



COVID-19 Variants and Transfer Learning for the Emerging Stringency Indices

Ayesha Sohail¹ · Zhenhua Yu² · Alessandro Nutini³

Accepted: 7 April 2022 / Published online: 10 May 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The pandemics in the history of world health organization have always left memorable hallmarks, on the health care systems and on the economy of highly effected areas. The ongoing pandemic is one of the most harmful pandemics and is threatening due to its transformation to more contiguous variants. Here in this manuscript, we will first outline the variants and then their impact on the associated health issues. The deep learning algorithms are useful in developing models, from a higher dimensional problem/ dataset, but these algorithms fail to provide insight during the training process and do not generalize the conditions. Transfer learning, a new subfield of machine learning has acquired fame due to its ability to exploit the information/learning gained from a previous process to improve generalization for the next. In short, transfer learning is the optimization of the stored knowledge. With the aid of transfer learning, we will show that the stringency index and cardiovascular death rates were the most important and appropriate predictors to develop the model for the forecasting of the COVID-19 death rates.

Keywords Artificial intelligence · Transfer learning · Stringency index · COVID-19 socioeconomic problems

1 Introduction

This era has remained the era of artificial intelligence and due to its success stories in almost all disciplines of sciences and technology, its demand has raised and besides that, people are more inclined to understand the blackbox nature of the machine learning tools.

✉ Alessandro Nutini
nutini@centrostudiattivitamotorie.it

Zhenhua Yu
zhenhua_yu@163.com

¹ Department of Mathematics, Comsats University Islamabad, Lahore Campus, Lahore, Pakistan

² Institute of Systems Security and Control, College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China

³ Centro Studi Attività Motorie - Biology and Biomechanics Department, Via di Tiglio 94, loc. Arancio, 55100 Lucca, Italy

The expaliable machine learning in this case has proved to be extremely helpful. It provides the researchers, from different disciplines, (with basic knowledge of AI tools) an insight of the tools and provides logic of the correct selection of the predictors and the response variables, suitable for their discipline, thus making the tools more transparent. The AI tool can then be presented as fair feature selecting algorithms, to identify potential bias/problems in the training data, and to ensure that the algorithms perform as expected.

Another evolving sub-field of data science is termed as the transfer learning field. It basically encompasses on the concept of learning knowledge from a given problem, that is rich in data, and followed by application to another problem interlinked with the given problem, and is not very rich in data due to limited sources/ limited experimental trials. Researchers have proposed that the transfer learning core is actually the domain adaptation, as a feature based approach [2, 4]. The transfer learning tools learns the transferable features across the given domains, in a “reproducing kernel Hilbert space (RKHS)”. With the advancement in this field, now the multiple kernels are also used to develop the transferable algorithm. The convolutional neural networks are designed on purpose, for the transfer learning approach, so that, the predefined networks can be utilized in a swift manner, for new problems. This step is also termed as the combined or hybrid approach of transfer learning with deep learning.

Coronavirus threat unlike other pandemics, proved to be extremely challenging for the human health, economy and progress worldwide. Several research articles have been published since the pandemic began [12, 15, 17]. The machine learning tools have been used in the literature to deal with the data bases linked with the human health [9–11]. The current research focuses on the factors linked with COVID-19 outbreak, the resulting responses and the relationship with different variants. The data based research is conducted and only one variant of coronavirus is discussed in detail, whereas the other variants and the associated data is carefully extracted and presented in this manuscript for the future research purposes.

To be more explicit, different variants of coronavirus exist in the world with different properties and characteristics since 1962 [1, 6, 14, 16, 17, 19–23]. According to the recent studies, the major variants that are extensively reported are: *the Alpha Variant (formerly called the UK Variant and officially known as B.1.1.7)*, β Variant (South Africa Variant/ B.1.351), γ Variant (Brazil Variant, officially known as P.1), and the Δ Variant (India Variant and officially known as B.1.617.2).

In Italy, at present, a pandemic wave caused by the Omicron variant is being faced, due to its very high transmissibility [7] but, despite the forecasts issued by IHME (IHME, 2022), the first studies conducted on this coronavirus variant confirmed a lower symptom severity (from 15 to 80%) compared to the delta variant [7, 13]. The scientific debate aimed at clarifying the effects of this new variant is still ongoing and, for this reason, in this article, we will refer to the variants already mapped and to the data available until 2021.

