

Epidemic Model Simulation for Corona Pandemic in Tamil Nadu

**Dr.V.Latha Jothi¹, Dr.S.Rajalakshmi², P.Ananthi³, C.Santhini⁴, A.H.Sheerin⁵
and Dr.S.Jabeen Begum⁶**

¹Velalar College of Engineering and Technology, Erode, Tamil Nadu, India,
lathajothi.s@gmail.com

² Malla Reddy Engineering College for Women, Hyderabad, Telangana, India,
mrjislml@gmail.com

³Velalar College of Engineering and Technology, Erode, Tamil Nadu,
India, ananthi2698p@gmail.com

⁴ Velalar College of Engineering and Technology, Erode, Tamil Nadu, India,
csanthini96@gmail.com

⁵Coimbatore Medical College, Coimbatore, Tamil Nadu, India, sheerin30499@gmail.com

⁶Velalar College of Engineering and Technology, Erode, Tamil Nadu, India,
sjabeenbegum@gmail.com

ABSTRACT

Forecasting and spreading of Covid-19 in Tamil Nadu state is demonstrated by simulating an epidemiology model Susceptible-Infectious-Recovered-Demised (SIRD) model. Implicit analytical solution is applied for some parts of the model and for other parts finite difference methods are used. On the basis of SIRD model, values of coefficient of infection, coefficient of morality and coefficient of recovery can be found. In addition to the calibration of the model the ratio of the average rate of death to the average rate of recovery for the pandemic in Tamil Nadu is calculated. For this model, data is collected from Tamil Nadu Health and Family Welfare Department. The data has number of infected cases, death cases, recovered cases and hospitalized cases per day in Tamil Nadu State. The prediction results give good analysis and better understanding of the spread of the disease in the state. Based on the results, it is obvious that as the number of days increases that cumulative count of infected patients are also increased. But in the pandemic a fall would be expected after certain period of time. However this prediction model enables us to make quick response of the pandemic and get more insight about the data.

Keywords

Forecasting, Epidemic model, SIRD model, Analysis, Pandemic, Prediction

Introduction

In early November 2019, an outbreak of “Pneumonia of unknown etiology” was found in Wuhan, Hubei province, China and it was ravaging China and this disease becomes a threatening to reach a pandemic state. A new strain of beta corona virus is the causative agent and it is closely related to Severe Acute Respiratory Disease (SARS) and Middle East Respiratory Disease (MERS). The origin of the disease was determined based on the preliminary investigations are suggestive of bats. Similar to SARS and MERS, the novel virus also transmitted from one person to another person principally by respiratory droplets. According to Centers for Disease and Prevention (CDC), the symptoms of the disease may include fever, cough, shortness of breath and fatigue. Preliminary data suggest that older people and people with co-morbidities are likely to be

infected severely while most of the infected patients experience mild symptoms some may develop to various organ failure and pneumonia which might result in death. In January World Health Organization named the virus “Covid-19” which refers to the meaning of “Corona Virus Disease 2019”. Covid-19 is a Public Health Emergency of International Concern and it was declared by the WHO Director-General. The virus has been continuous to spread in rapid rate all over the world. The rapid rise in the number of Covid-19 incidents worldwide has prompted the need for immediate counter measures to curb the catastrophic effects of the Covid-19 outbreak.

In India, the Covid-19 was reported on 30 January 2020. India has the largest number of confirmed cases in Asia and the second largest number of confirmed cases in the world. Tamil Nadu is one of the affected states of India. In March 7 2020, Tamil Nadu Government faced its first corona case. All the 38 districts are affected during the pandemic whereas Chennai the capital of state was affected the worst. Health Department reports that 88% of patients are asymptomatic while 84% of deaths were among those with co-morbidities. Tamil Nadu Government responded to the pandemic by following contact tracing, testing and surveillance model. Government has also imposed strict lockdown and restricted inter-state movement to control the pandemic.

As a contribution to the ongoing public health crisis, an epidemic model is built with the help of available data to get a small distance to bring the conclusion to the pandemic. The models and algorithm used here for simulating the forecasting and spreading of the virus are also used to find the other related coefficients of the pandemic such as reproduction number for a certain region.

