

Report of scientific research on the topic

Synthetic populations and multi-agent models of the spread of
acute respiratory infections

Performed by:

Carlos Andres Daza Rachen

Supervise by:

Vasiliy Leonenko

Saint-Petersburg

2021

Introduction	4
Related works	11
2.1 Machine Learning Models Quality and Interpretability for Chronic Heart Failure Patients[11]	11
2.2 Bayesian modeling of COVID-19 cases with a correction to account for under-reported cases[12]	12
2.3 Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt[14]	13
2.4 Machine learning-based predictors for COVID-19 disease severity[16]	14
2.5 A machine learning forecasting model for COVID-19 pandemic in India[18]	15
Bibliography	17

1. Introduction

Since the beginning of the development of living beings, including the awakening of humanity, we living beings have had to share our ecosystems with diverse species and we have generated diverse types of relationships with them and with our environment, which has caused that the flourishing of a group of individuals can affect the niche and its close neighbors and even the population in general.

This is true at all levels, including macroscopic, microscopic, and submicroscopic. The fact that we cannot see some of the smaller species does not imply that they are not there, that happens with viruses, that even being in the submicroscopic order which makes us not be able to see them, they are the most abundant form of life on the planet and are of great importance for the species, they are even considered a genetic vector that facilitates gene transfer between species and a trigger of evolution [1], however, viruses are also responsible for the origin and spread of many diseases, such as Ebola, AIDS, chickenpox, among others, as well as influenza and COVID-19 on which this article will focus.

In 2019 a new disease, COVID-19, caused by the SARS-CoV-2 virus, was announced to the world. On January 11, 2022, according to data from the World Health Organization, there have been 308,458,509 confirmed cases of COVID-19, including 5,492,595 deaths, on January 2020 the World Health Organization Emergency Committee declared a global health emergency based on growing case notification rates at Chinese and international locations [2].

Two years after the declaration of this emergency, humanity has not been able to recover from the damage to health systems, the economy, and in general in all areas of society. Most countries have declared different quarantine periods for their citizens, national and international health authorities have made great efforts to educate citizens on the importance of hygiene, the use of masks, and vaccination, the latter gaining great importance.

Table 1.

World Health Organization

Latest reported counts of cases and deaths (11.01.2022) [3]

	Name	Cases - total	Cases - last 7 days	Cases - last 24 hours	Deaths - total	Deaths - last 7 days	Deaths - last 24 hours
1	Global	308458509	16907207	1738701	5492595	45035	5150
2	United States of America	59848136	4697353	354163	831548	11862	393
3	India	35875790	915529	168063	484213	2196	277
4	Brazil	22523907	230679	24382	619981	848	44
5	The United Kingdom	14617318	1195432	142122	150230	1329	76
6	France	11907500	1854631	88607	123104	1486	279
7	Russian Federation	10684204	113992	17525	317687	5500	783
8	Turkey	10042797	446018	65236	83843	1048	141
9	Germany	7581381	342973	45690	114351	1772	322
10	Italy	7554344	1158234	117405	139265	1479	227

Vaccination is the most powerful weapon humanity has against the COVID-19 pandemic to January 11 there are 137 vaccines in clinical development and 194 in preclinical development [4], the speed at which scientific society has reacted and developed as many vaccines is encouraging, but at the same time, the virus in his fast rate of transmission has developed many variants.

Omicron is the last relevant variant, was initially reported to the World Health Organization (WHO) on November 24, 2021, by South Africa and it is a great challenge to scientists and existing vaccines, spreading around 70 times faster in the bronchi, and resistant to two doses of vaccination [5].

Table 2.