A mapping of the VOCs (variants of concerns) of the SARS-CoV2 virus in Italy has been carried out by the Istituto Superiore di Sanità since the beginning of the pandemic (<https://www.epicentro.iss.it/coronavirus/sars-cov-2-monitoraggio-varianti>). The variants analyzed are B.1.1.7 (alpha), B.1.351 (beta), P.1 (gamma) and B.1.617.2 (delta); in collaboration with the Bruno Kessler Foundation and the Italian Ministry of Health, sampling studies were carried out using RT-PCR tests.

The survey conducted examines samples obtained on 21 July 2021 and the sample size was calculated using the following equation:

$$n \geq \frac{N(z_{\alpha/2})^2 p(1-p)}{\epsilon^2(N+1) + (z_{\alpha/2})^2 p(1-p)} \quad (1)$$

Here the sample n necessary to observe the prevalence p of a variant on the national territory in a population of size N (population of notified positives) uses the estimate ϵ and a confidence level $“(1 - \alpha)\%”$. By dividing the entire national territory into four macro-areas: North-West, North-East, Center, South and Islands, and assuming we want to estimate a prevalence of 5% with 2% precision in these macro-areas, we obtain Table 1 which shows the values of the required sample size calculated on the basis of positive COVID-19 cases notified on 15 July 2021. The prevalence metrics are presented in Table 2.

From 1973 cases confirmed by RT-PCR, 1325 samples were sequenced and classified according to lineage (the details are provided in Table 3). In detail, among the 1309 sequences obtained for the analysis, the following were identified:

The prevalence estimates at national level, obtained as the average of the prevalences in the different Regions weighted by the number of regional cases notified on 20 July 2021, are as follows:

In this manuscript, the classification learners were first applied on the COVID-19 datasets to draw useful predictions, next, the predictions are verified, explained and interpreted in detail with the help of transfer learning algorithms. Initially, the classification learners divide the data into different classes, based on the definition (categorical) of the response variable.

Table 1 Macro-Area distribution and cases

Macro-area	Cases known as of July 15, 2021	Sample size ($p=5\%$; $\epsilon=2\%$)
North East	551	252
North West	480	237
Center	586	258
South and Islands	838	299
TOTAL	2455	1046

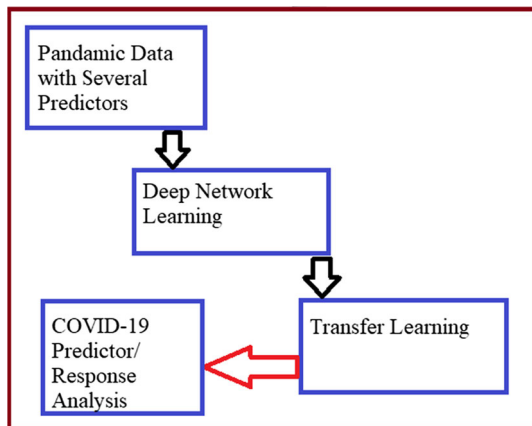
Table 2 Prevalence estimates at national level

Prevalence estimates	Linage
3.2% (range: 0.0–14.7%)	B.1.1.7
0.0% (range: 0.0–0.0%)	B.1.351
1.4% (range: 0.0–16.7%)	P.1
94.8% (range: 80–100%)	B.1.617.2
0% (range: 0.0–0.0%)	B.1.525
0% (range: 0.0–0.0%)	P.2

Table 3 List of sequences and the lineage

Sequences	Linage
49 attributable to SARS-CoV-2	B.1.1.7
0 traceable to SARS-CoV-2	B.1.351
16 traceable to SARS-CoV-2	P.1
1266 traceable to SARS-CoV-2	B.1.617.2
0 traceable to SARS-CoV-2	B.1.525
0 traceable to SARS-CoV-2	P.2

Fig. 1 COVID-19 data analysis flow chart with deep learning and transfer learning



Next, the transfer learning algorithms are applied on the pre-processed datasets. A schematic 1 is provided for better understanding of the research strategy.

2 Materials and Methods

In this manuscript, we have accessed two data sources, the first one provides information in depth, on the variants of the SARS-2 virus, reported in Italy since the virus initiated till date, and the second one is based on the linkage of the virus spread, the temperature and the stringency index, to the life expectancy, number of new hospital admissions and the new cases and new deaths reported.