SIRD Model Simulation

Methodology

This analysis is based on the concept of finding the relationship between the transportation that makes a vital role in spread of disease and the growth of patients count in a geographic location. In this model, certain population is considered for a particular geographic location i.e. Tamil Nadu state in this case. For this purpose Susceptible- Infectious-Recovered-Demised model is proposed. This compartmental model is a mathematical model, where the disease can be moved people between categories such as healthy, infected or sick, recovered or developed immunity and fatal. This is a truly resulting way of modeling an infection that is spread from one human to other human. In Covid case, the infection is both contagious and deadly.

SIRD model

The SIRD model considers the total population that belongs to one of the four states and derives the relationship between the states. The four states are. the susceptible state $S(t)$, the infectious state $I(t)$, the recovered state $R(t)$ and the demised state $\delta(t)$. Various other models are also have been used for this purpose, some may have differential equations and some others may have fractional differential derivatives. This model also takes fatality into account. In this attempt the mean values of the main parameters like basic reproduction rate R_0 , case recovery rate γ and the case infection rate β .

Contact rate $U(N)$ is the number of individuals contracted to the disease by an infective in a unit

time. The probability of infection to each contact is considered as β_0 then the adequate contact rate is traced as $\beta_0 U(N)$. The infection rate can be derived as the product of mean adequate contact rate of an infected individual to the susceptible $\beta_0 U(N) S/N$. The incidence of disease is the total number of newly infected patients by all the individual in the infected category per unit time $(\beta_0 U(N) S/N)I$.

The estimated parameters required for this model includes infection rate, recovery rate and demise rate of the disease in a population and it is gives as,

$$\beta = \frac{\text{the numbet daily confirmed cases at a time}}{\text{the number of accumulated confirmed cases at a time}} \quad (1)$$

$$\gamma = \frac{\text{the numbet daily recovered cases at a time}}{\text{the number of accumulated confirmed cases at a time}} \quad (2)$$

$$\delta = \frac{\text{the numbet daily death cases at a time}}{\text{the number of accumulated confirmed cases at a time}} \quad (3)$$

The SIRD model is described by four differential equations; the first equation represents the rate of susceptibility of cases in a certain population with respect to time period t. And it is represented by an equation,

$$\frac{dS}{dt} = \frac{-\beta s I}{N} \quad (4)$$

The rate of infected cases in a time are represented by the equation (5),

$$\frac{dI}{dt} = \frac{\beta s I}{N} - (\gamma + \delta)I \quad (5)$$

The recovery rate of the total population with respect to a time is given by,

$$\frac{dR}{dt} = I\gamma \quad (6)$$

The rate of demised cases with respect to a time is given the equation (7),

$$\frac{d\delta}{dt} = \delta I \quad (7)$$

Here, N is the total population of the geographic location and β , δ and γ are the rate of infection of patients, rate of demise and rate of recovery in that population. It is clear from all the four equations contain the infected cases I because it is the central aspect of the pandemic. The rate of population of susceptible, infection, recovery and demise are related to each other by the following equation,

$$\frac{dS}{dt} + \frac{d\delta}{dt} + \frac{dR}{dt} + \frac{dI}{dt} = 0 \quad (8)$$

The last equation represents the conservation of the total population in the system and it represents the critical condition of the system. It is very clear in the system that SIRD model

considers a population and divides the population into four categories.

An individual belongs to one of the four categories but not more than one. So the sum of the entire four categories gives the total population as well,

$$S + I + R + D = N \quad (9)$$

The basic reproduction number is taken as R_0 . R_0 can be used to predict whether the infectious disease will spread into a population in near future or die out. R_0 represents the average number of victim's affected secondary cases that results after the introduction of a single infectious patient into a susceptible population during the infectious period. The basic reproduction number can be easily calculated from the parameters of SIRD model.

$$R_0 = \frac{\beta}{\gamma} \quad (10)$$

The values of the parameter in Covid 19 is defined as, the unit of days. The official time for incubation period followed worldwide is 14 day. The SIRD system is defined in a discrete time period $t=1,2,\dots,n$, with the corresponding initial state of the system with its four sub-population is given as $S(0)=N-1$, $I(0)=1$, $R(0)$, $D(0)=0$. The parameters β and γ are not corresponds to the actual per day count in aspects.