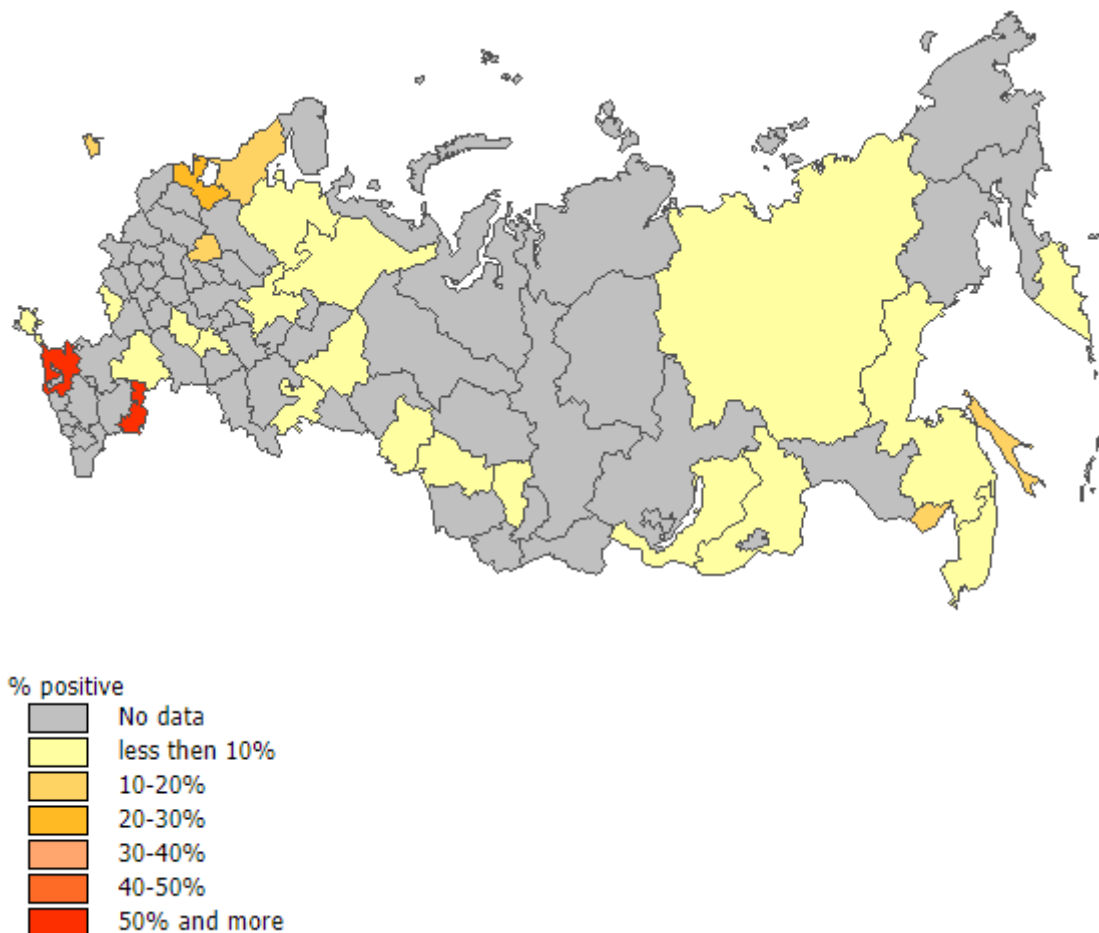
World Health Organization, Summary of vaccine performance against variants of concern (data as of 8 January 2022) [6]

