

Deep Analysis And Theoretical Investigation Of Covid-19 Pandemic In Iraq Using Data Mining Techniques

Atheer Y.O. Allmuttar^{1,2}, Ali B. Roomi^{3,4}, Sarmad K. D. Alkhafaji⁵

¹ Department of Computer Sciences, College of Education for Pure Science, University of Thi-Qar, Iraq.

² Al-Ayen University, Thi-Qar, Iraq.

³ PhD Biochemistry, Ministry of Education, Directorate of Education Thi-Qar, Thi-Qar-64001, Iraq.

⁴ Biochemistry and biological engineering research group, Scientific Research Center, Al-Ayen University, Thi-Qar, 64001, Iraq.

⁵ Department of Computer Sciences, College of Education for Pure Science, University of Thi-Qar, Iraq.

atheer.oudah@utq.edu.iq, Atheeryousif77@gmail.com, dr.ali_bader@alayen.edu.iq, Dr.sarmad@utq.edu.iq,

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

ABSTRACT

The aim of this paper is to deeply analyze the Corona-Virus Diseases (Covid-19) using data mining based K-Means Clustering technique. Medical Science in data mining is an emerging field that has proposed a lot of advanced techniques in analysis of a particular disease. Treatment of coronavirus is getting more and more challenging due to complex structure, shape and texture of the virus. Therefore, by advancing in data mining, K-Means methodology has been proposed to analyze the covid-19s in the world. The advancement in this field created an urge in me to research more on the techniques and methodologies developed for covid-19 extraction. During the outbreak of an epidemic, it is of immense interest to monitor the effects of containment measures and forecast of outbreak including epidemic peak. To confront the epidemic, a simple K-Means model is used to simulate the number of affected patients of Coronavirus disease in Iraq. The inhibition effect or precautionary measures also influence the spreading of a pandemic. If the inhibition factor increases up to 50%, then 0.5 million patients will be existing in Iraq till the end of this year. This number will exceed 1 million, if precautionary measures decrease to 50%. The worst effects of the disease appear in the community if we remove all the barriers. In such a case, this malady may increase by affecting 55% population till the end of this month. This number will start to decrease after September.

Keywords: Patients, K-Means, Clustering, Data Mining, Covid-19, Reported Cases, Analysis, Iraq.

1. Introduction

This research paper proposes K-Means Clustering based methodology to segment a covid-19 and reported cases from Covid-19 data and determine correlation for all of the disease. For this several data mining techniques have been implemented and analysis is provided regarding the efficiency of the data mining technique used. Each Covid-19 data is passed through an imaging chain where the data is preprocessed to remove noise and is further enhanced to improve the contrast of the data. This paper proposes five different data mining techniques which are then applied to the data to extract the covid-19 as mentioned in [1]. These data mining techniques include data mining, data preprocessing, feature extraction, watershed data mining and conversion. Applying each of the data mining techniques allows us to determine the most appropriate method to segment the covid-19 from each of the data. Normalized cross-correlation on each of the segmented covid-19s is then applied to determine the accuracy. Also, the foreground (covid-19) regions and the background regions are taken to compute cross-correlation between the target variable and covid-19 region of the texture data to determine how cases of the covid-19 region are closely correlated with each other as mentioned in [2]. The covid-19 region represents the case values for the foreground points extracted using the MATLAB command from texture data. The texture data is generated by applying the K-Means Clustering method. In order to enhance the texture characteristics of the data, a smoothing filter is applied to the texture data. This helps me to understand texture features more clearly.

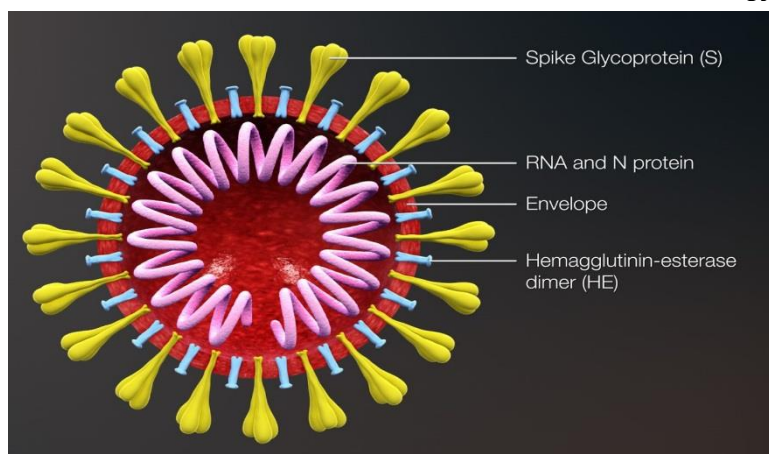


Figure 1: The structure of coronavirus [3].

1.1. Research Problem

The novel Coronavirus (Covid-19) pandemic was affirmed to have arrived in Iraq on February 21st, 2020. The first patient has been observed in Baghdad Iraq in the federal territory of the country [3]. Within a week of the appearance of the initial two cases, this pandemic started to increase in other areas of the country. On 21st of February 2020 first case of covid-19 was reported in Iraq, the quantity of affirmed cases in the nation is standing at 91456 till 19th July 2020, with (25.7% of the cumulative cases) recuperation and 33346 (2.2% of the cumulative cases) deceased and Iraq is, right now, the area with the most elevated number of cases at over 0.5 Million reported cases till December 2020. Applying appropriate data mining techniques followed by data mining operations helps in retaining only the relevant covid-19 portions and suppresses the covid-19 portions of the data to a large extent. This will help in analyzing meaningful portions of the data very efficiently.

