Deep Learning Based Prediction Of Covid-19 Daily New Hospitalization In France and Multiple Vaccines Strategies Simulation

Jing ZHAI
Université Toulouse 1 Capitole
2 Rue du Doyen-Gabriel-Marty
Toulouse,France
jing.zhai@ut-capitole.fr

Wenhan Yang
Université Toulouse 1 Capitole
2 Rue du Doyen-Gabriel-Marty
Toulouse,France
wenhan.yang@ut-capitole.fr

Amrutha Nair THALAPPAN
Université Toulouse 1 Capitole
2 Rue du Doyen-Gabriel-Marty
Toulouse,France
amrutha-nair.thalappan@ut-capitole.fr

Elijah ADEOGUN
Université Toulouse 1 Capitole
2 Rue du Doyen-Gabriel-Marty
Toulouse,France
elijah.adeogun@ut-capitole.fr

Abstract

The ongoing outbreak of the COVID-19 pandemic imposes human lives and economy at risk. Due to the increased enormity of the number of COVID-19 cases, the role of Artificial Intelligence (AI) is imperative in the current scenario. AI would be a powerful tool to fight against this pandemic outbreak by predicting the new daily hospitalization number in advance. In this paper,Long Short-Term Memory (LSTM) and Convolutional Neural Network(CNN) on time series based models are used for predicting the new daily hospitalization number both en period before the existence of vaccination and during vaccination process, more particularly to analyze and evaluate the efficiency of the vaccination policy in France.Initially, the data pre-processing and feature extraction is made with the official COVID-19 dataset published by French government. Moreover, as there are currently quite a range of different vaccines which varies in efficacy, it is important to implement agent-based modelling and simulation to stimulate the different vaccines in order to figure out the most efficient on the population with certain indicators under different parameters.

Key words

Deep learning, Convolutional Neural Network(CNN), Long Short-Term Memory (LSTM), Time series, Pandemic, COVID-19, Coronavirus, Multi Agent Simulation, Netlogo, Vaccination.

1 Introduction

1.1 Context

The World has been affected by a highly contagious virus called the Corona virus or SARS-COV-2. This virus originated in the wet markets of Wuhan, Hubei province of China during December 2019. This virus quickly spread to more than 160+ countries within a span of 3 months causing over 400,000 deaths with more than 8.9 million people affected globally.This virus has caused very distressing times across all the countries and significant disruptions in global economies. Several intervening measures have been taken by the affected countries such as quarantining people to stop the spread of the virus.

As a third wave of spread Covid-19 arrived, France has moved to the next step in the fight against the coronavirus epidemic: the vaccine campaign. It even began earlier than expected: from December 27, 2020 instead of January 4, 2021 as initially planned. The vaccination campaign takes place in three steps depending on the priority populations.

On December 27, 2020, France initiated the first phase of the vaccine campaign. In compliance with the recommendations from the Haute Autorité de Santé, it will first target a part of the population at risk: elderly people living in nursing homes, as well as the staff likely to develop a severe form of the disease.

Once the first step has been complete and as deliveries arrive, the vaccination scheme will be extended to vulnerable people. After people aged 75+, health workforce without age restrictions can be vaccinated, followed by people aged 50-65: "In compliance with the Haute Autorité de Santé's guidance, this vaccine will be used first to vaccinate all health workers, however old they are, then people of 50 to 65 years of age with comorbidities. And then, all people aged 50 to 65" Castex explained on Thursday February 4 during a press brief. He went on stating that "all people aged 50-65" will be vaccinated then, during this very step.

The last and third phase is set to begin in the spring 2021. This time, the access to vaccine will be extended to people aged 50-64, as well as security, education professionals, people vulnerable and in precarious situations, as people living in closed facilities. Then, it will be made available to the entire population.

To analyze the effect of vaccination on daily new hospitalization, data is dynamic in nature as the cases might differ based on the seasons, populations and more particularly the phase of French policy. Thus to study the vaccination effect for first phase in France, deep learning models based on Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) on time series could be used to predict the data accurately. While these tools are great in examining observations and reaching to conclusions, they come with some serious limitations.[2] In most cases the data is skewed and relativistic. Considering this a robust new method using deep learning models are inevitable to gain time series forecasting results with higher accuracy.

To effectively create a model and stimulation that compares the overall efficiency of different vaccines over another including vaccines that are promising in terms of efficacy not yet available, we used the latest vaccines information from [4] such as the availability date of the vaccine, the efficiency rate and also the price. These information are also used as adjustable comparative parameters in our model and in the scenario used to stimulate the model based agents.