	WHO Emergency Use Listing (EUL) Qualified Vaccines*							Vaccines without WHO EUL*		
	AstraZeneca-Vaxzevria/SII - Covishield	Beijing CNBG-BBIBP-CorV	Bharat-Covaxi	Janssen-Ad26.COV 2.S	Moderna-mRNA-1273	Pfizer BioNTech-Comirnaty	Sinovac-CoronaVac	Anhui ZL-Recombinant	Gamaleya-Sputnik V	Novavax-Covavax
Alpha^{39,40}										
Summary of VE*	Protection retained against all outcomes									
- Severe disease	↔ ₂	-	-	-	↔ ₂	↔ ₆	-	-	-	-
- Symptomatic disease	↔ to ↓ ₅	-	-	-	↔ ₁	↔ ₄	-	-	-	↓ ₁
- Infection	↔ to ↓ ₄	-	-	-	↔ ₃	↔ ₃	-	-	-	-
Neutralization	↔ to ↓ ₉	↔ ₁	↔ ₂	↔ ₅	↔ to ↓ ₁₅	↔ to ↓ ₄₈	↔ to ↓↓ ₈	↔ ₂	↔ to ↓ ₃	↓ ₂
Beta⁴¹⁻⁴⁴										
Summary of VE*	Protection retained against severe disease; reduced protection against symptomatic disease; limited evidence									
- Severe disease	-	-	-	↔ ₁	↔ ₁	↔ ₃	-	-	-	-
- Symptomatic disease	↔ to ↓↓↓ ₂	-	-	↔ ₁	↔ ₁	↔ ₂	-	-	-	↓↓↓ ₁
- Infection	-	-	-	-	↔ ₁	↓ ₁	-	-	-	-
Neutralization	↓ to ↓↓ ₁₁	↓ ₃	↓ ₂	↓ to ↓↓ ₉	↓ to ↓↓ ₂₅	↓ to ↓↓ ₅₇	↓ to ↓↓↓ ₇	↔ to ↓ ₃	↓ to ↓↓ ₄	↓↓↓ to ↓↓↓↓ ₂
Gamma										
Summary of VE*	Unclear impact; very limited evidence									
- Severe disease	↔ ₁	-	-	-	↔ ₁	↔ ₂	-	-	-	-
- Symptomatic disease	↔ ₁	-	-	-	↔ ₁	↔ ₁	-	-	-	-
- Infection	↔ ₁	-	-	-	↔ ₁	↔ ₁	↔ ₁	-	-	-
Neutralization	↔ to ↓ ₄	-	-	↔ to ↓ ₅	↓ ₂₀	↔ to ↓ ₂₈	↓ ₅	↔ ₁	↓ ₁	↓ ₁
Delta⁴⁵										
Summary of VE*	Protection retained against severe disease; possible reduced protection against symptomatic disease and infection									
- Severe disease	↔ ₃	-	-	↓ ₁	↔ ₃	↔ ₆	-	-	-	-
- Symptomatic disease	↔ to ↓↓ ₆	-	↓ ₁	-	↔ ₂	↔ to ↓ ₅	-	-	-	-
- Infection	↔ to ↓ ₄	-	-	↓↓↓ ₁	↔ ₃	↔ to ↓ ₃	-	-	-	-
Neutralization	↓ ₁₃	↓ ₂	↔ to ↓ ₃	↔ to ↓↓ ₉	↔ to ↓ ₃₄	↔ to ↓ ₃₉	↓ to ↓↓↓ ₈	↔ to ↓ ₂	↓ to ↓↓ ₃	↓ ₁
Omicron										
Summary of VE*	Reduced protection against infection and symptomatic disease; possible reduced protection against severe disease; limited evidence									
- Severe disease	-	-	-	-	-	↓↓↓/↓↓↓ ₁	-	-	-	-
- Symptomatic disease	↓↓↓ ₁	-	-	-	↓↓↓ ₁	↓↓↓ ₁	-	-	-	-
- Infection	-	-	-	-	↓↓↓ ₁	↓↓↓ ₂	-	-	-	-
Neutralization	↓↓↓ ₃	-	-	↓↓↓ ₁	↓↓↓ ₁₂	↓↓↓ ₂₀	↓ ₁	-	↓↓↓ ₁	↓↓↓ ₁

Like the rest of humanity, Russia has not been exempt from this problem, there are studies [8] that show that the climate of the area shared by Russia, Ukraine, and Belarus favors the spread of the virus, the government has made great efforts to slow the advance of the pandemic but as in the world in general, the speed of its advance has not been able to be slowed by government entities. Climatic factors such as temperature, relative humidity, seasonal temperature, sunlight, wind speed take on the characteristics of Russia's varied geography, which in turn imprint hemispheric characteristics and make them suitable for the spread of the virus[9].

According to World Health Organization (WHO) in Russian Federation, from 3 January 2020 to 12 January 2022, there have been 10.702.150 confirmed cases of COVID-19 with 318.432 deaths, with 17.946 new cases in the last 24 hours [7].

Figure 1.
Results of PCR detections of SARS-CoV-2 in Russia [10]



Although the number of cases ending in death and clinical occupancy is not as abundant, influenza remains an important public health factor, and is currently an additional vector for the transmission of COVID-19, its presence in Russia continues to be monitored by the National Influenza Center in Saint Petersburg, and even more so with the direct relationship to the current pandemic.

Figure 2.

Geographic distribution of RT-PCR detected influenza viruses in cities under surveillance in Russia, week 52 of 2021[10]



PCR results

	No data
	No viruses detected
	H1pdm09
	H3
	H3+H1pdm09
	B
	B+H1pdm09
	B+H3
	B+H3+H1pdm09
	A (not subt.)
	A (not subt.)+H1pdm09
	A (not subt.)+H3
	A (not subt.)+H3+H1pdm09
	A (not subt.)+B
	A (not subt.)+B+H1pdm09
	A (not subt.)+B+H3
	A (not subt.)+B+H3+H1pdm09

Figure 3.

