ASEAN Engineering Journal

Full Paper

AN ESTIMATION OF NUMBER OF DAILY DEATHS DUE TO COVID-19 IN UNITED STATES OF AMERICA, UNITED KINGDOM AND TURKEY

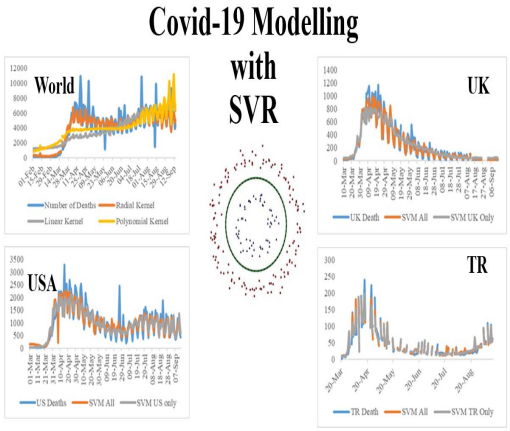
Ezgi Deniz Ülker[a], Sadık Ülker[b*]

[a]Department of Software Engineering, Faculty of Engineering, European University of Lefke, Mersin-10, Turkey
[b]Department of Electrical and Electronics Engineering, Faculty of Engineering, European University of Lefke, Mersin-10, Turkey

Article history
Received
24 September 2021
Received in revised form
05 July 2022
Accepted
06 July 2022
Published online
28 February 2023

*Corresponding author
sulker@eul.edu.tr

Graphical abstract



Abstract

Covid-19 virus is threatening the world with health, social and economic implications and all around the world data is obtained continuously with pandemic for modelling and predicting the future. In this work, support vector regression technique was used to make some predictions on the daily death values due to Covid-19 virus. The models were created for the world, United States of America, United Kingdom and Turkey. All the regression models were tested using coefficient of determination (R^2) and root mean square error (RMSE) values. The analysis was also conducted for comparing the suitability of linear, radial and polynomial kernels. The radial kernel produced relatively better results. In predicting the world data support vector regression with radial kernel produced 0.805262 R^2 value on test data. In the models created for United States of America 0.723376 R^2 value, for United Kingdom 0.95412 R^2 value and for Turkey 0.875343 R^2 value using test data were observed. Also, while the models were created for specific countries the comparisons were made between using only data for the country and also using the whole world data. In general modelling using the data for the world combined with the country data gave better prediction.

Keywords: Covid-19, Kernel, Pandemic spread, Modelling, Support vector regression.

1.0 INTRODUCTION

In January 2020, no one would have guessed that world would be in a state with almost all the countries being affected with the spread of a pandemic, Covid-19 virus. Firstly, thought to be spread from animals to humans in Wuhan City, China, the virus first affected the eastern part of the world in the first couple of weeks and then in March 2020 started to be a threat to the rest of the world. By October 2020, more than 42 million people got contaminated by the virus and more than 1,150,000 lives were claimed due to this virus. As of August 2021, more than 218 million people got contaminated by the virus claiming more than 4,500,000 lives. The virus is currently one of the most important health threats to the world. Nonetheless the Covid-19 has been considered pandemic due to several reasons; i) the disease has high transmission rate, ii) virus itself is stronger than other viruses, iii) the virus getting mutated to different variations. In order to prevent and control to a certain extent the spread, besides the efforts of medical research, all the people are advised to practice hygienic conditions and obey the social distancing rules.

Besides the medical and pharmaceutical research going on about the virus, many researchers around the world are trying to model the virus spread and make certain predictions beyond the virus. Vespignani et al. mentioned about different modelling cases for Covid-19 virus [1]. An epidemic model was formulated that describes the spread of the epidemic in a population [2]. A mathematical model using nonlinear fractional differential equations was recently suggested by Abdo et al. [3]. Temperature, humidity and latitude analysis to predict the potential spread of the virus was also researched by Sajadi et al. [4]. A different approach, by training the 2003 SARS data and using Susceptible-Exposed-Infectious-Removed model to derive the Covid-19 epidemic using artificial intelligence technique was performed by Yang et al. [5]. Analysis and forecasting of Covid-19 spread in China, Italy and France using non-linear fitting was also researched by Fanelli and Piazza [6].