System design

The system flow architecture represents the flow of implementation of the system. The first step in building the model is collecting the raw data from the real world. The data which relevant to Covid-19 is taken into account in this case. Then the data is processed for finding the missing values and replacing the redundant with data that are appropriate for building the model. The cleaned dataset is taken as input to the system and SIRD model can be built by evaluating the parameters from the available dataset ad considering the population. The visualization of resultant model enables for better understanding of the data. Product knowledge and insight can be brought from the developed model as a data product. These model generation and results will the authorities for better decision making. In addition to the generation of the model, the cleaned data set can be used to get several exploratory data analysis results and models are made for assisting the generation of the model.

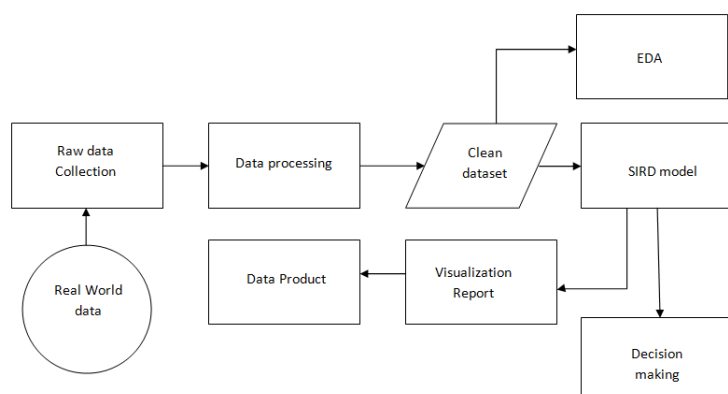


Figure 1. System flow architecture

Data set

The analysis is based on the publically available data that is being released by the Health and Family Welfare Department of Government of Tamil Nadu. The dataset is collected and compiled under circumstances from the daily official press release. Tamil Nadu government faced its first corona case on March 7 2020 and the dataset is continuously updated until September 21 2020. All the dataset is updated accurately as per the daily bulletin through state Government website www.stopcoronatn.com.

Table 1 has the summary details of corona case in Tamil Nadu. It has date in which the infection has occurred, how many patients has been confirmed to the infection, number of recoveries from the disease, how many patients are still in hospitalized state and number of death occurred due to disease.

Table 1.Case summary table of Tamil Nadu.

DATE	Confirmed Cases	Recovered cases	Active cases	Death cases
3/7/2020	1	0	1	0
3/8/2020	0	0	1	0
.....				
9/20/2020	5516	5206	46703	60

Table 2 gives the data about the number of confirmed cases for all the 38 districts in Tamil Nadu. This dataset is also recorded from 7 March to 21 September. The datasets are checked for the completeness and detected for the presence of NaN values. But as the disease continuous to spread in our region, the dataset is continuously updated daily. The result of finding missing values results null in the dataset. The datasets are collected in the form of Comma Separated Values. The dataset are then transformed for developing the model.

Table 2.District wise confirmed cases in Tamil Nadu.

DATE	Ariyalur	Chengalpattu		Villupuram	Virudhunagar
3/7/20	0	0		0	0
.....					
9/21/20	3517	32799		10525	14066

Results and Discussions

The available dataset is reported on a timeline. All the four parameters infected, recovered, active cases and deaths are reported for the available period of time.