1.2 Objective of the work

The objective of this study is to perform a comparative analysis on the prediction of new daily hospitalization number both in the period before the existence of vaccination and during vaccination process to evaluate the efficiency of the vaccination in France, by using different prediction models based on Convolutional Neural Network(CNN) or Long Short-Term Memory (LSTM) on time series. The second objective is the agent-based modelling and and simulation of the already available vaccines in comparison to other fourth coming vaccines according to their efficacy and other parameters which is done on the Netlogo platform and make a decision that we need to start using the existing vaccine or wait for next more efficinet vaccine.

2 Material and Method

2.1 Deep Learning based prediction of new hospitalization

2.1.1 Data inspection and preparation

The data is used in the models is open data source from French government. It contains the data from 19/03/2020 to 18/02/2021 [Figuer 1]. There are many columns in terms of different usage. In this paper, only the columns with the number of new patient hospitalised and the number of new people vaccinated are taking into account.

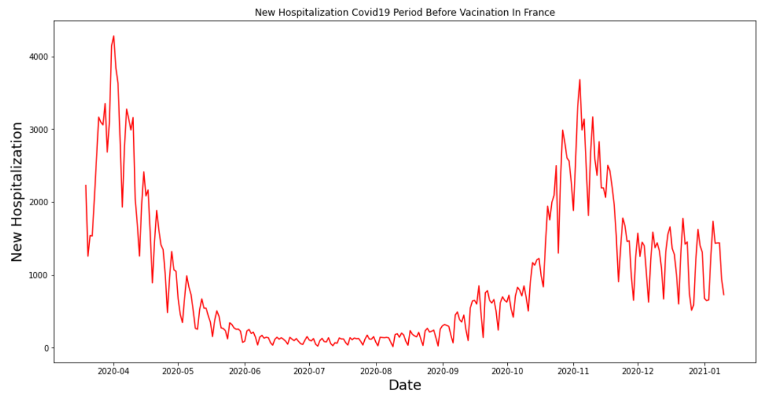

Figure 1: Dataset View,actual new hospitalization number

The first vaccination date in France happened in 27/12/2020. In order to represent the vaccination number before that date, the values are filled with 0. The data is split to two parts in order to train two different models and compare the results. First part until 27/12/2020 is the data without vaccination. Second part start from 10/01/2021 because there are two weeks for the vaccine effective.

2.1.2 Prediction based on Long short-term memory (LSTM) Model

LSTM was first proposed in 1997[3] motivated by an analysis of error flow in existing RNNs[4] , which found that long time lags were inaccessible to existing architectures. LSTM is suitable for processing and predicting interval and delay events in time series and also capable of catching data from past stages and use it for future predictions.

To build our model for predicting the daily new national level hospitalization number in France without the existence of vaccine at all, we took the dataset period from beginning of pandemic break up in France 19/03/2020 to 10/01/2021. With our LSTM model uses 80% of data for training and the other 20% of data for testing. For training we use mean squared error to optimize our model using 1 window, 1 batch size and 100 epochs on our model.

After training the dataset we got a model who allows to predict the national level new hospitalization number without any vaccination to evaluate the whether the existence of vaccination is efficient to decrease the new hospitalization number or not. Finally we got minimum 0.0036 as mean squared error (MSE).And the plot of comparison the actually actual daily new hospitalization number and predicted one on period before vaccination[Figure2].

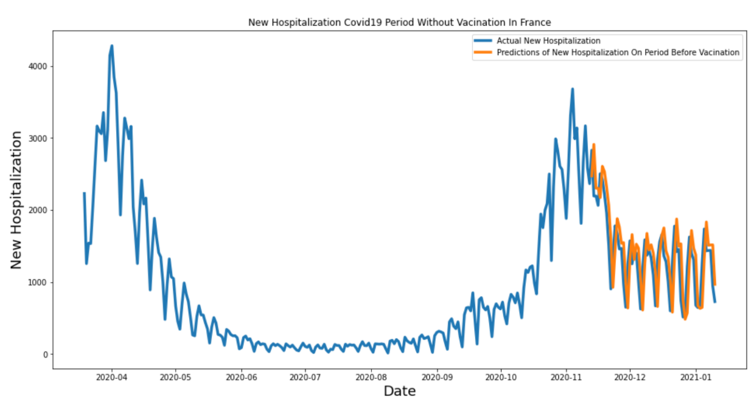

Figure 2: comparison plot between the actual daily new hospitalization number and predicted one on period before vaccination

We also build our model for predicting the daily new national level hospitalization number in France during the period of vaccination. we took the dataset period from 2 weeks after the begining of vaccination in France 10/01/2021 to 18/03/2021. With our LSTM model uses 80% of data for training and the other 20% of data for testing. For training we use mean squared error to optimize our model using relu as activation function,5 window, 1 batch size and 100 epochs on our model.