A predictive mathematical model was suggested by Djilali and Ghanbari for South Africa, Turkey and Brazil recently [7].

In the light of these studies, we noticed one more time the importance of the infectious diseases that can spread so easily and the damage that they cause not only to the health systems of the countries but the economies and social aspects of the countries as well. In this work, the aim was to construct a model for the world and in specific three countries separately to predict the number of daily deaths. The work was conducted for considering the data for the whole world and also three countries namely United States of America (USA), United Kingdom (UK) and Turkey (TR). These three countries are from the different continents and naturally have different Covid-19 spread and case/death ratios. The models were created using support vector regression techniques and using various parameters the model suitabilities were studied.

2.0 SUPPORT VECTOR REGRESSION

Support vector algorithm, originally developed as a nonlinear generalized portrait algorithm in 1960s by Vapnik and Lerner [8]. Later, the support vector machine was developed by Vapnik, became very popular in classification and prediction problems. Their importance and success in prediction became more apparent with the development of computers and programming.

The basic methodology of support vector regression was outlined in some studies [9, 10]. Use of different kernels in support vector regression can lead to differences in predicted values and many applications and comparisons can be found in literature. Conditionally positive definite kernels for image recognition was performed by Boughorbel et al. [11], approximation with polynomial kernels was performed by Zhou and Jetter [12], customizing kernel functions for image classification problems was researched by Guo et al. [13] and use of different kernel functions for comparing performance in unemployment rate and GDP prediction was performed by Ulker and Ulker [14]. Also, it is proved to be a powerful artificial intelligence technique for regression and classification and found many different applications in the literature ranging from health sciences to complex design problems [15-20].

In this work our aim was to create a prediction model to observe the number of daily deaths in the world and in three different countries with different characteristics, USA, UK and TR. The extent of prediction of number of daily deaths was to observe various findings. First, we wanted to see if the model can estimate a regression model for the world, using the date data as well as the number of daily newly confirmed cases to predict the number of daily deaths. Secondly, we wanted to create a model for each country using their own statistics and compare it with the model that can be created with the world data. Also, as a third comparison, we wanted to see the difference between radial, linear and polynomial kernels in terms of modelling and performance.

The methodology was as follows: use the day and number of daily confirmed cases to predict the number of daily death value in the world. In addition to this, repeat this procedure for three countries using support vector regression technique using different kernels and compare the coefficient of determination (R^2) and root mean square error (RMSE) values to predict the accuracy of the models. Also, as a second analysis for the countries, the daily data for the world (i.e. number of confirmed cases daily in the world and number of daily deaths in the world) as well as the number of daily confirmed cases for the country in question were used as the input parameters and number of daily deaths was used as the output parameter.

2.1. Modelling

As mentioned previously models are created for the world and for three specific countries. The data was obtained from virusncov.com [21]. The data obtained included: number of new daily confirmed cases in the world, number of daily deaths in the world, number of daily newly confirmed cases for USA, UK, TR, number of daily deaths for USA, UK and TR.

2.1.1. Modelling For The World

In modelling the world, data was obtained between the dates of 1st of February 2020 till 14th of September, 2020 considering 225 days. The daily data included the number of new cases of infection due to Covid-19 virus in the world and number of daily deaths in the world.

In setting up the model we used the number of new daily confirmed cases of Covid-19 as the input parameter and we considered the number of daily deaths as the output parameter which needed to be predicted. Out of these 225 data, 205 of them were used in training and 20 of them in testing. Also, with the created model all the results are plotted for better comparison of the model and the real data. Figure 1 shows the comparison and Table 1 tabulates the summary of results. As it can be observed clearly from the figure, radial kernel showed better modelling compared to linear and polynomial kernels. This was also reflected in the R^2 and RMSE values. Support vector regression with radial kernel produced 0.805262 R^2 value using the test data and 1210.15 RMSE value which indicated good modelling. Similarly, when the whole data is considered the R^2 value was 0.805095 and RMSE value was 1074.755.