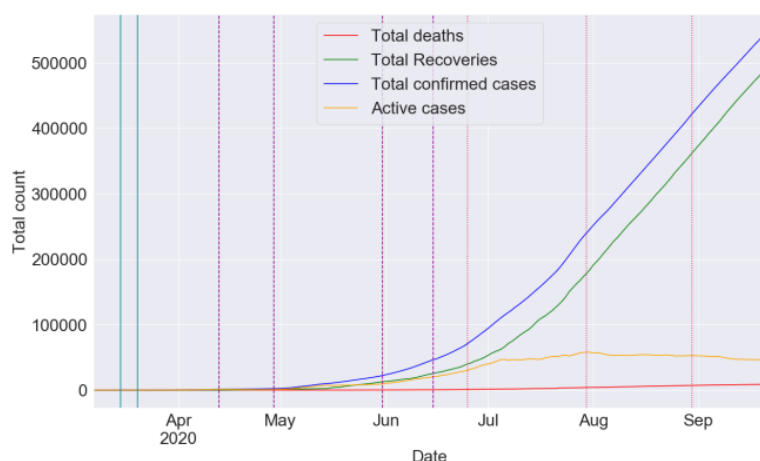


Figure 2. Timeline report of Covid-19 in Tamil Nadu

The vertical lines depict the dates of lockdown. The first two solid lines are for strict lockdown. The dashed lines are for extension of lockdown with some restriction. The rest dotted lines represent the extension of lockdown without border restriction and national level transportation.

The parameters are compared with their moving average (MA) for analyzing the data points for multiple subsets of the entire dataset in a time series. The moving average is computed with the scale of 7 days. The comparison results help to find the outliers. Presence of outlier may cause poor results in performance. Outlier is the data point that does not fit with the rest of the data points.

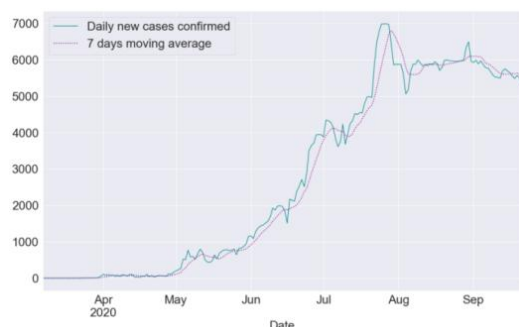


Figure 3. Confirmed cases vs. MA

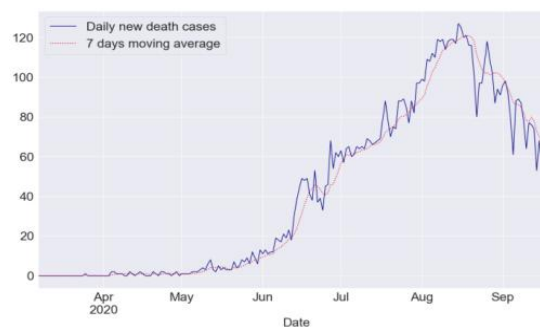


Figure 4. Death cases vs. MA

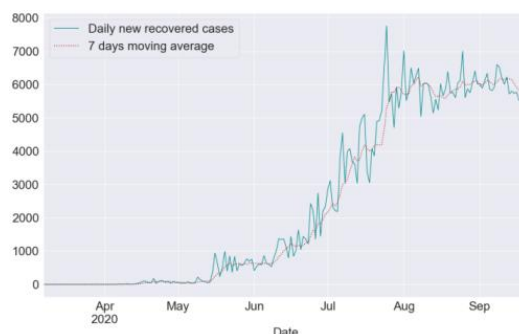


Figure 5. Recovered cases vs. MA

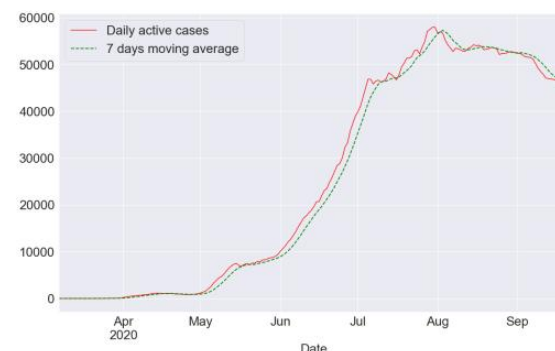


Figure 6. Active cases vs. MA

From Figure 3, 4, 5 and 6 it is clear that actual confirmed cases and its moving are similar from initial stage of disease. The comparison of recovered and death cases with its moving average shows that curve is not so smooth. The curves also have some fluctuation but there is no presence of outlier. The active case comparison with the moving average gives that main idea that both the curves are similar from initial stage to the end of available data period.