Influenza and ARVI morbidity per 10.000 population in 61 cities under surveillance in Russia, seasons 2020/21 and 2021/22[10]

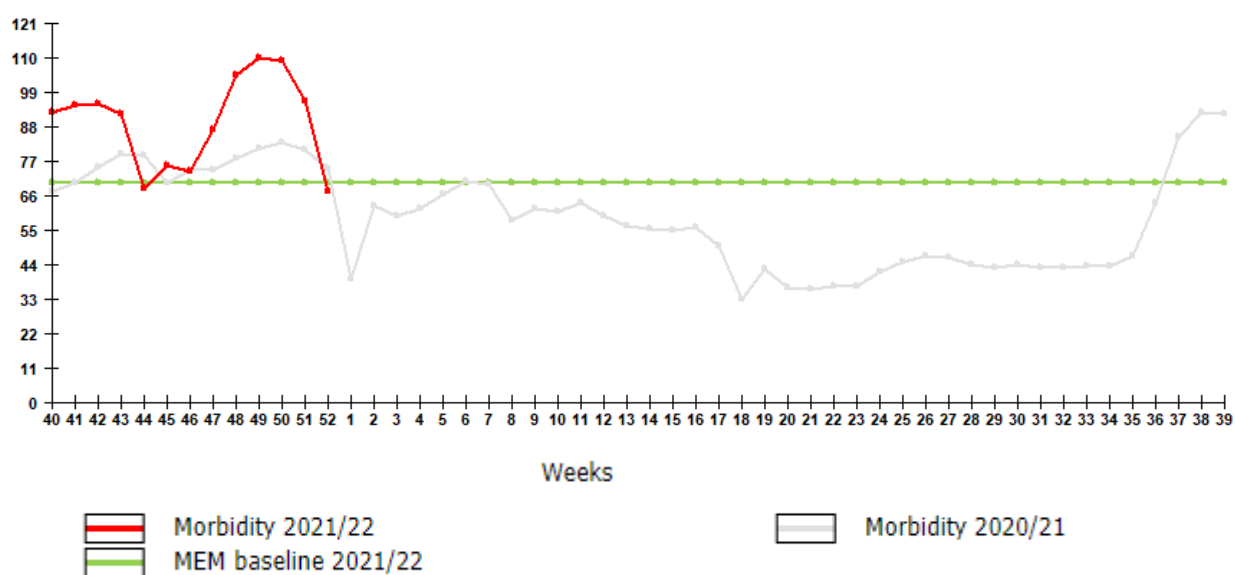


Table 3.

Results of influenza and other ARVI detection by RT-PCR in Russia, week 52 of 2021

	Number of specimens/number of positive cases	% positive
Influenza		
Number of specimens tested for influenza	4621	-
Influenza A (not subt.)	146	3,2%
Influenza A(H1)pdm09	0	0,0%
Influenza A(H3)	820	17,7%
Influenza B	4	0,09%
All influenza	970	21,0%
Other ARVI		
Number of specimens tested for ARVI	4316	-
PIV	29	0,7%
ADV	37	0,9%

RSV	86	2,0%
RhV	98	2,3%
CoV	43	1,0%
MPV	3	0,07%
BoV	10	0,2%
All ARVI	306	7,1%
SARS-CoV-2 (COVID-19)		
Number of specimens tested for SARS-CoV-2	14288	-
SARS-CoV-2	2981	20,9%

With the above, we can see how Russia and the world are facing the difficult task of curbing the spread of the virus and face the challenges it poses to their health systems, the different government entities, and its effects on the economy, among others. And as the efforts of different scientists has resulted in the development of various vaccines, protocols, and many study material, as professionals and members of the academy, it is our duty to ask ourselves the question of how from our area we can contribute to the victory over this great challenge, more precisely, how, through the use of big data and machine learning, we can generate knowledge that helps to understand and counteract this pandemic.

To provide a solution to the above, we propose the development of a model to demonstrate applications of AI and machine learning in understanding and forecasting the spread of acute respiratory infections, influence, and COVID in St. Petersburg, by now there are studies for the entire country but not specifically for St. Petersburg.

The great extension of Russia makes that there are great geographic changes between its different cities and therefore its environmental factors vary, which makes it difficult to apply a disease prediction model to the whole country without generating errors, therefore it is necessary to make a specific approach for St. Petersburg given its importance as one of the main cities, and therefore, one of the most populated, and affected by the different respiratory diseases, and currently by the pandemic of COVID-19, this is a great contribution to the city, the country, and society in general.

2. Related works

Since the emergence of more powerful computers has allowed us to make greater use of the different techniques of artificial intelligence with neural networks leading them, developers and scientists have wanted to make use of them for analysis and projections, this is how in the field of medicine have generated much research that makes use of artificial intelligence and neural networks, however, the fact that machine learning models require a high level of detail to ensure their reliability and that is an area still under development, generates distrust in their results by many health professionals, This does not stop scientists and developers from continuing their efforts in the creation of new and improved models for the analysis and prediction of results. Below are cases of application of artificial intelligence in the field of medicine with special emphasis on the analysis and prediction of the COVID-19 pandemic, the main topic of this article.