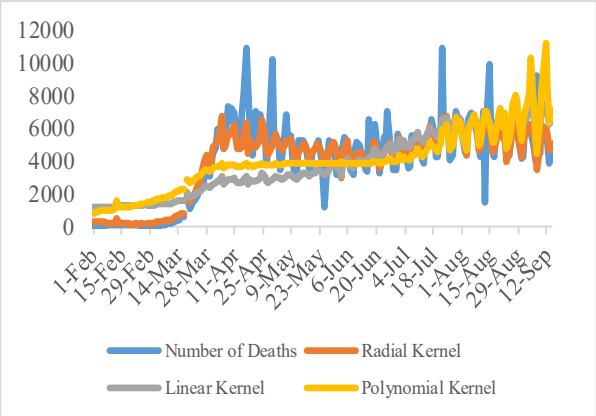


Figure 1 The number of daily deaths predicted for the world (1st February -14th September 2020)

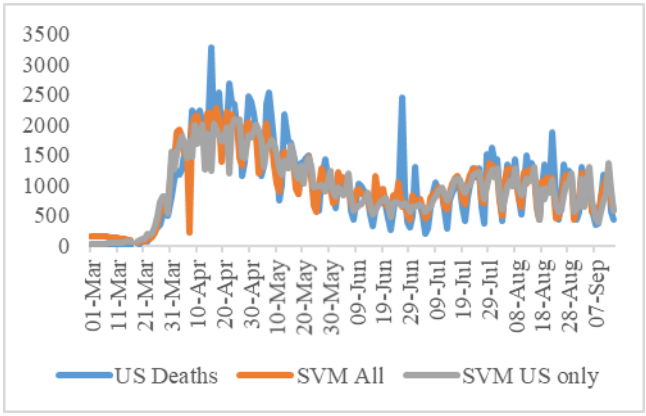


Figure 2 The number of daily deaths predicted for the United States of America (1st March -14th September 2020)

Table 1 Predicted Error in Numbers of Daily Death Values Using Three Different Kernels

Kernel	All Data		Test Data	
	RMSE	R^2	RMSE	R^2
Radial	1074.75	0.805095	1210.15	0.805262
Linear	1857.26	0.417959	1735.41	0.4555786
Polynomial	1764.69	0.474535	2175.22	0.4858999

2.1.2. Modelling For The United States Of America

In modelling for USA, the data was taken between the dates 1st March 2020 till 14th of September 2020 consisting of 197 data points. Among these data points 180 of them were used for training and 17 of them for testing.

In this case, we used support vector regression with radial kernel. During the regression, two different methods are used using the radial kernel. In the first case: using the number of daily confirmed cases in the United States of America as the independent parameter and predicting the number of daily death values in the United States of America accordingly was the method. In the second case, the input parameters were extended as the number of daily confirmed cases in the United States of America, the number of daily confirmed cases in the world and the number of daily death values in the world.

These were used in predicting the number of daily death values in the United States of America. Similar to previous analysis in the world, both the real data and all predicted data are plotted together in Figure 2. Additionally, the summary of regression evaluation parameters is tabulated in Table 2.

Table 2 Predicted Error in Numbers of Daily Death Values Using Two Different Methods for World vs. US Only

Dataset	All Data		Test Data	
	RMSE	R^2	RMSE	R^2
World	306.630	0.787649	263.7262	0.723376
US Only	349.397	0.724282	293.2111	0.517248

When the test data and all data are considered both in terms of R^2 value and RMSE, the method using the extended parameters (i.e. inclusion of world data) produced much better results. In this case using just the test data R^2 value turned out to be 0.723376 when the world data was included, which is higher compared to 0.517248 when the world data was not included. Similar observation could be made when the regression model is used with all the data. R^2 value was 0.787649 with a smaller RMSE 306.630 compared to 0.724282 R^2 value and 349.397 RMSE, when the extended parameters and only USA parameter were considered respectively.