The SIRD model is built with four sub population. Susceptible fraction where people do not contract to the disease but vulnerable to get disease from an infected patient. In Covid situation, the disease is contagious and total population was considered in a susceptible state. The infected population is the number of people who are affected by the virus and remains in hospitalized state. The patient is potential of carrying the disease to an unaffected population and spread the disease. In the recovered set of population the patients develops immunity against the disease and hence recovered from the disease. But the recovered patient may or may not get the disease for the second time. In death set, the disease is a fatal disease; the patients are unable to develop immunity which may lead to death. This category of people will never be recovered or transported to other category of population.

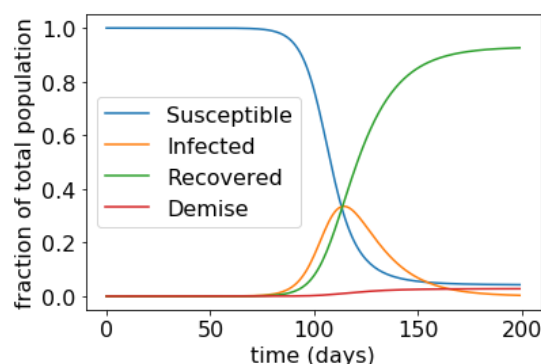


Figure 7. SIRD model without lockdown

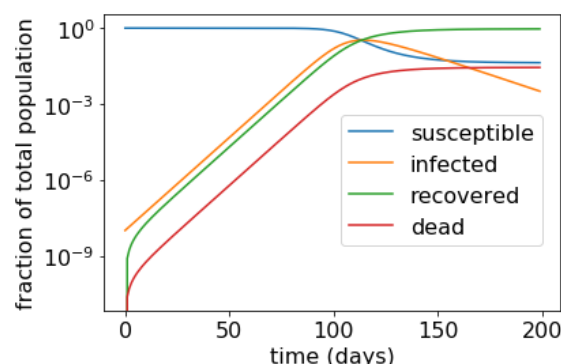


Figure 8. SIRD model with lockdown

Here the model is developed under lockdown scenario, where the total population for under the infected cases would be less when people have less contract rate to the disease. The rate of spreading the disease is very much less when compared to the worst scenario.

From the model, is observed that number of infected cases increases from a certain period of time and then eventually decreases. The susceptible fraction of population decreases as the virus is transmitted and then eventually drops to the absorbent state 0. For recovered cases where it starts from the state 0 it raises to a maximum value. But in the demised case, it starts from 0 and attain the difference value between the infected and recovered value.

Considering the worst case scenario, the population is considered in an open condition i.e. without lockdown. Large number of population will be in susceptible state and many people would get disease easily in this scenario. The number of infected cases will be large and recoveries would be less because of minimum available health facilities for a large affected population.

The parameters used to calculate the model is estimated by the using above differential equations. Tamil Nadu geographic location is considered in our case. The table below shows the parameters for the state.

Table 3. Parameters estimated.

Parameters	Parameter Description	Parameter Value
N	Total population	7.7×10^7
β	Infection rate	0.35
δ	Death rate	0.02
γ	Recovery rate	0.07
R_0	Reproduction number	5

When trying to fit the Tamil Nadu dataset with the calculated model. It is observed that a drop would be expected in the bell curve after 170 days. But the results show that the drop was actually happened after 190 days.

The comparison is made for infected cases as well as recovered cases. The death case scenario, both Figure 2 and Figure 7 shows that number of death would be low. In Figure 9 and Figure 10 the blue solid lines are the actual data points and orange lines are for the smooth predicted curve of SIRD model.

