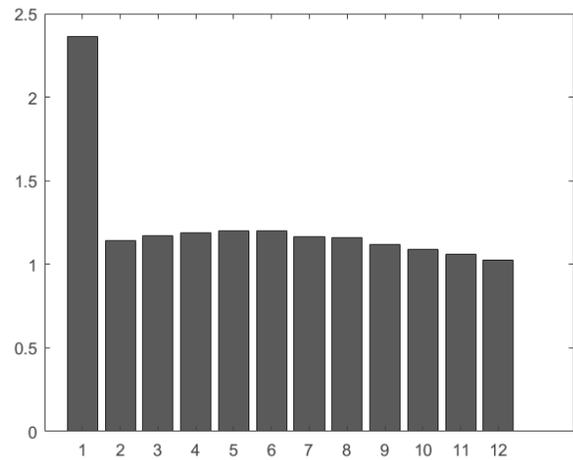
**FIGURE 12: FIGURE SHOWING THE MODEL RESULTS OF THE RISE OF RECOVERED POPULATION WITH TIME**

Figure 11 and Figure 12 shows the trend of increasing dead population and recovered population from the model. The results from the model makes it evident that the death rate spikes after a certain period of spread of COVID 19. The exact time and the spike number can be brought about by much more detailed modeling using the real time data. The results from model shown in Figure 12 makes it evident that the recovered sub population becomes a major part of the sub population as time progresses. This results makes it evident that the dynamic sub population namely recovered and expired trends can be matched with real time data using the data from local topography and real time intervention measures.

### 3.9 Reproduction number



**FIGURE 13: FIGURE SHOWING THE VARIATION OF REPRODUCTION NUMBER  $R_0$  WITH TIME WITHOUT SOCIAL INTERVENTION MEASURES**



**FIGURE 14: FIGURE SHOWING THE VARIATION OF REPRODUCTION NUMBER  $R_0$  WITH TIME DURING STRICT SOCIAL INTERVENTION MEASURES**

The basic reproduction number ( $R_0$ )[9] is a statistical parameter that helps predict the longevity of the virus spread in a population. The physical interpretation of  $R_0$  is the average number of secondary cases that emerge from a single primary infected diagnosed/non-diagnosed case in a totally susceptible population. The results from the model shows the variation of the reproduction number  $R_0$  with weeks. As seen from Figure 13, the reproduction number  $R_0$  spikes during the initial time of the virus spread and seen to oscillate for the case of no social intervention measures. For the case of high social intervention measures the value of  $R_0$  is seen to decrease gradually and reach a plateau with time as shown in Fig 14. The final value of  $R_0$  becomes slightly less than 1 (0.96) which is a good indication of slowdown of the virus spread. The results from the model also makes it evident that the mortality rate associated with the sudden spike of  $R_0$  is very high. The model predicts a  $R_0$  of 2.3 initially and reduces gradually with social intervention. Li et al. [20] reported a very close  $R_0$  value of 2.4. The value of the  $R_0$  obtained in this study is consistent with theoretical finding by [9]

#### 4. Effect of Binary string length

the model was run with different string lengths with non-dimensional parameters and found variation between cases in the acceptable range as shown in Table 1. However, for the ease of computation complexity binary string length of "8" was demonstrated in the article.

Table 1: Effect of binary string length on the predictive analysis

| S No | Binary string length | Social intervention condition                       | Time to reach 50% infected population | Ro   | Computational time (s) |
|------|----------------------|---|---------------------------------------|------|------------------------|
| 1    | 8                    | With social intervention, Moderate PPE, PD = 0.8    | 85                                    | 0.98 | 23                     |
| 2    | 16                   | With social intervention, Moderate PPE, PD = 0.8    | 81                                    | 0.97 | 62                     |
| 3    | 32                   | With social intervention, Moderate PPE, PD = 0.8    | 80                                    | 0.97 | 120                    |
| 4    | 64                   | With social intervention, Moderate PPE, PD = 0.8    | 80                                    | 0.97 | 320                    |
| 5    | 8                    | Without social intervention, Moderate PPE, PD = 0.8 | 62                                    | 2.4  | 11                     |
| 6    | 16                   | Without social intervention, Moderate PPE, PD = 0.8 | 55                                    | 2.32 | 25                     |
| 7    | 32                   | Without social intervention, Moderate PPE, PD = 0.8 | 52                                    | 2.33 | 45                     |
| 8    | 64                   | Without social intervention, Moderate PPE, PD = 0.8 | 52                                    | 2.31 | 75                     |

Furthermore, the authors would like to state that the parametric study on the interactive nature of the string length with different population metrics could emerge as a whole new research study in future. From Table 1 it is evident that the effect of string length was more prominent on the cases with no social intervention cases rather with social intervention cases. However, the computational time and complexity involved with increasing string length is also evident from the results

## 5. Validation

Our present theoretical model is validated for the accuracy of the prediction of Reproduction number  $R_0$ .  $R_0$  validation on reduction in  $R_0$  with literatures available. This validates our method

| Work                          | $R_0$ value   | Method description       |
|-------------------------------|---|--------------------------|
| Q Li et al [20]               | 2.4   | Statistical (Data based) |
| Christopher Bronk Ramsey [9]  | 2.4 (before intervention),<br><1 (after intervention)   | Theoretical              |
| Srikanth et al (Present work) | 2.4 (before intervention),<br>0.98 (after intervention) | Theoretical              |