1.2. Aim of Study

We want to develop state-of-the-art coronavirus cases and infections based on the available data using K-Means in data mining applied to human body data samples. We provide justifications for selecting K-Means as our base method, which is further extended to directly support multi-scale classification. The work also provides an insight into the theory behind K-Means applied to human body. We describe the input and output representation used in our network and provide the derivation of clustering for the used loss function. In the end, the network architectures are used in this work. This makes the classifier capable of predicting potential infections of covid-19 and reported Covid-19 cases of data from dataset.

- Provide an overview of previous works and achievements on classification of coronavirus and reported cases.
- Apply K-Means Clustering methods to the unified dataset to perform classification with less data loss and more accuracy.
- Evaluate and compare used K-Means Clustering, as well as compare with other studies from the literature.
- Develop an K-Means Clustering based novel methodology with good results for covid-19 and predict the cases up to December 2020.

2. LITERATURE REVIEW

Modeling is a science of creative capabilities connected with profound learning in a variety of strategies to represent physical phenomena in the form of mathematical relations. In the prevailing situation, agencies, which control the diseases and maintain all the data of diseases, are publishing data of Covid-19 on daily bases as mentioned in [4]. This data includes number of people having positive corona tests, number of deaths, number of recoveries and active number of cases, and also cumulative data from all over the world. So, an appropriate model, with much accuracy, is needed at this level. Low dimensional models, with a small

number of compartments and having parameters which can be determined with the real data with good precision, are better to study and forecast the pandemic [5]. A high dimension's model requires a huge number of parameters to describe it but this huge number of parameters cannot be found with enough precision [6]. In the absence of details, compartmental epidemic models describing the average behavior of the system can be a starting point. Even the simplest models contain several variables, which are hard to determine from the available data. The minimal K-Means model describes the behavior of the susceptible $S(t)$, the infected $I(t)$, and the removed (recovered or deceased) $R(t)$ populations [7, 8]. Numerous models have been published on Covid-19 [9]. To the best of our knowledge, it has not been focused on the implications of mathematical model to guess the future trend of Covid-19 disease in Iraq as well as in Iraq. Thus the present study is taken to fill this gap. Hence it was not possible to calculate normalized cross-correlation for the covid-19 extracted using K-Means technique. The author in [11] also tried to superimpose the region using the data mining method. The author in [10] was not able to use the features generated from the Covid-19 data using the data mining method. Therefore, using normalized cross-correlation, Researchers in [13] have determined if the covid-19 extracted has some noise present in it or not for all of the data mining techniques except Covid-19 data.

To estimate the early dynamics of Covid-19 infection in Iraq, we modeled the transmission through a deterministic K-Means model. We are choosing the K-Means model because in present situation worldwide data contains the infectious patients, recovered and deaths only, so from that data we can have an average death rate and recovery. We estimate the size of the epidemic for both countries. We also forecast the maximum level of Covid-19 patients and the time period for approaching the endemic level through model simulations. The dreadful effects of the pandemic, if precautionary measures or social distancing were ended, have also been analyzed. We also perform the sensitivity analysis of the parameters by varying the values of transmission rate, disease-related death rate, recovery rate and the inhibition effect.

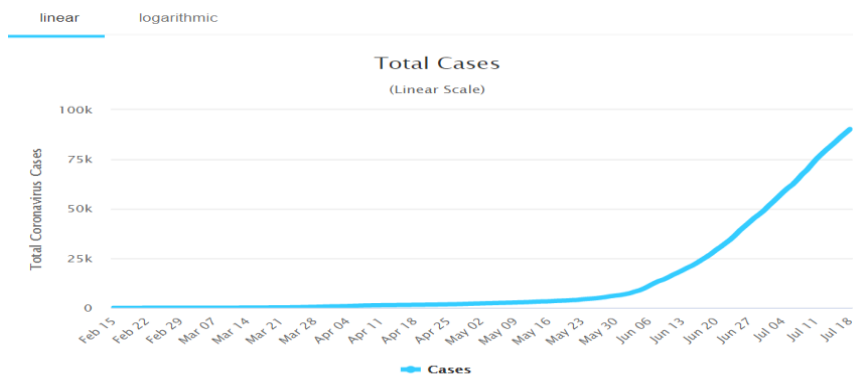


Figure 2: Total covid-19 cases from February 15 – July 18 represented by linear line [14].

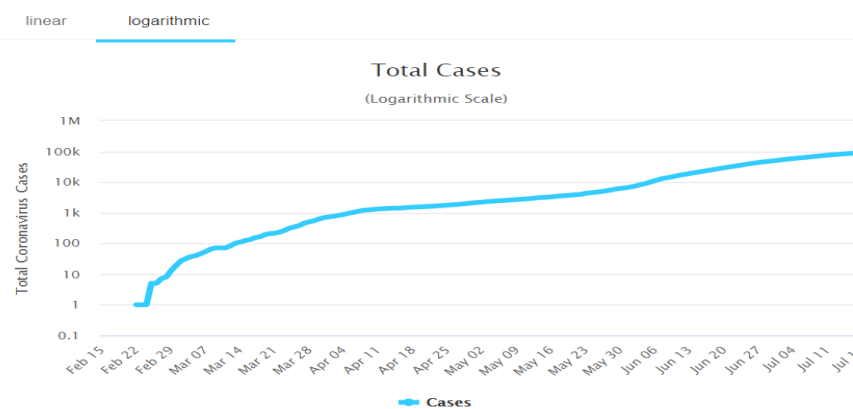


Figure 3: Total covid-19 cases from February 15 – July 18 represented by logarithmic line [14].