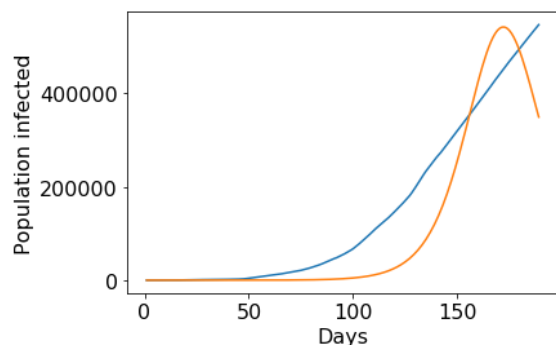


Figure 9. SIRD model fit with infected case in Tamil Nadu

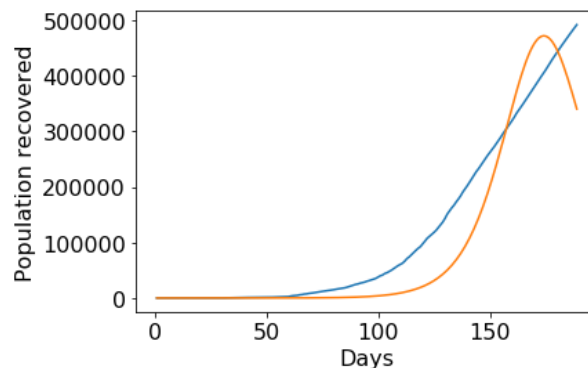


Figure 10. SIRD model fit with recovered case in Tamil Nadu

The goodness of fit of the curve is estimated by Mean square error (MSE), Root mean square error (RMSE) and R^2 . The accuracy gives the how much effectively the model was built against original data.

The table describes the error rate of Confirmed cases and recovered cases. From the calculate results the recovered cases prediction is way better than the infected cases. The error are taken in the way that smaller the estimated error value greater will be the accuracy of calculation.

Table 4. Error estimations.

Attribute	MSE	RMSE	R^2
Confirmed case	21035.92	145.03	0.820
Recovered cases	193.13	13.89	0.63

Conclusion

The SIRD prediction system has been proposed for predicting the risk of Covid-19 in Tamil Nadu. The system analyses dataset containing the day-wise actual data and makes predictions for upcoming days using the compartmental model. The goodness of fit for the curves is also estimated. Observation makes a clear view that date and number of cumulative number of infection are positive correlated. The forecast model shows that the number of death cases will be very less when compared to the recovered cases in Tamil Nadu. Imposition of lockdown is the best move to control the pandemic and avoid the spread of disease for a large population geographic location. There are significant limitations to this model; the fact of how the disease is spread or does not spread throughout the space is neglected. In such case social distancing will be helpful. How different age group and with different health condition are susceptible to the disease. Health advisory reports that aged people with co-morbidities are likely to be affected

more than other people.

Based on the results that the while large population is susceptible to the disease, the number of infected cases grows exponentially. A critical number of people are infected and begins to recover, the number of cases decays exponentially. There would be drop in susceptible and infected cases. The recovered and death cases will be zero at initial state and grows gradually to a maximum values.

As the infection rate is growing day by day the parameter that also growing similar to infected case is the recovered cases. When the number of infected cases the number of recoveries are also less. When the number is high out of the number of infection the number of deaths will be less.

Thus for a new virus pandemic situation SIRD model can be applied to get the vulnerability. In addition to this the same algorithm can be used to get reproduction rate. The primary need for this model is to get updated dataset with the number of observed case scenario for a certain population. The same algorithm can also be used to exact solution by using other epidemic model. At the time writing, vaccination was not taken into considered. In future vaccinated state can be included where susceptibility would be less in such case results in less number of infection in future which influence the number of recovered and death cases also.