2.1 Machine Learning Models Quality and Interpretability for Chronic Heart Failure Patients[11]

The aim of this paper was to determine whether historical data or the latest data is more relevant and efficient to analyze a disease in a particular moment of time, for that 3 machine learning algorithms were trained: GBoost, logistic regression, and random forest for classification and linear regression, decision tree and XGBoost for the regression task.

The models were interpreted using “SHAP (Shapley Additive exPlanations) algorithm, as classification models, they used XGBoostClassifier, Logistic Regression, Decision Tree, and Random Forest. And as a quality metric, they used F1-score with macro averaging, since classes were imbalanced”.

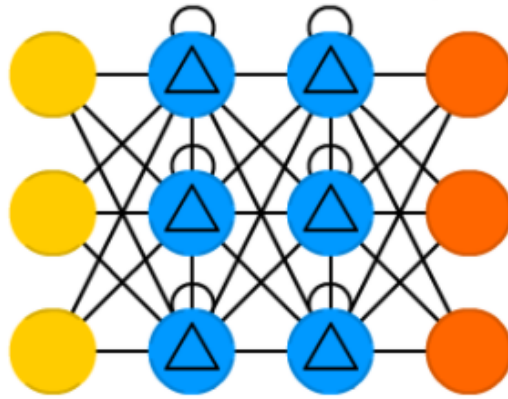
As conclusions of the article, we have that, “in binary and multi-classification tasks, where target variables do not change at a high rate, ML models can boost the quality of performance using dynamic features from the past. However, when it comes to regression tasks with highly volatile targets, the latest static information might be more valuable”

The algorithm that had the best quality results was XGB, which related to neural networks, GRU (RNN with gated recurrent units), demonstrate to be the best in medical predictions, “due to long temporal dependencies and missing time series values efficient treatment”, but it requires a lot of data (amount in hundreds to thousands of records)to

be well trained, for the investigators deep neural networks are difficult to use due they can be hardly interpretable.

Figure 4.
Neuronal Networks, Gated Recurrent unit[15]

Gated Recurrent Unit (GRU)

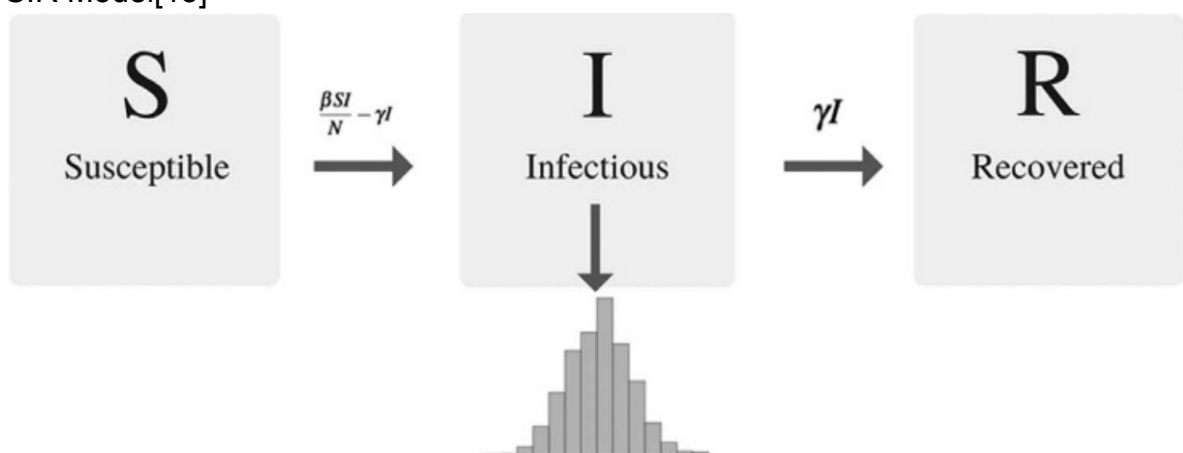


2.2 Bayesian modeling of COVID-19 cases with a correction to account for under-reported cases[12]

In this article the investigators tried to make a Bayesian approach to the SIR model, a compartmental model proposed by Kermack and McKendrick in 1927 in which population is assigned to compartments with labels:

- **S:** Susceptible
- **I:** Infectious
- **R:** Recovered

Figure 5.
SIR Model[13]



This with the aim to counter the lack of data due to unreported cases.