The another validation was in terms of population group

| Work                  | Super spreader %                                    | % infected by super spreader | Method description       |
|-----------------------|---|------------------------------|--------------------------|
| Amit and Rajneesh[17] | 5   | 80                           | Validated with real time |
| Srikanth et al        | 8 (generated automatically using fitness threshold) | 72                           | Theoretical              |

## 5. Conclusions

This study demonstrates a novel methodology in the form of evolutionary algorithm to simulate the spread and transmission of COVID 19 pandemic. The evolutionary based model is able to produce the trend of the curve, however accurate prediction is not possible until the real time data is processed accordingly. The level and intensity of social interaction are incorporated into our interaction probability over the whole population, irrespective of age. This study concludes that the model results indicate testing along with social distancing is seen to flatten the curve much earlier. The results from the model demonstrates that a strict lockdown, social-distancing measures in conjunction with more number of testing and contact tracing is required to flatten the ongoing COVID-19 pandemic curve. Furthermore, closing borders would help flatten the curve much more effectively. The theoretical model developed in this manuscript could be possibly employed to analyze the aspects of the dynamics of pandemic progression under various scenarios.

### Future plans

The authors would like to state that, the model presented is fully flexible and suited to include, age distinction among the individuals in the community. The authors would also wish to express their interest

for collaboration from interested researchers in healthcare field to develop this algorithm further to match the real-time data.

### Conflict of interest declaration

On behalf of all authors, the corresponding author states that there is no conflict of interest.

### References

- [1] “Lawler J. What healthcare providers need to know: preparing for the COVID-19. American Hospital Association webinar.”
- [2] D. Adam, “Special report: The simulations driving the world’s response to COVID-19,” *Nature*, vol. 580, no. 7803, pp. 316–318, 2020, doi: 10.1038/d41586-020-01003-6.
- [3] C. J. M. Whitty, “What makes an academic paper useful for health policy?,” *BMC Med.*, vol. 13, no. 1, p. 301, Dec. 2015, doi: 10.1186/s12916-015-0544-8.
- [4] Q. Lin *et al.*, “A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action,” *Int. J. Infect. Dis.*, vol. 93, pp. 211–216, Apr. 2020, doi: 10.1016/j.ijid.2020.02.058.
- [5] A. J. Kucharski, T. W. Russell, C. Diamond, S. Funk, and R. M. Eggo, “Early dynamics of transmission and control of 2019-nCoV: a mathematical modelling study,” p. 17.
- [6] N. P. Jewell, J. A. Lewnard, and B. L. Jewell, “Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections,” *JAMA*, vol. 323, no. 19, p. 1893, May 2020, doi: 10.1001/jama.2020.6585.
- [7] G. Giordano *et al.*, “Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy,” *Nat. Med.*, vol. 26, no. 6, Art. no. 6, Jun. 2020, doi: 10.1038/s41591-020-0883-7.
- [8] CDC, “Coronavirus Disease 2019 (COVID-19),” *Centers for Disease Control and Prevention*, Feb. 11, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html> (accessed Oct. 09, 2020).
- [9] C. Bronk Ramsey, “Human agency and infection rates: Implications for social distancing during epidemics,” *Plos One*, vol. 15, no. 12, p. e0243699, 2020.
- [10] “De Jong, K.A., Genetic Algorithms: A 10 Year Perspective, in [167], pp.169—177.”
- [11] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd ed. Berlin Heidelberg: Springer-Verlag, 1996.
- [12] “Fogel, L.J., Owens, A.J., and Walsh, M.J., Artificial Intelligence Through Simulated Evolution, John Wiley, Chichester, UK, 1966.”
- [13] “Holland, J.H., Adaptation in Nature and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.”
- [14] “Schwefel, H.-P., Numerical Optimization for Computer Models, John Wiley, Chichester, UK, 1981.”

- [15] “Davis, L., Adapting Operator Probabilities in Genetic Algorithms, in [344], pp.61—69.”
- [16] “Devis, L. and Steenstrup, M., Genetic Algorithms and Simulated Annealing: An Overview, in 73], pp.1—11.”
- [17] A. Agrawal and R. Bhardwaj, “Probability of COVID-19 infection by cough of a normal person and a super-spreader,” *Phys. Fluids*, vol. 33, no. 3, p. 031704, 2021.
- [18] “Covid-19 Social Distancing Scoreboard — Unacast.” <https://www.unacast.com/covid19/social-distancing-scoreboard> (accessed Oct. 09, 2020).
- [19] “Emeruwa, U. N., S. Ona, J. Shaman, A. Turitz, J. D. Wright, C. Gyamfi-Bannerman, A. Melamed. Associations between built environment, neighborhood socioeconomic status and SARS-CoV-2 infection among pregnant women in New York City. *JAMA*.”
- [20] Q. Li *et al.*, “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia,” *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020, doi: 10.1056/NEJMoa2001316.

## Nomenclature

|     |                             |
|-----|-----------------------------|
| G   | - Generation/Iteration/Time |
| H   | - Host sub population       |
| I   | - Incubation sub population |
| NSI | - No Social Intervention    |
| NR  | - No recovery               |
| P   | - Population                |
| PD  | - Population density        |
| V   | - Virus sub population      |
| WSI | - With social intervention  |
| WR  | - With recovery             |

## Declarations

### Funding

Not applicable

### Conflicts of interest/Competing interests (include appropriate disclosures)

Attached

### Availability of data and material

Not applicable

**Code availability**

Not applicable