After training the dataset we got a model who allows to predict the national level new hospitalization number during the period of vaccination to evaluate the whether the existence of vaccination is able to decrease the new hospitalization number in the future. Finally we got minimum 0.004 as mean squared error (MSE).And the plot of comparison the actually actual daily new hospitalization number and predicted one on period before vaccination[Figure3].

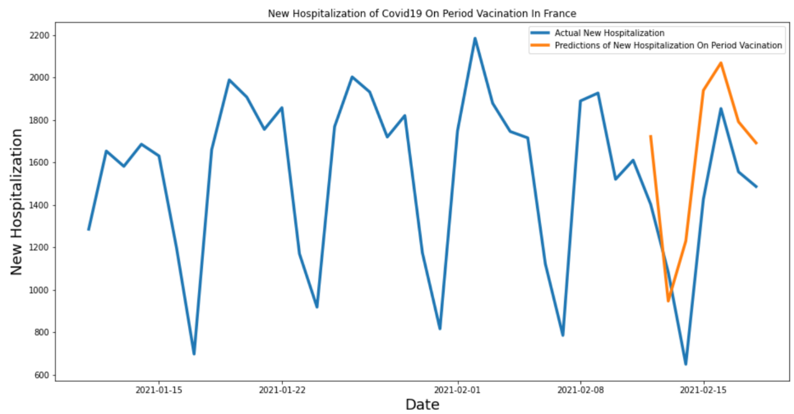

Figure 3: comparison plot between the actual daily new hospitalization number and predicted one during period vaccination Phase 1

We combianed the models using LSTM by the daily new hospitalization before and during the vaccination to obtain a complete model and also with visualization of the model validation[Figure4].

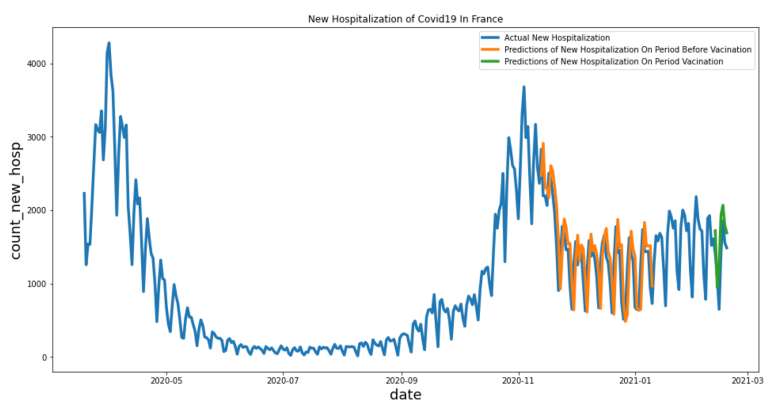

Figure 4: Combinaison of LSTM models prediction of daily new hospitalization before and during the vaccination

2.1.3 Prediction based on Convolutional neural network Model

Convolutional Neural Network models can be applied to time series forecasting. In this section, there are three 1D CNN models based on different data-sets. The general architecture is a 1D CNN layer 64 filters, 2 kernel size and ReLU activation function, followed by a size 2 MaxPool layer. Then there is a flatten layer and a dense layer with 50 neurones and ReLU activation function. The last layer is a dense layer with 1 neurones and linear activation function. It's compiled with adam optimizer and trained 100 epochs.

The input sequence is extracted from data-set with size 5 window. The data is split to 75% training data-set and 25% testing data-set. The MSE validation loss is 0.017. Then the model is used to predict the training data-set and testing data-set and incrementally taking predicted data as input to predict future number of hospitalized people.