2.1.3. Modelling For The United Kingdom

In modelling of the daily number of deaths for UK, the data was taken between the dates 10th March 2020 till 14th of September 2020 consisting of 176 data points. Among these data points, 160 of them were used for training and 16 of them for testing. Similar to USA modelling, we modelled the UK cases using support vector regression and radial kernel. The analysis was conducted with only UK data and also with UK data combined with the world data. To be more explicit, we used number of daily new confirmed cases in the world, number of daily deaths in the world and number of daily new confirmed cases in UK as input parameters and the output parameter was number of daily death values in the United Kingdom in one method, and used the number of daily observed cases in UK as only parameter to determine the number of daily deaths in UK in the other method. Similar to the previous analysis, both the real data and all predicted data are plotted together in Figure 3. Also, the summary of regression evaluation parameters is tabulated in Table 3.

Similar observations as with USA analysis can be observed. Using multiple input parameters produced a more accurate regression

modelling compared to the only input parameter of using number of daily confirmed cases in UK. Using test data R^2 value of 0.95412 and RMSE value of 82.94953 were observed, and when all the data is considered these values were 0.932362 and 79.3694 respectively. These results are much better than using the single parameter to UK, 0.871249 R^2 value with 108.1134 RMSE value when test data is considered and 0.82323 R^2 value with 128.310 RMSE value when all of the data is considered.

Table 3 Predicted Error in Numbers of Daily Death Values Using Two Different Methods for World vs. UK Only

Dataset	All Data		Test Data	
	RMSE	R^2	RMSE	R^2
World	79.3694	0.932362	82.94953	0.95412
UK Only	128.310	0.82323	108.1134	0.871249

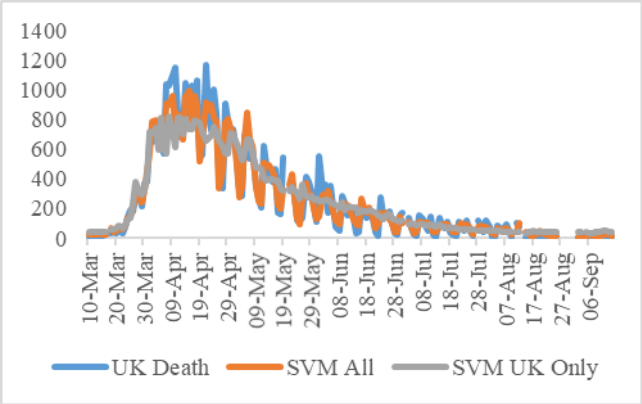


Figure 3 The number of daily deaths predicted for the United Kingdom (10th March -14th September 2020)

2.1.3. Modelling For Turkey

In modelling for Turkey, the data was taken between the dates 20th March 2020 till 14th of September 2020 consisting of 147 data points. There should have been more data during those dates, but since this data was not provided by the data center those dates were unavailable and not used. Among these data points 135 of them were used for training and 12 of them for testing. Similar to the other modellings, support vector regression with radial kernel is used as the basis and two different approaches, one by using only TR data and one by using TR data combined with world data is used. Both the real data and all predicted data using support vector regression are plotted together in Figure 4.

Also, the summary of regression evaluation parameters is tabulated in Table 4. In this case, some of the input parameters were not available from the source which we skipped in showing them in the graph and these appeared as blank since we wanted to show all the daily data. When the test data is considered two models using different parameters produced well modelling results: R^2 values were 0.875343 and 0.938179 for using extended parameters and only number of daily confirmed cases in Turkey respectively. It is important to note that using test data only Turkey only produced a more accurate model, this can be due to the fact that Turkey's data alone might not be very compatible with the whole world data since the number of death values were relatively lower compared with the other two countries. When using all data, the models for using only number of daily confirmed cases in Turkey only and model using the extended world parameters produced almost identical results, as 0.904195 compared to 0.904147. The reason for these high values of R^2 which indicated very good modelling can be due to the fact that the start date for the Covid-19 cases for Turkey was actually relatively later compared to the other two countries and mainly the values were very stable throughout the days. The other reason can be the smaller value in number of daily deaths in Turkey compared to the other two countries that were examined.