2.1 SARS-2 Variants

Currently the data for Omicron, one of the variants of SARS-2, is not fully available, we have therefore conducted numerical experiments on other variants listed above. To explore the data based on the other variants, the following details were analyzed:

2.1.1 Variants of Concern (VOC)

- The lineage B.1.617.2 (Delta variant) has a prevalence of more than 94% and is reported in all Regions, replacing the de facto variant Alpha, now present only at 3.2% in this rapid survey compared to 57.8% of the previous survey;
- The P.1 lineage (Gamma variant) has a prevalence of 1.4% (in the previous it was 11.8%). In absolute numbers it appears to be decreasing in all Regions/PAs.

2.1.2 Variants of Interest (VOI)

- Lineage B.1.525 (Eta variant) was not reported in any Region / PA compared to the 10 cases of the previous survey;
- Lineage B.1.621 was reported in only one case in Veneto;
- The lineage P.3 (Theta variant) in a case in Sicily.

2.1.3 Variants Under Monitoring (VUM)

- The lineage P.2 (Zeta variant) has not been reported in any Region;
- Lineage C.36 + L452R (sub-lineage C.36.3) has been identified in 3 cases all in Lombardy;
- lineage B.1.1.318 in 1 case in Lombardy.

Researchers are working on the data of these variants and the details of these variants are available at the source repository (<https://www.acrobiosystems.com/>). Some useful results are drawn from these variants and are presented in the next section.

2.2 Transfer Learning and SVM

During this research, we have selected the predictors and response and have run the numerical experiments. The motivation is to understand which classifiers can classify well the response such as the stringency index during the pandemic, under the influence of the most relevant predictors. This analysis can help the governments, to forecast the level of control measures (or to be prepared to fight) against the pandemic in a better manner.

During this research, among the other classifiers, such as RUSboosted trees, random trees and gradient methods, we have applied the fine Gaussian support vector machine learning algorithm for the classification. This classifier works in determining a hyperplane, that will be used as a decision boundary. For this specific type of classifier, the hyperplane is affected by the support vectors and the outliers have very mild impact. Since the data was highly nonlinear, the nonlinear kernels were used. The Python scikit-learn package along with warm start parametric approach was used and the kernels used were Gaussian Radial Kernel Function (rbf) and sigmoid.

Since the COVID-19 data was unpredictable throughout the outbreak, the transfer learning is applicable since it is more appropriate when the domain or the task, i.e. the predictors and the response are expected to be changed. The research group led by Fan et al. [3] used the conventional softmax model for this purpose. In a similar manner, we used softmax, trained it with original pairs of features, label to obtain the label from the features.

While using the transfer learning, the data was prepared with transfer learning, it was split and the SVM nonlinear kernel was applied.

3 Results and Discussion

3.1 SARS-2 Variants Understanding

Based on the data key features presented in Sect. 2.1, we have derived the following conclusion:

- the diffusion of variants with greater “transmissibility” can have a significant impact. The Delta variant (B.1.167.2) is to be considered predominant in this quick survey, having replaced the Alpha variant for the first time (B.1.1.7). This data, however expected, is in line with what has been observed in other European countries. The Delta variant is, in fact, characterized by a “transmissibility” from 40 to 60% higher than the Alpha variant, and is associated with a relatively higher risk of infection in unvaccinated or partially vaccinated subjects;

- the prevalence of the Gamma variant (P.1) has drastically decreased throughout the country;
- in the current national scenario, characterized by the circulation of different variants of SARS-CoV-2, it is necessary to continue to monitor with great attention their diffusion and, in particular, of those with greater “transmissibility” or with mutations related to potential evasion of the immune response.

3.2 SARS-2 Spread Parameters

In the previous sections, we have highlighted the importance of different factors responsible, directly or indirectly, for the increase or decrease (positive or negative gradient) of COVID-19 cases reported over a period of fifty-five weeks from February 2020 till February 2021.

3.2.1 Impact of Temperature

The family of Corona Viruses is really sensitive to the sunlight. The virus dies rapidly when it is exposed to the ultraviolet light i.e. the sunlight is not favorable to the corona viruses. On the other hand, it has the ability to survive well at room temperature as it is termed as an enveloped virus. It survives well at low humid environment.