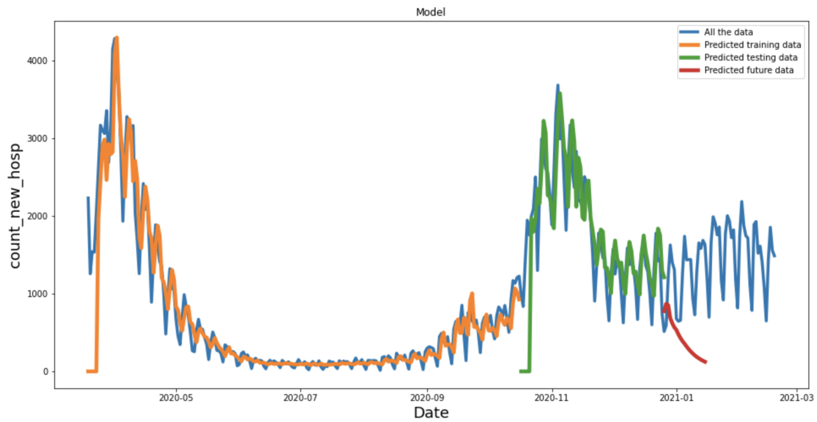

Figure 5: CNN model using only the number of people hospitalized everyday

The tendency and the values of predicted data is close to the original data. However, the red line is using predicted data to predict future data with our model and in this case it works really bad. It predicts the number of hospitalised people would decrease but in fact it increased.

The second model is to predict the number of hospitalised people by adding a new sequence of data which is the number of vaccinated people everyday. It is used multivariate CNN structure. There are two separated CNN layer taking two input sequence. And then they are merged before the dense layer. And each of them is used a general architecture described above.

In order to compare the results of weather taking into account the data before the vaccination but filled with 0 value in vaccination data. The second model only used data after two weeks of vaccination. The MSE validation loss is 0.059.

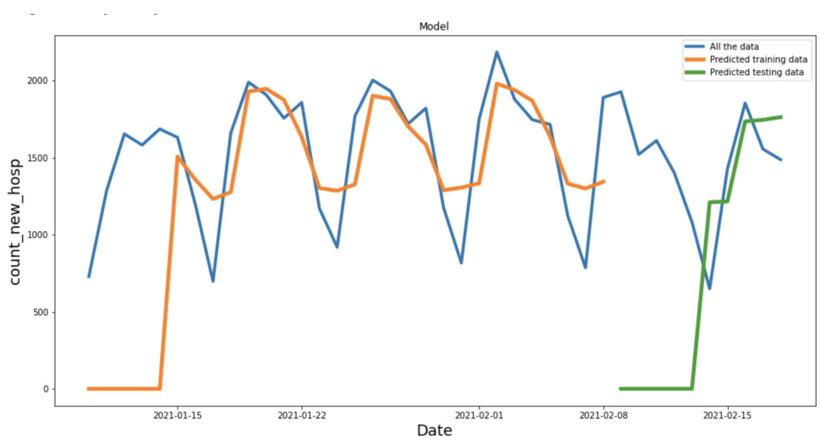

Figure 6: CNN model using only the number of people hospitalized everyday

Given by the small amount of data, the model can predict the tendency and the value is deviated not much.

The third model is try to verify will the model learn the impact of vaccine to the number of hospitalised people if giving the model more data before vaccination. Considering there are much more data before vaccination than after, the data is split as 92% to training data-set which includes around half of data after vaccination. And the MSE validation loss is 0.112.

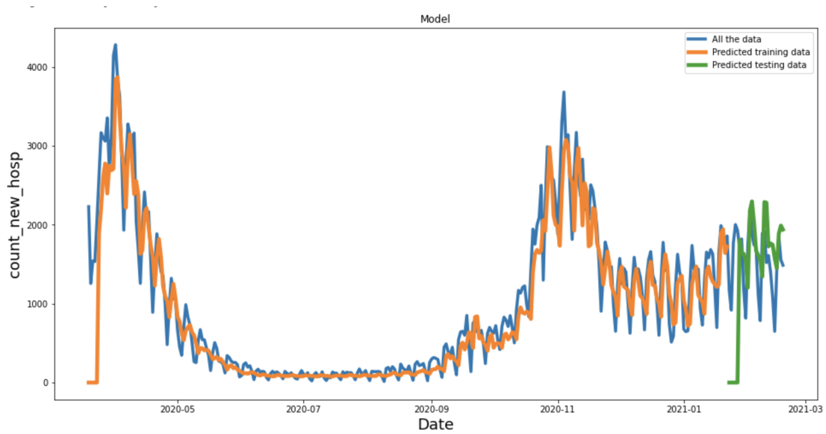

Figure 7: CNN model using only the number of people hospitalized everyday

Despite the validation loss is higher than the second model, the model can predict well the tendency. And in these two models, we don't use the model incrementally predict the future data because there are two input sequence and the number of vaccination is missing.