In this paper [15], the author used the data that is converted to double and appropriate texture filter method is applied to determine the texture features of the data. In the program, the best texture filter found was Random

Forest technique. In [16] the author tried to enhance the textures of the covid-19 region (foreground) and structure region (background), smoothing filter is applied. In [17] the program, user-defined filter matrix is provided to smoothen the data. Greater the size of the filter, better the features can be seen. The foreground (covid-19) regions and background (structure) regions are extracted using the GMM technique. A target variable vector is initialized which consists of two classes: 0 for structure region and 1 for covid-19 region. Now cross-correlation is applied to compute the correlation between target variable and covid-19 region to determine the degree to which the cases of the covid-19 region are closely correlated with each other. In this paper [18], the cross-correlation is computed using MATLAB. The covid-19 region represents the case values for the foreground points extracted using the data mining from the texture data. In this way, the authors have determined how the cases in the covid-19 region are closely related to each other. However again this value of cross-correlation completely depends on the way we select the foreground and background points using the data mining method. If we do not select the points properly; that is for example if the points of the covid-19 region (foreground) are considered at the edges or borders, there is a possibility that we may not get an accurate cross-correlated value.

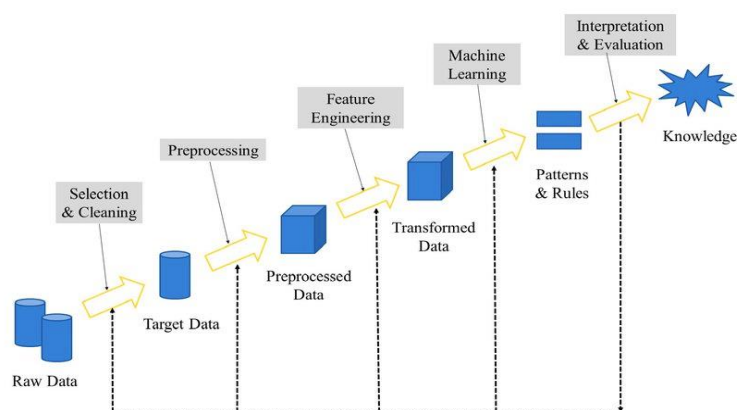


Figure 4: The typical flow of converting raw data into knowledge [19].

Author in [20] has used the methods except manages to extract the covid-19 correctly. Also compared to the first output, we can see some noise in the output. Due to this, the normalized cross-correlation values displayed are not tending towards. This is because of the inefficient threshold values used to segment the covid-19. This is because there is hardly any noise seen in the data. However, this normalized cross-correlation technique is not applied to the covid-19 generated from the data mining technique. The reason is that in reported cases, researchers determined the covid-19 features and plot the regions based on the features extracted using the data mining techniques. Due to this, the region is not superimposed on the original data. The region is just highlighted to display the result. Hence it was not possible to calculate normalized cross-correlation for the covid-19 extracted using the data mining technique. However, in [21] output of the covid-19 data mining is so clear that by just looking at the output, one can predict the accuracy. For eliminating the noise and display better results, better threshold values can be selected. Also in [22], only depicts one covid-19 is extracted instead of both the covid-19s. This has resulted due to two possibilities. First is the area range provided is smaller which neglects the required covid-19 portion of the data. The second possibility is due to a higher threshold value. Therefore, in this paper [23] which is presented on increasing the area range and decreasing the threshold, the region of interest can be extracted accurately. In this way unnecessary noise can be eliminated and correlation values that are tending towards one are displayed. However, in paper [24] the author devised any of the data mining techniques, doing any changes may result in accurate results for some data and produce noisy output for some data. Also after extracting the foreground and background portions of the data, researchers have determined how the cases in the covid-19 region are closely related to each other by computing the cross-correlation between target variable and covid-19 region.

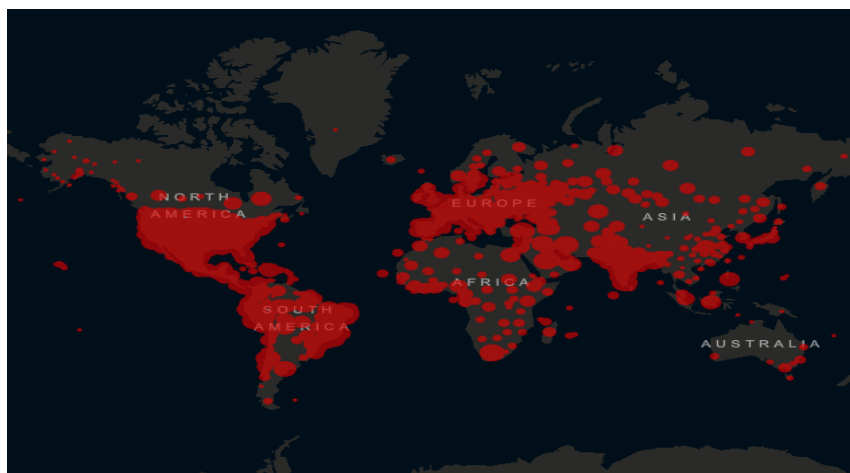


Figure 5: Cumulative confirmed cases around the world [25].

In this paper [26], the idea was not successful because author was unable to extract the proper features accurately using those methods. Therefore, this idea did not help me in computing the correlation coefficient. These problems would have occurred due to the exposure of the data. Due to this reason, author extracted the foreground and background cases accurately and determined cross-correlation between foreground cases of the texture data and target variable. This foreground and background points are obtained using the data mining from the texture filtered data. Also using K-Means data mining method, authors in [27] has extracted the covid-19 and structure regions to compute cross-correlation between the target variable and covid-19 region. The target variable is a vector consisting of human body represents the covid-19 region in [19] which represents Iraq region. In this paper [20], the covid-19 region represents the case values for the foreground points extracted using the python programming from the texture data. This texture filtered data is obtained using the texture filter command, using data mining method. In order to enhance the texture of the foreground and background region, smoothing filter is applied on the texture generated data as described in [28]. In this way cross-correlation is used to determine how cases in the covid-19 region are closely related to each other presented in [29]. However again this value of cross-correlation completely depends on the way we select the foreground and background points using the data mining method. If we do not select the points properly; that is for example if the points of the covid-19 region (foreground) are considered at the edges or borders, there is a possibility that we may not get an accurate cross-correlated value.