According to some research findings, Coronavirus spread is inversely proportional to the temperature. For higher temperature, it dies out or its transmission rate is reduced.

The researchers [14, 18] assessed the impact of the environmental stressors such as the impact of ozone, impact of nitrogen present in the environment, and other factors, on the spread of this enveloped virus. Another group of researchers [5] discussed the impact of pollution as well as the temperature, on the spread of this virus.

Recent studies have worked on the temperature range that is favorable for the spread of this virus. It is reported that the virus can not transmit rapidly, if the humidity is above “6 g/kg” and the average air temperature is above 51 F.

Response

We conducted to major numerical experiments. In first study, the response is taken to be as the temperature whereas, in the second study, response is kept as the stringency index. The results are presented in Tables 4 and 5 respectively.

Predictors

Three different groups of predictors, G_1 , G_2 and G_3 are used during this research. In group G_1 , the No of cases; vaccinations; No of deaths; ICU patients; Positive rate; No of hospital admissions; Population density were considered as the predictors.

Next, in group G_2 , the No of deaths; ICU patients are considered as the predictors. In group G_3 , Reproduction rate; Positive rate and total vaccination are taken as the predictors.

In group G_4 , Total tests and the Positive rate are taken to be the predictors relative to the stringency index, as the response variable.

In group G_5 , No of deaths and Weekly hospital admissions are selected as the predictors.

These combinations of predictors are selected after careful analysis of the correlations between the key variables. Another justification of this grouping is that the variables are grouped according to their direct or indirect linkage to the pandemic variants statistics.

Table 4 Groups of predictors for blackbox models with response as Temperature

Predictor (unit: per week)	Model	Accuracy (%)
G ₁	Fine Gaussian SVM	67.3
G ₁	Random Tree	61.8
G ₁	Ensemeled RUS boosted tree	63.6
G ₂	Fine Gaussian SVM	67.3
G ₂	Random Tree	63.6
G ₂	Ensemeled RUS boosted tree	63.6
G ₃	Fine Gaussian SVM	78.2
G ₃	Random Tree	78.2
G ₃	Ensemeled RUS boosted tree	76.4

Table 5 Groups of predictors for blackbox models with response as stringency index

Response	Predictor (unit: per week)	Model	Accuracy (%)
Stringency index	G ₄	Fine Gaussian SVM	69.1
–	G ₄	Random Tree	72.7
–	G ₄	Ensemeled RUS boosted tree	74.5
Stringency index	G ₅	Fine Gaussian SVM	67.3
–	G ₅	Random Tree	67.3
–	G ₅	Ensemeled RUS boosted tree	72.7

3.2.2 Stringency Index

The stringency index (S_I) is defined as the parameter that can gauge the level of strictness in making the policies that basically restrict individuals daily routine. The formula is based on the ordinal containment and closure policy indicators. According to the data presented by the source [8], the key metrics used for the evaluation of the Stringency Index were closure of the schools, offices, public places and cancellation of social events and gatherings, public transport (local and international travel). Its value for Italy spanned from 0 to 130. We have grouped (classified) the data into three categories, group A with $S_I < 50$; group B with $50 \leq S_I < 80$; and group C with $80 \leq S_I \leq 130$. From Fig. 2 we can see that the trend of the stringency index remained nonlinear, when the predictions were made (top panel) relative to the tests and positivity of the COVID-19 tests, similarly, for bottom panel, when the number of deaths were plotted relative to the hospitalization frequency, the nonlinear trend was observed. Here the red box indicated the domain, where there was saturation of the three classes. The cross represents the incorrect values and dot represents correct values achieved from the model.