Table 4 Predicted Error in Numbers of Daily Death Values Using Two Different Methods for World vs. TR Only

Dataset	All Data		Test Data	
	RMSE	R^2	RMSE	R^2
World	15.1104	0.904147	12.60291	0.875343
TR Only	15.1066	0.904195	15.43535	0.938197

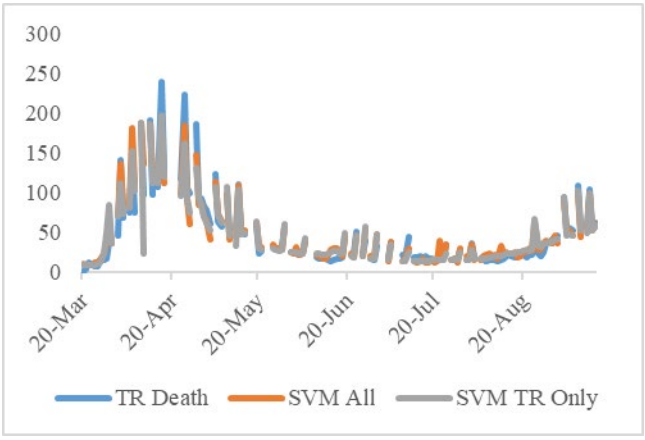


Figure 4 The number of daily deaths predicted for Turkey (20th March - 14th September 2020)

4.0 CONCLUSION

In this work we tried to create a model to predict the number of daily deaths in the world and also in three countries to make better predictions for future and so that the governments can take necessary precautions and do the necessary actions accordingly. In creating these models, linear, radial and polynomial kernels were used but it was noticed that the radial kernel produced better modelling. Also while doing analysis for different countries, considering data of the country together with the world data produced more accurate regression modelling compared to using data from only the country itself.

The models can be considered to be in good agreement although the data was limited to only seven months or less. Using different number of data for each country did not produce an obvious conclusion about the prediction results. Also, we can observe that although the country predictions are diminishing in an oscillatory way, the world data is still

not showing a clear end to the pandemic in terms of the number of daily deaths.

Acknowledgement

The authors are thankful to European University of Lefke due to the support provided for this work.

References

[1] Vespignani A., Tian H., Dye C., Lloyd-Smith J. O., Eggo R. M., Shrestha M., Scarpino S. V., Gutierrez B., Kraemer M. U. G., Wu J., Leung K. and Leung G. M. 2020. Modelling COVID-19. *Nature Reviews Physics, 2020.* 279-281. DOI: https://doi.org/10.1038/s42254-020-0178-4

[2] Arino J., and Portet S. 2020. Simple Model for COVID-19. *Infectious Disease Modelling.* 5: 309-315. DOI: https://doi.org/10.1016/j.idm.2020.04.002

[3] Abdo M. S., Shah K. Wahash H. A. and Panchal S. K. 2020. On a comprehensive model of the novel coronavirus (COVID-19) under Mittag-Leffler derivative. *Chaos, Solitons, Fractals.* 135,109867 2020. DOI: https://doi.org/10.1016/j.chaos.2020.109867

[4] Sajadi M. M., Habibzadeh P. Vintzileos A., Shokouhi S., Wilhelm F. M., and Amoroso A. 2020. Temperature, humidity, and latitude analysis t predict potential spread and seasonality for COVID-19. *SSRN 2020.* 3550308. DOI: https://doi.org/10.2139/ssrn.3550308

[5] Yang Z., Zeng Z., Wang K., Wong S., Liang W. Zanin M., … He J. 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease.* 12(3): 165-174. DOI: https://doi.org/10.21037/jtd.2020.02.64