Also the proposed data mining technique fails to produce accurate results if the data is over exposed and has no covid-19 in the data. This is because each of the data mining techniques segments the brightest portion of the data based on the area range and the threshold value specified as described in [30]. Therefore, these data mining algorithms consider the over exposed white regions as covid-19 and display the needed result. Hence, it does not produce an output as expected as shown in [31]. Also there are data where they are over-exposed and covid-19 regions are present. In such cases, along with the covid-19 regions, the over-exposed portion is also considered as a covid-19 which leads to inaccurate results presented in [32]. Similarly, this technique also fails if the data is under exposed and has a covid-19 present. This is because, the covid-19 region that is expected to be extracted will not be segmented due to inappropriate threshold value as described in [33]. These problems occur due to the exposure of the data. Hence if we are successful in controlling the exposure of the data by using histogram equalization, there is a possibility that we could get better results. This texture filtered data is obtained using the texture filter command, data mining method. This computation is done to determine how cases of the covid-19 region are closely related to each other. Compared the efficiency and accuracy of covid-19 extracted after every data mining technique using data mining. The covid-19 cases extracted are compared with the data to check if noise is present in the data. If noise is present, we can conclude that covid-19 segmented is not accurate as it has some noise in it. Otherwise, the covid-19 segmented is accurate and correct as presented in [34].

3. Materials & Methods

The proposed methodology initially involves applying different preprocessing techniques the data to sharpen the relevant portions of the covid-19 data after that K-Means Clustering is applied. We also forecast the maximum level of Covid-19 patients and time period for approaching the endemic level through model simulations. The dreadful effects of the pandemic, if precautionary measures or social distancing were ended, has also been analyzed. We also perform the sensitivity analysis of the parameters by varying the values of transmission rate, disease-related death rate, recovery rate and the inhibition effect.

3.1. Analyzing the Covid-19 Outbreak

The coronavirus 2019-20 (COVID-19) pandemic was affirmed to have arrived in Iraq on 24th February of this year. Due to the spread of coronary disease in China, the government of Iraq reported two weeks of isolation, starting from 21st April, for its residents which were coming back from the influenced regions. On the very next day, the Iraq government declared a few preventive measures, including assignment of five clinics as separation-habitats for new cases, arrangement of warm scanners on airport terminals, and uniquely assigned lines for travelers originating from zones influenced by COVID-19 outbreak. For avoiding the virus expansion several steps were taken by the government like on 9th May the authorities reported discontinuance of trips to and from China via all terminals also the Special National Emergency Situations Committee ordered to close all schools on the same day. Two days later, on 11th May, the Government distributed a rundown of the fifteen rules in regards to the mindful social conduct in forestalling the spread of Covid-19. Specialists have forced a prohibition on all religious, scientific, sports, social or diversion occasions with more than 100 members for the next three weeks.

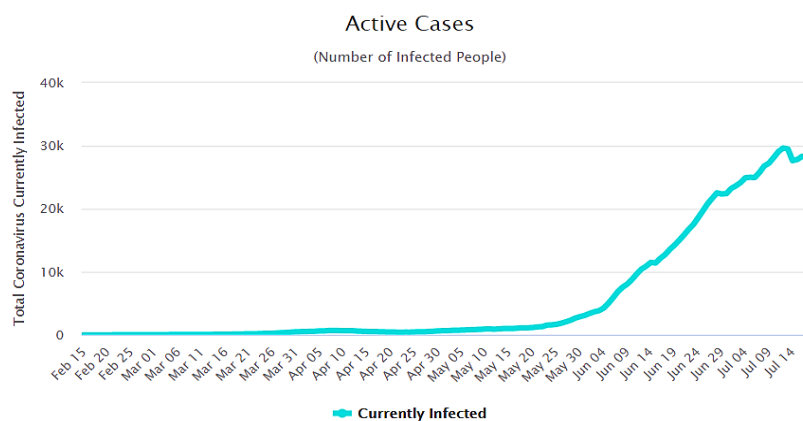


Figure 6: The total active cases in Iraq as of July 18, 2020 [14].

3.2. Dreadful Effects of Removal of Social Distancing and Precautionary Measures

We know that the major factor to avoid from the Covid-19 are social distancing and precautionary measures, in our model we have considered v as this major factor. Now, if we have the present scenario and we consider do not take care of v then we can see from the figure that almost 33% population of the whole country will be infected till 19th of July, 2020 and this is the peak of infection after this it will start decreasing. According to the present recovery rate, disease-related death rate, and estimated values of transmission rate, we observe that if we remove the social distancing and adopted precautionary measures then the worst effects appear in the population. Almost 55% population will be infected up to 31st May and then infected people will begin to decrease. Note that this situation will according to the current position. It means that it will happen only according to current transmission rate, recovery rate, disease related death rate. However, the situation may vary with the variation of these parameters.