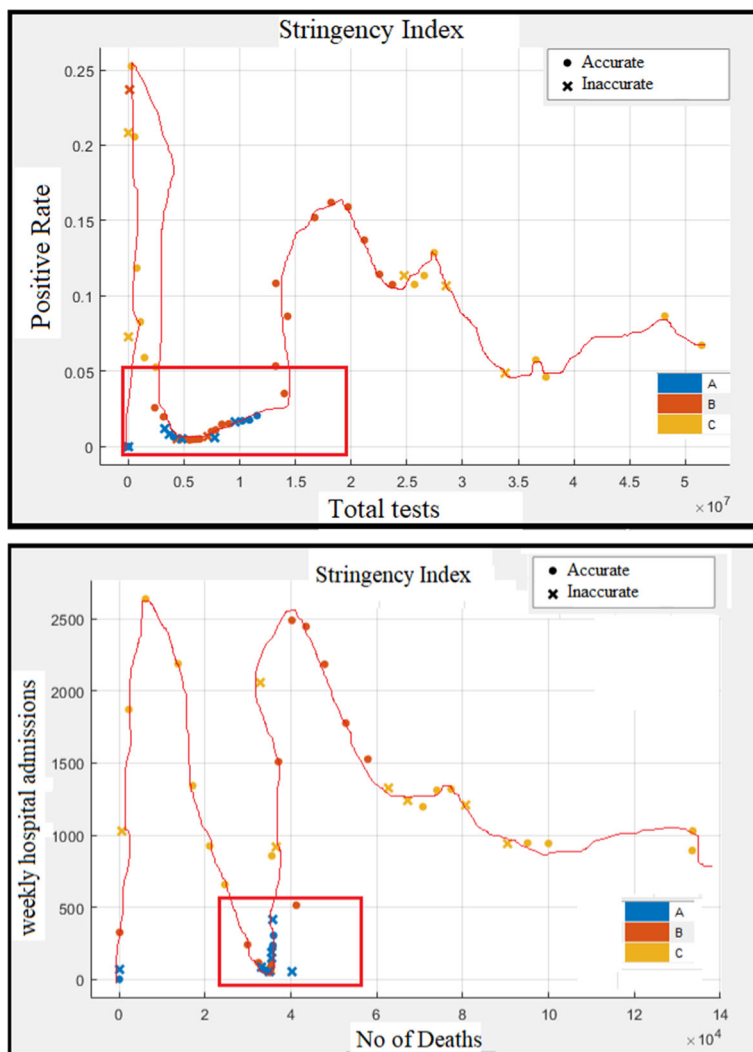


Fig. 2 Stringency index over period of 55 weeks Feb 2020 till Feb 2021. Group A with $S_I < 50$; group B with $50 \leq S_I < 80$; and group C with $80 \leq S_I \leq 130$

4 Conclusions

From this research, we conclude that lot more data repositories are desired to be established for the development of better precautionary measures. Since the impact of temperature on the spread of virus is of great significance according to the results obtained. Similarly, the stringency index and its impact on several predictors are discussed in detail in this manuscript. The data was based on the specific country of Italy and we hope that models can be developed for other countries as well, provided that detailed data repositories become freely available for the researchers to design, interpret and explain the machine learning models. With the aid of expalanabale artificial intelligence and its ability to deal with transfer learning tools,

we have provided the readers a state of the art framework, that can help to extract the most impulsive predictors, from a given nonlinear high dimensional data base. It is highly desired to obtain compact data, and to analyze it with the aid of transfer learning, for the variation in stringency index relative to different variants of COVID-19, since it is observed that the index varied in different regions of the world, relative to the α , β , γ , or Δ variants. For these reasons, the work we have presented can serve as a basis for the development of further algorithms predicting the effects and expansion of viral infection.

Acknowledgements The authors would like to acknowledge the source of data and data analysis repositories: (1) "<https://ourworldindata.org/covid-vaccinations?country=~ITA>", (2) <https://www.acrobiosystems.com/> and their cooperation, (3) <https://github.com/topics/transfer-learning?l=jupyter+notebook>, (4) <https://github.com/statmlben/Variant-SVM>, (5) <https://github.com/liquidSVM/liquidSVM>.