2.1.4 Comparison of CNN and LSTM

A convolutional neural network is traditionally used for image classification, and do not account for sequential dependencies in the way that a recurrent neural network is able to do. However, the main advantage of CNNs that make them suited to forecasting time series is the ability to use filters to compute dilations between each cell[1]. That is to say, CNNs can better understand the relationships between the different observations in the time series.

As we can see the results, LSTM is working better than CNN in this time series task. CNN can predict well the tendency but less accurate on the values. In conclusion, it is still early to see the effect of vaccination on the daily hospitalised people, given by the small amount of data.

2.2 Decision and Multi Agent Simulations

For modelling in Netlogo, We chose to answer and model the question of figuring out if it is better(more efficient or cheaper) to wait for other vaccines or start the vaccination with the current available vaccines. For modelling and and simulation, we first carried out a sensitivity analysis to identify the sensitive parameters of our model idea to better understand our model behaviour and the interpretation of the result. These parameters are: the prices of the vaccines, the efficacy percentage of the vaccines, the available date of the vaccines and also the rate of reinfected individuals after vaccination. The main idea behind our modelling to to be able to stimulate 2 different vaccines based on the indicators and parameters above.

2.2.1 Interface Description

We designed the movement dynamics of individual agents and a slider bar to change this parameter. Also, we created each slider to change hospitalization rate, the price and efficacy percentage parameters of both vaccines that is to be stimulated. For good visualization, A graph to illustrate the vaccine distribution between the two vaccines, and also Monitors to display the number of individuals vaccinated by each vaccine, number of reinfected people among vaccinated people and the reinfection rate of the vaccines. Also we have a monitor to show which vaccine is more efficient by comparing the reinfection rate.

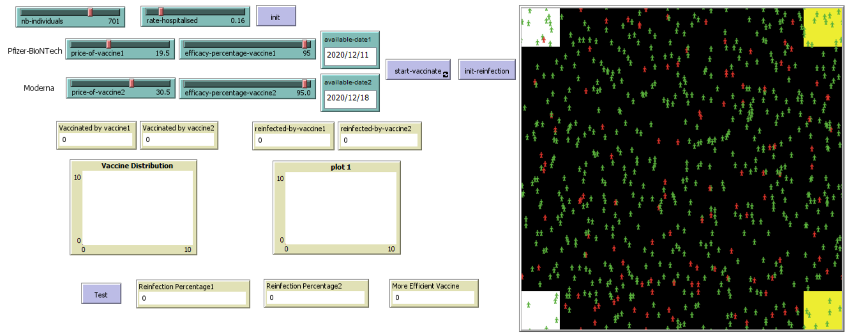

Figure 8: The interface before simulation

For the simulation process, the green agents are the healthy population while the red agents are the hospitalized agents, after the parameters are set and the simulation starts, the blue agents are the population vaccinated by Vaccine 1 while the violet agents are those vaccinated by Vaccine 2. To also note are the 4 boxes at the corners of the interface which was implanted as the vaccination centres availability, the white boxes are for vaccination centres for Vaccine 1 and the Yellow for Vaccine 2.

2.2.2 Scenario Simulation

Now we have our model ready for simulation, we came up with scenarios to explore our model under different conditions. Since there are many covid vaccinations already out and in use already, we put ourselves in the positions of experts and decision makers in regards to the vaccination and their evolution, we decided to do a simulation to recommend if it is more efficient to start with the current available vaccines or wait till the next more efficient vaccine. To have the best parameters and conditions, we found in [4] all the information (price, efficacy rate and availability date) about the currently available vaccines and vaccines that are due to be in use soon.

Firstly, we made a simulation of the currently available top two vaccines (Pfizer-BioNTech and Moderna)[5] to discover the most efficient pf them on our population agents. The parameters taken account in this simulation are the prices and efficacy percentage of these vaccines. Below is the result of this simulation.