3.3. K-Mean Clustering

K-Mean is clustering algorithm that works on vector quantization by selecting appropriate data levels, we can successfully extract the covid-19 from the dataset. However, this method sometimes misses out relevant portions of the data or sometimes considers unwanted portions of the data. Therefore, there may be situations where only a portion of covid-19 may be extracted or certain structure regions along with the covid-19 may be displayed. The regions that are displayed are completely dependent on the global threshold values that are

computed by the data mining function in MATLAB. This data includes number of people having positive corona test, number of deaths, number of recoveries and active number of cases, and also cumulative data from all over the world. So, appropriate model, with much accuracy, is needed at this level. Low dimensional models, with small number of compartments and having parameters which can be determined with the real data with good precision, are better to study and forecast the pandemic

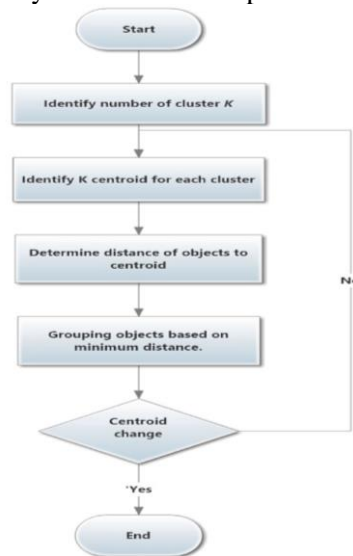


Figure 7: The basic k-mean clustering algorithm.

3.4. Structure of Model

In K-Mean type clustering model, the total population is partitioned into three categories, the susceptible (S), the infectious (I), and the recovered (R). If the homogeneous mixing of people is assumed, the mathematical form of the model is given as:

$$\begin{aligned} \frac{dS}{dt} &= \mu - \frac{\beta IS}{1 + \nu I} - \mu S, \\ \frac{dI}{dt} &= \frac{\beta IS}{1 + \nu I} - (\alpha + \mu + \delta)I, \\ \frac{dR}{dt} &= \alpha I - \mu R. \end{aligned}$$

In the above model, we assume that birth and death rate is equal and is denoted by μ . The parameter β is the transmission rate as a result of the contact of susceptible individuals with the infected ones. The incidence term is assumed to be nonlinear and is represented as $\frac{\beta IS}{1 + \nu I}$. The parameter ν represents the inhibition effect or precautions that have been adopted to prevent the mixing of susceptible and infectious individuals. We assume that the recovery rate of infectious individuals is α and δ is the disease-related death rate. This returns a labeled matrix which consists of positive integer values for different regions and for ridge lines. This data is not very useful as there is only one catch basin spanning the entire data. Here catchment basins are the regions we want to identify. Therefore, we take the compliment of our data and apply K-Means Clustering on the complimented data after which we negate the distance in order to determine the bright catchment basins that represents individual regions.

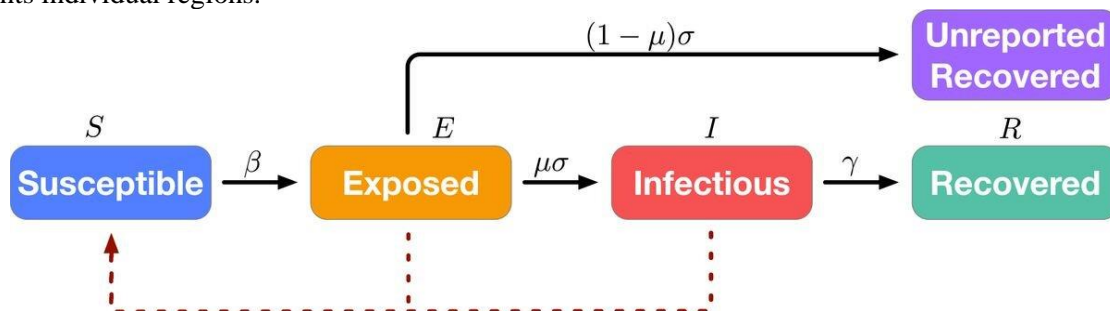


Figure 8: The structure of mathematical model represented by the block diagram.

4. Results

The K-Means Clustering manages to show fairly accurate results for all of the datas. The covid-19 region is represented by a white portion for thresholding and data data mining. In data data mining, the covid-19 is represented by a gray region whereas for analysis of the covid-19 extracted. For K-Means Clustering, each of the layer is sub plotted from where we can understand the layer in which the covid-19 is present. The normalized correlation values that is obtained by computing the case values between the covid-19 extracted and Covid-19 data also has values. This is because, there is hardly any noise seen in the data. However this technique is not applied on the covid-19 generated from the K-Means Clustering technique. The reason is because in K-Means Clustering, we determine K-Means Clustering features and plot the regions based on the features extracted using the Matlab functions. Due to this, the region is not actually superimposed on the original data. The region is just highlighted to display the result. Hence it was not possible to calculated normalized cross-correlation for the covid-19 extracted using K-Means Clustering technique. However by looking at the output, K-Means Clustering produces the best result as compared to other results.

Also after extracting the foreground and background portions of the data, we have determined how the cases in the covid-19 ans seizure region are closely related to each other by computing cross-correlation between target variable and covid-19 region. The covid-19 region represents the case values for the foreground points extracted using the Matlab built-in commands whereas the target variable consists K-Means Clustering technique are used represent structure region and represent covid-19 region and seizure region. This foreground and background points are obtained using the Matlab commands from the texture filtered data.