Two important factors were determined in the research were the case fatality ratio and the potential for under-reporting which were calculated this way:

$$CFR_c = m / \mu$$

Where m is the cumulative number of deaths and μ the number of notified cases

They applied Bayesian statistics with Markov Chain Monte Carlo to make integrations over the unreported data of Brazil on February 2020.

The article concluded with the affirmation that the method of adjusting cases by delay was successful and can be applied to forecasting COVID-19 spread, dealing with the lack of data due to unreported cases.

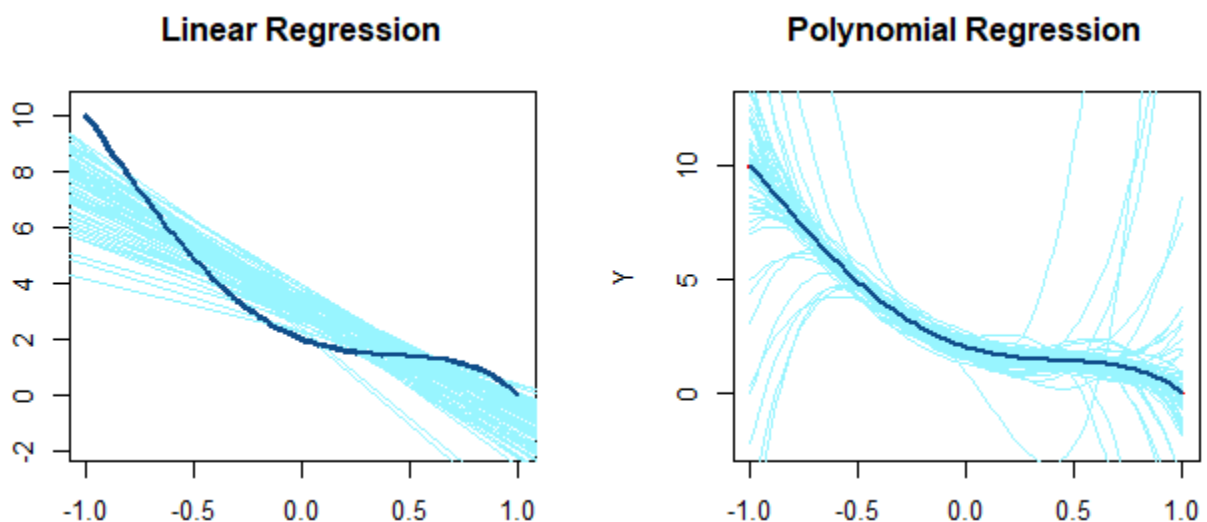
2.3 Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt[14]

The aim of this investigation was to forecast the number of patients that will be infected with COVID-19 in Egypt, the data analyzed corresponded to the period from February 15 of 2020 to June 15 of the same year, seven regression analysis-based models were applied:

- Exponential polynomial,
- Quadratic
- Third-degree
- Fourth-degree (The most effective)
- Fifth-degree
- Sixth-degree
- Logit growth

Figure 6.

Linear - Polynomial Regression



According to the conclusions of the paper, the forecast of a final number of cases is 1.667.600.000, at 11.01.2022 the number of cases according to the WHO(World Health Organization) is 308.458.509.

The final date of the epidemic was predicted to 8-September 2020, this forecast was previous to the appearance of the new variants, so the expected end date for the pandemic was wrong, and a little more than a year we find ourselves facing new variants, and again with the challenge of estimating a date for the end of this pandemic.

2.4 Machine learning-based predictors for COVID-19 disease severity[16]

In this paper the authors exposed the way they worked with population data to estimate the need for intensive care and mechanical ventilation on COVID-19 patients, they split the dataset into two parts, 60% training, and testing sets of 20%. To train the supervised learning models they used Fivefold cross-validation. The relative importance of the input features was evaluated by computing their Gini importance.

The following algorithms were studied:

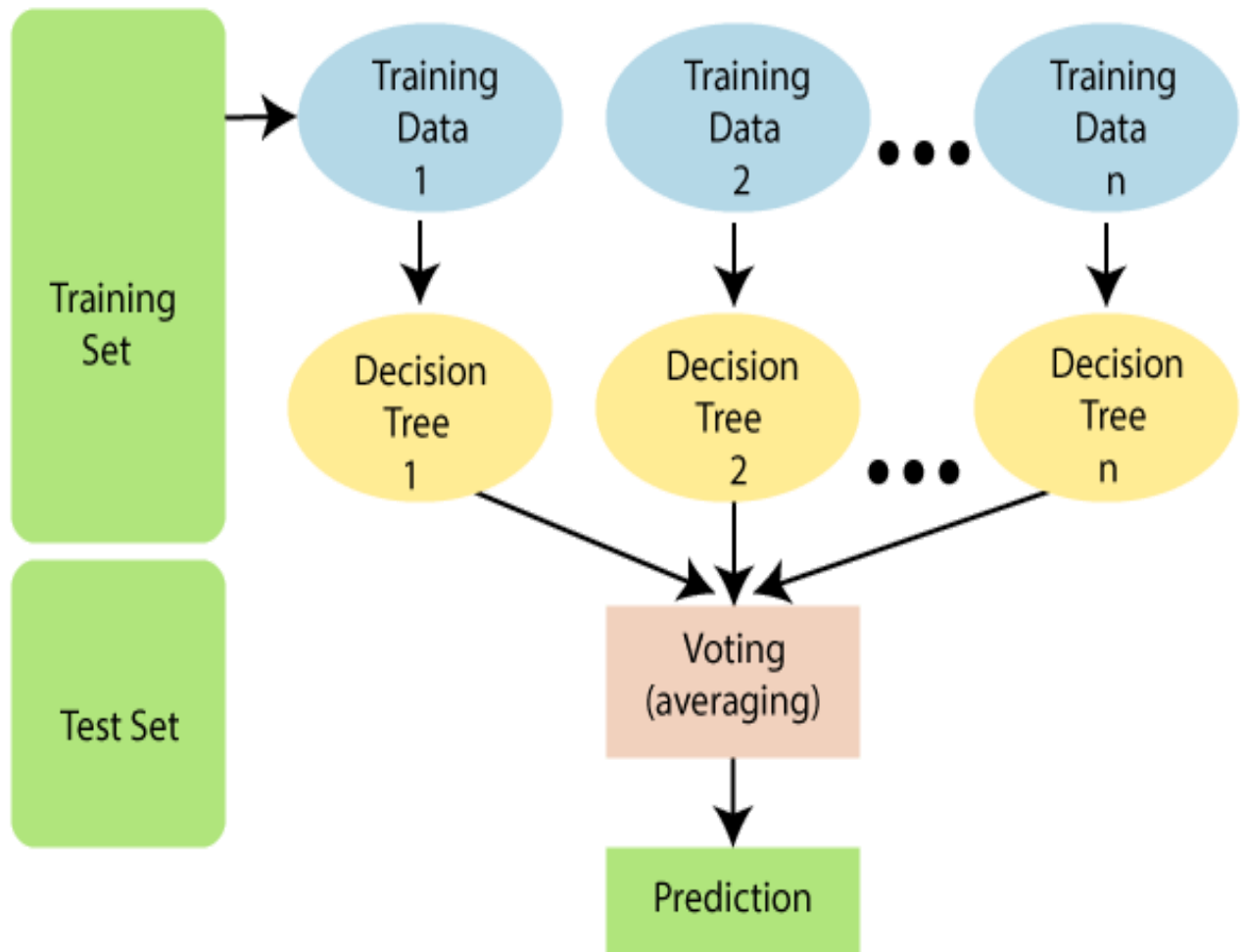
- Random forest
- Multilayer perceptron
- Support vector machines
- Gradient boosting
- Extra tree classifier
- Adaboost

The best performance and accuracy were obtained with:

- Random forest
- Extra tree classifier
- Gradient boosting

With Random forest being the most accurate, the fault in Support vector machines algorithm could be probably it doesn't have the capacity to capture the complexity in the prediction. Multilayer perceptron didn't work as expected due that it requires a large amount of data for efficient training.

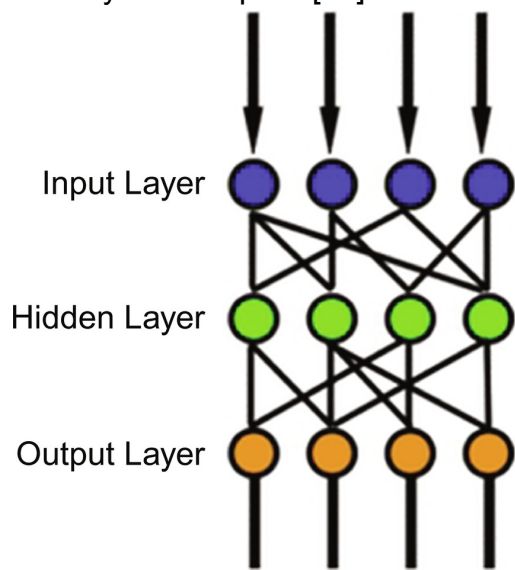
Figure 7.
Random Forest algorithm[17]



2.5 A machine learning forecasting model for COVID-19 pandemic in India[18]

This paper is from 30 May of 2020, one of the first in India in trying to make a forecast of the magnitude of the pandemic, it applied an attempt at prediction through the implementation of Multilayer perceptron, also applied Pearson and Spearman to measure the correlation between the input factors of the model. The paper concluded by giving a forecasted rate of recuperation of 95% and indicating that it is necessary for more data to give accurate forecasting numbers for a number of cases.