[6] Fanelli D. and Piazza F. 2020. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons, Fractals.* 134, 109761, 2020. DOI: https://doi.org/10.1016/j.chaos.2020.109761

[7] Djilali S. and Ghanbari B. 2020. Coronavirus pandemic: A predictive analysis of the peak outbreak epidemic in South Africa, Turkey and Brazil. *Chaos, Solitons, Fractals.* 138, 109971, 2020. DOI: https://doi.org/10.1016/j.chaos.2020.109971

[8] Vapnik V. and Lerner A. 1963. Pattern recognition using generalized portrait method. *Automation and Remote Control*. 24: 774-780.

[9] Smola A. J., Scholkopf B. 2004. A Tutorial on Support Vector Regression. *Statistics and Computing.* 14: 199-222. DOI: https://doi.org/10.1023/B:STCO.0000035301.49549.88

[10] Basak D., Pal S. and Patranabis D. C. 2007. Support Vector Regression. *Neural Information Processing –Letters and Reviews.* 11(10): 203-224.

[11] Boughorbel S., Tarel J. P., Boujema N. 2005. Conditionally positive definite kernels for SVM based image recognition. *IEEE International Conference on Multimedia and Expo, July 2005*. DOI: https://doi.org/10.1109/ICME.2005.1521373

[12] Zhou D. X. and Jetter K. 2006. Approximation with polynomial kernels and SVM classifiers. *Advances in Computational Mathematics.* 25:323-344. DOI: https://doi.org/10.1007/s10444-004-7206-2

[13] Guo B. Gunn S. R. Damper R. I. and Nelson J.D.B. 2008. Customizing kernel functions for SVM-Based Hyperspectral Image Classification. *IEEE Transactions on Image Processing*. 17(4):622-629. DOI: https://doi.org/10.1109/TIP.2008.91895

[14] Ulker E.D. and Ulker S. 2019. Unemployment rate and GDP prediction using support vector regression. *Proceedings of the 1st International Conference on Advanced Information Science and System November 2019*. 17:1-5. DOI: https://doi.org/10.1145/3373477.3373494

[15] Zhong H., Wang J., Jia H., Mu Y., and LV S. 2019. Vector field-based support vector regression for building energy consumption prediction. *Applied Energy*. 242:403-414. DOI: https://doi.org/10.1016/j.apenergy.2019.2019.03.078

[16] Nourali H. and Osanloo M. 2018. Mining capital cost estimation using support vector regression (SVR). *Resource Policy*. 62:527-540. DOI: https://doi.org/10.1016/j.resourpol.2018.10.008

[17] Zhao D., Liu H., Zheng Y., He Y., Lu D. and Lyu C. 2019. A reliable method for colorectal cancer prediction based on feature selection and support vector machine. *Medical & Biological Engineering & Computing*. 57:901-912. DOI: https://doi.org/10.1007/s11517-018-1930-0

[18] Nguyen H. 2019. Support vector regression approach with different kernel functions for predicting blast-induced ground vibration: a case study in an open-pit coal mine of Vietnam. *SN Applied Sciences.* 1, 283 2019. DOI: https://doi.org/10.1007/s42452-019-0295-9

[19] Tan Z., Zhang J., He Y., Zhang Y., Xiong G. and Liu Y. 2020. Short-Term Load Forecasting Based on Integration of SVR and Stacking. *IEEE Access*. 8: 227719-227728, 2020. DOI: https://doi.org/10.1109/ACCESS.2020.3041779

[20] Liu Q., Liu W., Mei J., Si G., Xia T. and Quan J. 2021. A New Support Vector Regression Model for Equipment Health Diagnosis with Small Sample Data Missing and Its Application. *Shock and Vibration*. 6675078, 2021. DOI: https://doi.org/10.1155/2021/6675078

[21] Coronavirus statistics reports. 2020. http://www.virusncov.com [Accessed September 16, 2020].