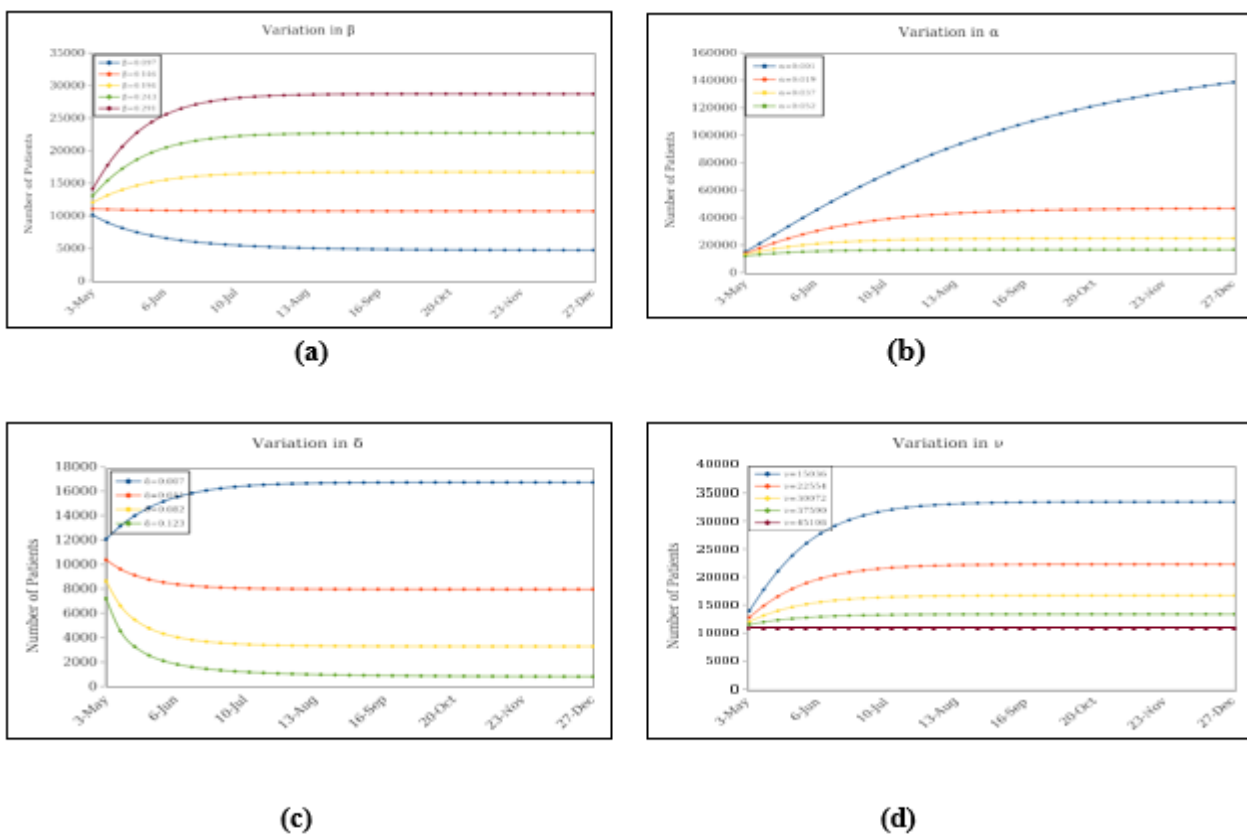


Figure 9: Variation in the number of active patients on transmission rate β , death rate δ , recovery rate α and the inhibition effect ν .

Figure 9a represents the dependence of number of patients on the variation of transmission rate β . This rate tells that how many people are getting infection per day. For example if $\beta = 0.097$ then it means that 97 people are getting infection per day per 1000 people. We have taken five different values of β including the

model fitted value $\beta = 0.194$, we can see that by increasing the transmission rate number of cases are also increasing as expected, Table 1 contains all the possible number of patients.

Next, we will check the dependence of number of active cases on recovery rate, α . It is the rate which tells that how many people are getting immunity from this disease. For example if $\alpha = 0.001$ then it means that out of 1000 people one person is recovered per day. We have taken four different values of α , one is our model fitted value which is $\alpha = 0.015$ and three from the real data, by observing the real data we perceived that average recovery rate is maximum for the week 19th - 25th June, 2020 which is 0.037 and minimum for the week 15th- 21st June, 2020 which is 0.001, so we have considered these two values and fourth is the average of 0.037 and 0.001. Figure 9b represents the trend of active cases depending on α , we can see that number of Covid-19 cases is inversely proportional to the recovery rate α , which makes sense. All the possible number of cases for all these values of α .

Next, we will see that how the death rate δ effects the number of Covid-19 cases. It is the rate which tells that how many people die from this disease. For example if $\delta = 0.007$ then it means that out of 1000 people seven people die per day. We have taken four different values of δ , one is our model fitted value which is $\delta = 0.00703844071$ and three from the real data. We have seen that average death rate is minimum for the week 19th - 25th June, 2020 which is 0.004 and maximum for the week 15th - 21st June, 2020 which is 0.00122985. Fourth is 0.0008, it is the average of 0.004 and 0.001. Figure 9c is depicting the number of active cases as a function of δ . We have calculated the number of Covid-19 cases for all these values of δ . In Figure 9d, we present our results for the number of patients as a function of the inhibition effect ν . The model fitted value of ν is 30072. Since this number can also vary, we have taken four other values of ν in Figure 9d. Since ν is proportional to the precautionary measures adopted by the Covid-19 patients along with the general population, higher values of ν means lower the number of active patients. The values that we have chosen for ν other than the model fitted value are $\nu = 15036.1, 22554.2, 37590.2, 45108.3$. We can see in Figure 9d that the total number of Covid-19 patients ranges 0.5 million cases.

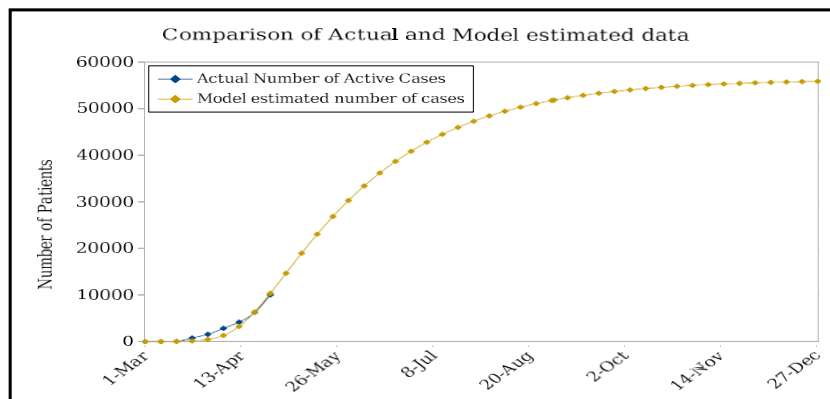


Figure 10: Epidemic curve of COVID-19 patients in Iraq.