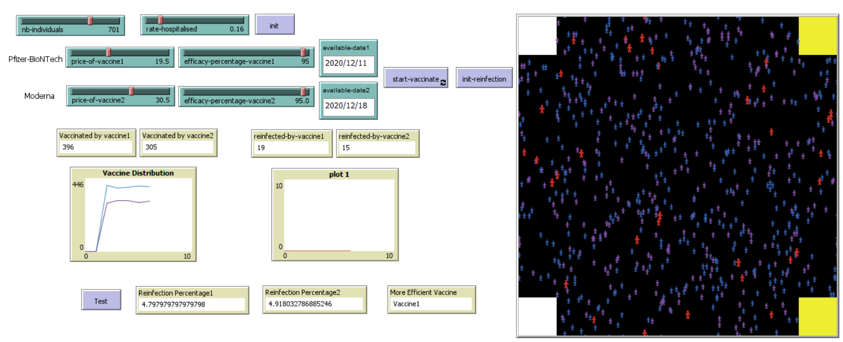

Figure 9: Simulation result between the Pfizer-BioNTech and Moderna Vaccines

In the above simulation, both Vaccines have the same efficacy rate of 95% and prices of \$19.5 and an average of \$30 respectively. Both Vaccines are great in efficiency but due to the notable price difference, the vaccination distribution of the Pfizer-BioNTech is quite greater than Moderna which means that with the same efficiency, the BioNtech vaccine would be much more available to the population because of its price advantage. We therefore recommend the Pfizer-BioNTech as the most efficient vaccine current available at the moment.

The second simulation we did is between our recommended efficient already available vaccine, the Pfizer-BioNTech and the most promising non yet available vaccine which is the Russian's sputnik V. Below is the output of this simulation:

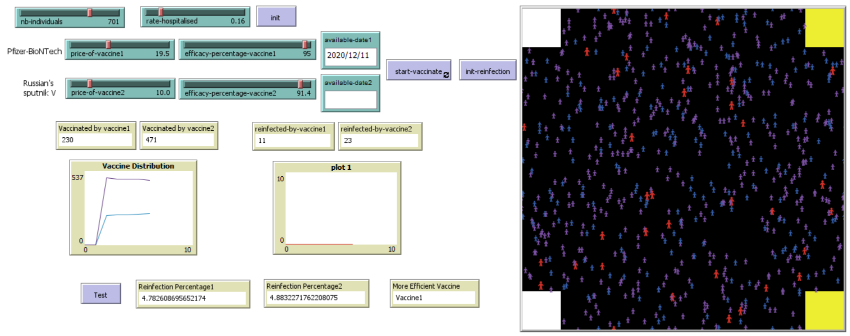

Figure 10: Simulation result between the Pfizer-BioNTech and the Russian's sputnik V Vaccines

In the simulation above, the Pfizer-BioNTech has a lower reinfection rate and higher efficiency rate (95%) compared to the Russian's sputnik (91.4%). But nevertheless, the vaccination distribution of the latter is higher due to the notable price difference, as it is twice as cheaper.

To answer the question of the choice of the two above simulated vaccines, either to go ahead with the already available Pfizer-BioNTech vaccine or to wait for promising soon to come Russian's sputnik V, we recommend that: Since the Pfizer-BioNTech efficacy is better, it should be used on the population for the main time, while we await the promising Russian's sputnik V vaccine which also has quite a very good efficacy and promises greater availability to the population because of its relative low price.

3 Conclusion and future work

This paper proposes based on LSTM and CNN built to forecast future values for new daily hospitalization at national level in France, the result of our model has shown some promising result. With the LSTM model we got few data in the dataset which could cauze the problem when predicting the future might be exist a huge bias with the future actual daily new hospitalization.

For the decision and multi-agent simulations, we were able to successfully create a model in which we can simulate, discover and make comparisons and decisions about the choice of vaccine to be used on the population, this model can be used to compare between already existing vaccines or even with fourth coming vaccines as far as the modelling parameters are already known. This model can be extended by addition of other comparative scenarios and parameters other than the ones illustrated in this paper.

References

[1] Michael Grogan (2020) *CNN-LSTM: Predicting Daily Hotel Cancellations*, `https://towardsdatascience.com/cnn-lstm-predicting-daily-hotel-cancellations-e1c75697f124`

[2] Mohamed Hawas (2020) *Generated time-series prediction data of COVID-19s daily infections in Brazil by using recurrent neural networks,Data in Brief 32 (2020) 106175*

[3] Sepp Hochreiter & Jürgen Schmidhuber(1997)*Long short-term memory,[J]Neural Computation. 9 (8): 1735–1780.*

[4] S.C. Kremer, J.F. Kol.(2001) *A field guide to dynamical recurrent neural networks,IEEE Press*

[5] Mark Terry (Feb 08, 2021) *Comparing COVID-19 Vaccines: Timelines, Types and Prices, BioSpace , https://www.biospace.com/article/comparing-covid-19-vaccines-pfizer-biontech-moderna-astrazeneca-oxford-j-and-j-russia-s-sputnik-v/*