Figure 8.
Multilayer Perceptron[19]



Bibliography

1. [Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L., & Brüssow, H. \(2003\). Phage as agents of lateral gene transfer. Current Opinion in Microbiology, 6\(4\), 417–424.
https://doi.org/10.1016/S1369-5274\(03\)00086-9](#)
2. [Velavan, T. P., & Meyer, C. G. \(2020\). The COVID-19 epidemic.
https://doi.org/10.1111/tmi.13383](#)
3. [World Health Organization, Latest reported counts of cases and deaths \(11.01.2021\)\)](#)
4. [WHO COVID-19 - Landscape of novel coronavirus candidate vaccine development worldwide 12.01.2022](#)
5. [Chowdhury, S., Bappy, M. H., Chowdhury, S., Chowdhury, M. S., & Chowdhury, N. S. \(2022\). Omicron Variant \(B.1.1.529\) of SARS-CoV-2, A Worldwide Public Health Emergency!. European Journal of Clinical Medicine, 3\(1\), 5-9.
https://doi.org/10.24018/clinicmed.2022.3.1.159](#)
6. [WHO, Weekly epidemiological update on COVID-19 - 11, Edition 74, January 2022](#)
7. [World Health Organization \(WHO\). Russian Federation Pandemic Overview, \(12.01.2022\)](#)
8. [Anders Åslund \(2020\) Responses to the COVID-19 crisis in Russia, Ukraine, and Belarus, Eurasian Geography and Economics, 61:4-5, 532-545, DOI: 10.1080/15387216.2020.1778499](#)
9. [Malay Pramanik, Parmeshwar Udmale, Prafullit Bisht, Koushik Chowdhury, Sylvia Szabo & Indrajit Pal \(2020\): Climatic factors influence the spread of COVID-19 in Russia, International Journal of Environmental Health Research, DOI: 10.1080/09603123.2020.1793921](#)
10. [Ministry Of Health Of The Russian Federation, NATIONAL WEEKLY INFLUENZA BULLETIN OF THE RUSSIAN FEDERATION](#)
11. [Balabaeva, K., & Kovalchuk, S. \(2019\). Comparison of Temporal and Non-Temporal Features Effect on Machine Learning Models Quality and Interpretability for Chronic Heart Failure Patients. Procedia Computer Science, 156, 87–96.](#)
12. [de Oliveira, A. C. S., Morita, L. H. M., da Silva, E. B., Zardo, L. A. R., Fontes, C. J. F., & Granzotto, D. C. T. \(2020\). Bayesian modeling of COVID-19 cases with a correction to account for under-reported cases. Infectious Disease Modelling, 5, 699–713.
https://doi.org/10.1016/J.IDM.2020.09.005](#)

13. [Daughton, A. R., Generous, N., Priedhorsky, R., & Deshpande, A. \(2017\). An approach to and web-based tool for infectious disease outbreak intervention analysis OPEN. Nature Publishing Group.](#)
14. [Amar, L. A., Taha, A. A., & Mohamed, M. Y. \(2020\). Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt. Infectious Disease Modelling, 5, 622–634.](#)
15. [<https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464>. \(11.01.2022\)](#)
16. [Patel, D., Kher, V., Desai, B., Lei, X., Cen, S., Nanda, N., Gholamrezanezhad, A., Duddalwar, V., Varghese, B., & Oberai, A. A. \(123 C.E.\). Machine learning based predictors for COVID-19 disease severity. Scientific Reports |, 11, 4673. <https://doi.org/10.1038/s41598-021-83967-7>](#)
17. [Random Forest Algorithm. <https://www.javatpoint.com/machine-learning-random-forest-algorithm>](#)
18. [Sujath, R., Jyotir, •, Chatterjee, M., Aboul, •, & Hassanien, E. \(n.d.\). A machine learning forecasting model for COVID-19 pandemic in India. <https://doi.org/10.1007/s00477-020-01827-8>](#)
19. [Abirami, S., & Chitra, P. \(2020\). Energy-efficient edge based real-time healthcare support system. Advances in Computers, 117\(1\), 339–368. <https://doi.org/10.1016/BS.ADCOM.2019.09.007>](#)