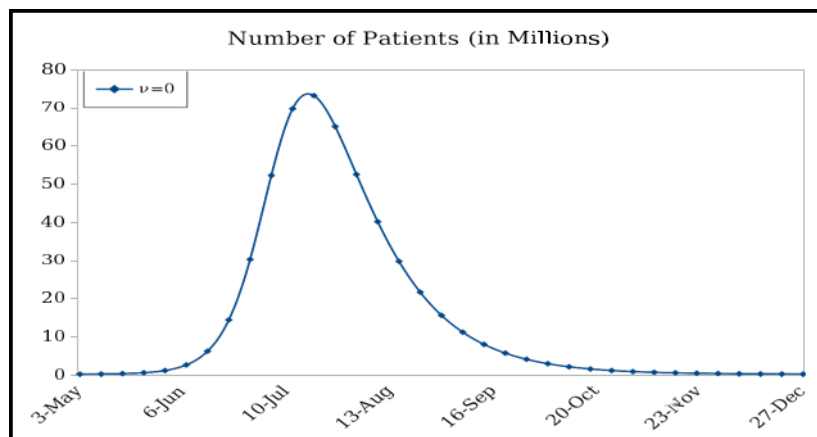


Figure 11: Comparison of actual data with estimated data and future prediction.

Table 1: Weekly Expected Number of active cases, for Iraq, for the next month's according to current situation.

Date	Estimated Number of Cases	Date	Estimated Number of Cases
03-May	2296	06-Sep	182257
10-May	2603	13-Sep	192767
17-May	3404	20-Sep	203211
24-May	4469	27-Sep	213599
31-May	6179	04-Oct	223937
07-Jun	12366	11-Oct	244230
14-Jun	20209	18-Oct	264487
21-Jun	30868	25-Oct	294710
28-Jun	45802	01-Nov	304903
05-Jul	60476	08-Nov	315073
12-Jul	77506	15-Nov	325218
19-Jul	91456	22-Nov	355345
26-Jul	97212	29-Nov	375458
02-Aug	108358	06-Dec	395552
09-Aug	149361	13-Dec	425636
16-Aug	159237	20-Dec	485708
23-Aug	164005	27-Dec	515770
30-Aug	171674	31-Dec	555803

5. Discussion

The result shows that K-Means Clustering is a powerful tool for data mining analysis. In the research, the classification using K-Means Clustering technique proposed have been tested only on dataset. The data collected is taken under different lighting conditions and at different cases reported. Also normalized cross-correlation has been performed to determine the efficiency of the covid-19 extracted. This normalized cross correlation is applied on all the data mining techniques except in K-Means Clustering. Cross correlation has been computed between foreground and background cases of the texture feature extracted data to determine how the cases of the covid-19 region are closely correlated with each other. Each methodology has some drawbacks and may work for some data and work for all data. Using K-Means Clustering technique to classify the covid-19 shows best result of the covid-19 region for the area range and data mining value specified for most of the data as compared to other data mining techniques. If the transmission rate in Iraq increases 50% and recovery rate and disease related death rate is taken for 30th April, according to reported data, then there will be 9000 persons carrying Corona malady and if this rate decreases 50%, then 0.5 Million infected persons will exist in the Iraqi community by the end of this year. If we take previous average maximum weekly recovery rate and disease related death rate, then there will be 5000 and 10,000, patients, respectively in Iraq. Similarly, by assuming minimum weekly average recovery and disease related death rate will result 515770 and 555803, respectively. The inhibition effect or precautionary measures also influence in the spreading of pandemic. If the inhibition factor increases up to 50%, then 0.5 million patients will be existing in Iraq till the end of this year. This number will exceed to 1 million, if precautionary measures decreases to 50%. The worst effects of the disease appear in the community, if we remove all the barriers. In such case, this malady may increase by effecting 55% population till the end of this month. This number will start to decrease after September.

6. Conclusion

In this study, we used a mathematical model to assess the feasibility of the appearance of Covid-19 cases in Iraq as well as the ultimate number of patients according to the current situation. We have used the K-Means

clustering on MATLAB for analysis. By comparing model outcomes with the confirmed cases, it has been observed that our estimated values have good correspondence with the confirmed numbers. If the current pattern is going on then, according to our estimate, there will be 0.5 million infectious individuals in Iraq by the end of this year. Iraq will bear the burden of 555803 till the end of December, 2020. The situation will vary by the variation of transmission rate, death rate, recovery rate, and further implementation of social distancing in Iraq. It has been observed that average weekly recovery rate and average weekly disease related death vary for Iraq. In such case, this infection may increase by effecting 33% population till the end of this month. This number will start to decrease after December, 2020. Although these estimates may vary with the passage of time but it will really help us to observe the most influential factors that causes to increase the epidemic. On the basis of this analysis competent authorities may design the most effective strategies in order to control the epidemic.