References

- [1] Furqan Rustam , Aijaz Ahmad Reshi , Arif Mehmood ,Saleem Ullah , Byung-Won On, Waqar Aslam and Gyu Sang Choi (2020) *COVID-19 Future Forecasting Using Supervised Machine Learning Models*, IEEE ACCESS, 101489-101499
- [2] Romney B Duffey and Enrico Zio (2020) *Analysing Recovery From Pandemics by Learning Theory: The Case of CoVid-19* in IEEE Engineering in Medicine and Biology Society Section, vol.8, 110789-110795
- [3] Petar Radanliev, David De Roure and Rob Walton (2020) *Data mining and analysis of scientific research data records on Covid-19 mortality, immunity, and vaccine development - In the first wave of the Covid-19 pandemic*, Science direct, vol.3, 1121-1132.
- [4] Jooyeon Park, Jinhwa Jang and Insung Ahn (2017) *Epidemic Simulation of H1N1 Influenza Virus using GIS in South Korea*, in 2017 International Conference on Information and Communication Technology Convergence (ICTC), 58-60
- [5] Huadong Xia, Kalyani Nagaraj, Jiangzhuo Chen and Madhav Marathe (2013) *Evaluating strategies for pandemic response in Delhi using realistic social networks* in IEEE International Conference on Healthcare Informatics, 121-130
- [6] Vasilis Z. Marmarelis (2020) *Predictive modeling of Covid-19 data in the US: Adaptive phase-space approach*, IEEE Open Journal of Engineering in Medicine and Biology, pp. 207-211
- [7] Fotios Petropoulos and Spyros Makridakis (2020) *Forecasting the novel coronavirus COVID-19*, (pp 1-8) , ARGENTINA Lidia Adriana Braunstein, Universidad, Nacional de

Mar delPlata

- [8] Hiba Asri, Hajar Mousannif, Hassan Al Moatassime and Thomas Noel (2016) *Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis* in The 6th International Symposium on Frontiers in Ambient and Mobile Systems, ScienceDirect, (pp. 1064 – 1069)
- [9] Spyros Makridakis, Evangelos Spiliotis and Vassilios Assimakopoulos (2018)*Statistical and Machine Learning forecasting methods: Concerns and ways forward* Plos one (Alejandro Raul Hernandez Montoya,Universidad Veracruzana MEXICO), pp 1-26
- [10] Shaobo He. Yuexi Peng and Kehui Sun (2020)*SEIR modeling of the COVID-19 and its dynamics*Springer Nature B.V.
- [11] Sharma and Milan Coronavirus: (2020) Second ICMR report on random sampling test results shows possible community transmission India Today.
- [12] Tamil Nadu State Health & Family Welfare Department daily Bulletin, www.stopcornatn.in , Tamil Nadu Government
- [13] Chen Y., Liu Q. and Guo, D. (2020) *Emerging coronaviruses: genome structure, replication, and pathogenesis*. J. Med. Virol.92, pp 418–423.
- [14] Zhu N., Zhang D and Wang W., (2019) *A Novel Coronavirus from patients with pneumonia in China*, New Engl.J Med, pp 727-733.
- [15] Fanelli D., and Piazzia F (2020) *Analysis and forecast of Covid 19 spreading in China, Italy and France*, Chaos Solitons Fractals, pp 109761-109761.
- [16] Ceraolo C. and Giorgi, F.M (2020) *Genomic variance of the 2019-ncov Coronavirus*, J. Med. Virol, pp 522–528.
- [17] G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne (2012) *Machine learning strategies for time series forecasting* (pp 62-77)Proc. Eur. Bus Intell. Summer School. Berlin, Germany Springer.
- [18] Ji, W., Wang, W., and Zhao x. (2020) *Cross-species transmission of the newly identified coronavirus 2019-ncov*.(pp 433–440) J. Med. Virol. 92.
- [19] J.C. Miller, (2017)*Mathematical models of SIR disease spread with combined non-sexual and sexual transmission routes Infectious Disease Modelling*,vol 2 pp 35-55.
- [20] J. M. Carcione , J. E. Santos, C. Bagaini and J. Ba. (2020) *A simulation of a COVID-19 epidemic based on a deterministic SEIR model*, Front. Public Health
- [21] Hyndman R and Khandakar Y (2008)*Automatic time series forecasting: the forecast package for R*. Journal of Statistical Software, pp 1-22.