Declaration

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Al-Utaibi KA, Sohail A, Yu Z, Arif R, Nutini A, Abdel-Salam A-SG, Sait SM (2021) Dynamical analysis of the delayed immune response to cancer. *Results Phys* 26:104282
2. Banos O, Calatroni A, Damas M, Pomares H, Roggen D, Rojas I, Villalonga C (2021) Opportunistic activity recognition in iot sensor ecosystems via multimodal transfer learning. *Neural Process Lett* 53:1–29
3. Fan J, Lee J, Lee Y (2021) A transfer learning architecture based on a support vector machine for histopathology image classification. *Appl Sci* 11(14):6380
4. Lee Y, Chung M, Cho S, Choi J (2019) Extraction of product evaluation factors with a convolutional neural network and transfer learning. *Neural Process Lett* 50(1):149–164
5. Li H, Xu X-L, Dai D-W, Huang Z-Y, Ma Z, Guan Y-J (2020) Air pollution and temperature are associated with increased covid-19 incidence: a time series study. *Int J Infect Dis* 97:278–282
6. Malik YA (2020) Properties of coronavirus and sars-cov-2. *Malays J Pathol* 42(1):3–11
7. Pascarella S, Ciccozzi M, Bianchi M, Benvenuto D, Cauda R, Cassone A (2021) The electrostatic potential of the omicron variant spike is higher than in delta and delta-plus variants: A hint to higher transmissibility? *J Med Virol*
8. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B, Giattino C, Appel C, Rodes-Guirao L, Roser M (2020) Coronavirus pandemic (covid-19). *Our World in Data* <https://ourworldindata.org/coronavirus>
9. Sherin L, Sohail A, Mustafa M, Jabeen R, Ul-Hamid A et al (2020) Facile green synthesis of silver nanoparticles using terminalia bellerica kernel extract for catalytic reduction of anthropogenic water pollutants. *Colloid Interface Sci Commun* 37:100276
10. Sherin L, Sohail A, Shujaat S (2019) Time-dependent ai-modeling of the anticancer efficacy of synthesized gallic acid analogues. *Comput Biol Chem* 79:137–146
11. Sohail A (2019) Inference of biomedical data sets using bayesian machine learning. *Biomed Eng Appl Basis Commun* 31(04):1950030
12. Sohail A, Nutini A (2020) Forecasting the timeframe of 2019-ncov and human cells interaction with reverse engineering. *Prog Biophys Mol Biol* 155:29–35
13. Wolter N, Jassat W, Walaza S, Welch R, Moultrie H, Groome M, Amoako DG, Everatt J, Bhiman JN, Scheepers C et al (2022) Early assessment of the clinical severity of the sars-cov-2 omicron variant in south africa: a data linkage study. *The Lancet*
14. Yu Z, Abdel-Salam A-SG, Sohail A, Alam F (2021) Forecasting the impact of environmental stresses on the frequent waves of covid19. *Nonlinear Dyn* 106(2):1509–1523
15. Yu Z, Arif R, Fahmy MA, Sohail A (2021) Self organizing maps for the parametric analysis of covid-19 seirs delayed model. *Chaos Solitons Fractals* 150:111202
16. Yu Z, Ellahi R, Nutini A, Sohail A, Sait SM (2021) Modeling and simulations of covid-19 molecular mechanism induced by cytokines storm during sars-cov2 infection. *J Mol Liq* 327:114863

17. Yu Z, Sohail A, Nutini A, Arif R (2020) Delayed modeling approach to forecast the periodic behaviour of sars-2. *Front Mol Biosci* 7:386
18. Zoran MA, Savastru RS, Savastru DM, Tautan MN (2020) Assessing the relationship between ground levels of ozone (o3) and nitrogen dioxide (no2) with coronavirus (covid-19) in milan, italy. *Sci Total Environ* 740:140005
19. Yu Z, Gao H, Wang D et al (2022) SEI2RS malware propagation model considering two infection rates in cyber-physical systems. *Phys A: Stat Mech Appl* 597:127207
20. Yu Z, Sohail A, Nofal TA, Tavares JMRS (2021) Explainability of neural network clustering in interpreting the COVID-19 emergency data. *Fractals* 10:S0218348X22401223
21. Wang F, Sohail A, Wong WK, Azim QUA, Farwa S, Sajad M (2022) Artificial intelligence and stochastic optimization algorithms for the chaotic datasets. *Fractals*. <https://doi.org/10.1142/S0218348X22401752>
22. Nutini A, Zhang J, Sohail A, Arif R, Nofal TA (2022) Forecasting of the efficiency of monoclonal therapy in the treatment of CoViD-19 induced by the Omicron variant of SARS-CoV2. *Results in Physics* 35:105300
23. Sohail A, Arif R, Azim QUA, Nutini A, Gepreel KA (2022) Piecewise fractional order modeling of the breast cancer epidemiology after the Atezolizumab treatment. *Fractals*. <https://doi.org/10.1142/S0218348X22401764>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.