References

- [1] L. D. Izquierdo, et al., Informe tecnico nuevo coronavirus 2019–ncov, Ph.D. thesis, Instituto de Salud Carlos III (2020).
- [2] W. H. Organization, Novel coronavirus (2019–ncov) situation reports. 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- [3] Al-Salih M, Samsudin S, Alsalih SW, Arshad SS, Warid F, Sfoog AA, Abed RE, Roomi AB. Identify human cluster of differentiation 147 (CD147), a new target of SARS-CoV-2 invasion. International Journal of Pharmaceutical Research. 2020 Apr 1;12(2).
- [4] C. I. Paules, H. D. Marston, A. S. Fauci. Coronavirus infections more than just the common cold, *Jama* 323. (8);707-708,2020.
- [5] L. Saif, Animal coronavirus vaccines: lessons for sars., *Developments in biologicals*.119;129-140, 2004. URL <https://europepmc.org/article/med/15742624>
- [6] H. Jard'on-Kojakhmetov , C. Kuehn , A. Pugliese , M. Sensi. A geometric analysis of the SIR, SIRS and SIRWS epidemiological models. arXiv preprint arXiv:2002.00354. 2020 .
- [7] J. T. Wu, K. Leung, G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*. 395(10225): 689-697,2020. doi:10.1016/S0140-6736(20)30260-9.
- [8] W. O. Kermack, and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of London. Series A*, 115(772), pp.700-721,1927.
- [9] F. Brauer , C. Castillo-Chavez , Z. Feng . *Mathematical Models in Epidemiology*. Springer New York; 2019.
- [10] A. J. Kucharski, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*. 2020; Online First.
- [11] M. Kochan'czyk, F. Grabowski, and T. Lipniacki, Dynamics of COVID-19 pandemic at constant and time-dependent contact rates. *Mathematical Modelling of Natural Phenomena*, 15, 28, 2020. doi:10.1016/S1473-3099(20)30144-4.
- [12] Roomi AB, Al-Salih RMH, Ali SA. The effect insulin therapy and metformin on osteoporosis in diabetic postmenopausal Iraqi women. *Indian J Public Health Res Dev* 2019;10(4):1544-1549.
- [13] M. Banerjee, A. Tokarev and V. Volpert. Immuno-epidemiological model of two-stage epidemic growth. *Mathematical Modelling of Natural Phenomena*, 15, 27, 2020.

- [14] Available Online: <https://www.worldometers.info/coronavirus/country/iraq/>
- [15] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. MedRxiv. March 2020. doi:<https://doi.org/10.1101/2020.02.11.20022186>.
- [16] Roomi AB, Nori W, Hamed RM. Lower Serum Irisin Levels Are Associated with Increased Osteoporosis and Oxidative Stress in Postmenopausal. Rep Biochem Mol Biol 2021;10(1):13-19.+
- [17] V. Volpert, M. Banerjee, and S. Petrovskii, On a quarantine model of coronavirus infection and data analysis. Mathematical Modelling of Natural Phenomena, 15, 24,2020.
- [18] Radio Iraq International - Measures against the coronavirus. Radio Iraq International. Archived from the original on 26th February 2020.
- [19] On the feature engineering of building energy data mining - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-typical-data-mining-procedure_fig1_323674394,2020. [accessed 17 Jul, 2020]
- [20] Y. Tan, An improved KNN text classification algorithm based on K-medoids and rough set. In Proceedings of the 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 25–26 August; Volume 1, pp. 109–113,2018.
- [21] C. Conner, J. Samuel, A. Kretinin, Y. Samuel, L. Nadeau. A Picture for The Words! Textual Visualization in Big Data Analytics. Northeast Bus. Econ. Assoc. Annu. Proc. 46, 37–43,2019.
- [22] Y. Samuel, J. George, J. Samuel. Beyond STEM, How Can Women Engage Big Data, Analytics, Robotics & Artificial Intelligence? An Exploratory Analysis of Confidence & Educational Factors in the Emerging Technology Waves Influencing the Role of, & Impact Upon, Women. arXiv, arXiv:2003.11746,2020.
- [23] K. Svetlov, K. Platonov. Sentiment Analysis of Posts and Comments in the Accounts of Russian Politicians on the Social Network. In Proceedings of the 2019 25th Conference of Open Innovations Association (FRUCT), Helsinki, Finland, 5–8 November; pp. 299–305,2019.
- [24] H. Saif, M. Fernández, Y. He, H. Alani. On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter; European Language Resources Association (ELRA): Reykjavik, Iceland, 2014.
- [25] Available Online: <https://coronavirus.jhu.edu/map.html>
- [26] K. Ravi. V. Ravi. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowl. Based Syst., 89, 14–46,2015.
- [27] V. K. Jain, S. Kumar, S. L. Fernandes. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. J. Comput. Sci. 21, 316–326,2017.
- [28] Nori W, Abdulghani M, Roomi AB, Akram W. To operate or to wait? Doppler indices as predictors for medical termination for first trimester missed abortion. Clin Exp Obstet Gynecol 2021;48(1):169-175..
- [29] I.C.H. Fung, J. Yin, K. D. Pressley, C. H. Duke, C. Mo, H. Liang, K. W. Fu, Z.T.H. Tse, S. I. Hou. Pedagogical Demonstration of Twitter Data Analysis: A Case Study of World AIDS Day, 2014. Data. 4, 84,2019.
- [30] E.H.J. Kim, Y. K. Jeong, Y. Kim, K. Y. Kang, M. Song. Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. J. Inf. Sci. 42, 763–781,2016.

- [31] J. Samuel, N. Ali, M. Rahman, M.; Y. Samuel, A. Pelaez. Feeling Like it is Time to Reopen Now? COVID-19 New Normal Scenarios Based on Reopening Sentiment Analytics. arXiv, arXiv:2005.10961,2020.
- [32] R. Nagar, Q. Yuan, C. C. Freifeld, M. Santillana, A. Nojima, R. Chunara, J.S. Brownstein. A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *J. Med Internet Res.* 16, e236, 2014.
- [33] B. K. Chae, Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *Int. J. Prod. Econ.* 165, 247–259,2015.
- [34] J. P. Carvalho, H. Rosa, G. Brogueira, F. Batista. MISNIS: An intelligent platform for twitter topic mining. *Expert Syst. Appl.* 89, 374–388,2017.