Comparative Study of Forecasting Models for COVID-19 Outbreak in Turkey

Mert Nakıp
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
Gliwice, Poland
ORCID: 0000-0002-6723-6494

Onur Çopur
Department of Statistics
Sapienza Universiy of Rome
Rome, Italy
ORCID: 0000-0002-5564-8504

Cüneyt Güzeliş
Dept. of Electrical-Electronics Engineering
Yaşar University
Izmir, Turkey
ORCID: 0000-0001-5416-368X

***Abstract*—This paper gives an explanation for the failure of machine learning models for the prediction of the cases and the other future trends of Covid-19 pandemic. The paper shows that simple Linear Regression models provide high prediction accuracy values reliably but only for a 2-weeks period and that relatively complex machine learning models, which have the potential of learning long-term predictions with low errors, cannot achieve to obtain good predictions with possessing a high generalization ability. It is suggested in the paper that the lack of a sufficient number of samples is the source of the low prediction performance of the forecasting models. To exploit the information, which is of most relevant with the active cases, we perform feature selection over a variety of variables such as the numbers of active cases, deaths, recoveries, and population. Furthermore, we compare Linear Regression, Multi-Layer Perceptron, and Long-Short Term Memory models each of which is used for prediction of active cases together with various feature selection methods. Our results show that the accurate forecasting of the active cases with high generalization ability is possible up to 3 days because of the small sample size of COVID-19 data. We observe that the Linear Regression model has much better prediction performance with high generalization ability as compared to the complex models but, as expected, its performance decays sharply for more than 14-days prediction horizons.**

***Index Terms*—COVID-19, forecasting, machine learning, feature selection, generalization**

I. Introduction

Since the first COVID-19 case confirmed on December 2019 in Wuhan, the COVID-19 outbreak has been spreading with acceleration all around the world. According to rapid spread of this pandemic, in the most of the countries, the first concern was that the medical facilities may not be sufficient to handle with the massive number of patients. To plan necessary actions such as increasing the facilities or taking preventive decisions to flatten the curve of daily cases, the determination of the future pattern of the active cases has become one of the most important issues. So, many studies on the forecasting of the number of active cases have been published in the literature [1]–[3]. Although there are many valuable results in the published work, some of the publications optimistically make long term predictions for the pandemic [2], [4]. Furthermore, in most of the works, the test performance of the forecasting results has not been demonstrated well due to the restricted size of the available time series data covering several months only [1], [2], [4].

In this study, we perform an analysis on the generalization ability of the data-dependent forecasters to explain why forecasting models used for determining the trend of COVID-19 cases possess poor medium and long term prediction performances in the special case of machine learning models. For this purpose, a forecasting system that consists of a feature selection module and a machine learning based forecasting module is designed and implemented. In the early phase of our studies, we observed that such an architecture provides the best forecasting performance among the considered models including the standard architectures of Linear Regression (LR), Multi-Layer Perceptron (MLP), and Long-Short Term Memory (LSTM) state-of-the-art models with or without feature selection. The rational behind the choice of these models relies on the following three facts: 1) LR is a linear static model which is the least complex architecture, so possessing the high generalization ability [5]. 2) MLP is a nonlinear static neural network model which has universal function approximation property and can be said to be the most widely used neural network model with producing successful results in many applications [6]. 3) LSTM is a recurrent neural network model which is capable of approximating to the nonlinear dynamics and has proved itself as the best model in many challenging applications requiring to capture the temporal relations hidden in inherently nonlinear dynamics [7]. In order to determine the best performance provided by this feature selection based forecasting model in terms of the cross-validation error, we trained and tested all of the feature selection and forecaster pair combinations: For the feature selection module, No Feature Selection (No FS), iterative feature selection based on the Pairwise Correlation (PCorr), Recursive Feature Selection (RFS), and feature selection by using the Lasso regression (Lasso, in short) [8] are used. For the machine learning module, LR, MLP, and LSTM are chosen as the forecasters to process the selected features.

The rest of this paper is organized as follows: In Section II, we presented the related works. In Section III, we state the problem and our method proposed for the forecasting of the number of active cases. In Section IV, we present the feature selection methods and the parameter optimization for each method. In Section V, we present the implementation of the considered forecasting methods. In Section VI, we present our results on the forecasting of the number of active cases in COVID-19 pandemic. In Section VII, we present our conclusions.

II. Relationship to the State of the Art

Now, we present the relationship between our study and the works that aim to forecast the active cases in COVID-19 outbreak. According to the best of the authors' knowledge, with respect to the method of forecasting, we classify the studies that forecast the active cases in COVID-19 outbreak into 3 categories as follows: (1) SIR (Susceptible, Infected, Recovered) family [9]; (2) statistical time series analysis methods (for example, Auto-Regressive Integrated Moving Average (ARIMA)) [1]; (3) machine learning methods [10]–[12].

The SIR model is a dynamical compartmental model for describing the time evolution of a disease transmitted from human to human within a population by a set of nonlinear ordinary differential equations. In the SIR model, the total population is assumed to be constant and divided into the following classes: Susceptible (S), Infected (I), and Recovered (R) [9]. The works in [2], [9], [13] use the SIR model as the estimator for the number of active cases. These works show that the SIR model performs much better than the SEIR (Susceptible, Exposed, Infectious, Recovered) model in representing the information contained in the confirmed-case data. This indicates that predictions using more complex SIR-like dynamical models may not be reliable in comparison to the ones using simpler SIR-like models. On the other hand, although SIR-like models explains rise-and-fall nature of growth of the pandemic, they fail to capture the peak and the whole time-evolution of the disease within a reasonable accuracy due to the sensitive dependence of the time waveform of the solutions to the SIR differential equations on model parameters such as the average number of contacts per person per time.

For the COVID-19 outbreak in Italy, ARIMA which is a linear time-invariant dynamical model with stochastic input is used in [1], [14]–[16], and Seasonal ARIMA (SARIMA) is used in [17]. In [3], the exponential smoothing based models are used to predict future of the cumulative number of cases. The results of these works show that although statistical forecasters are able to capture the increasing trend of the active cases until the peak point, they are not capable of determining the whole time evolution of the disease.

Machine learning methods are also used in order to forecast the active cases in the COVID-19 pandemic. In [10], [18]–[20], the LSTM based models are used to forecast the future of the pandemic by training the model for the past COVID-19 data for each of the selected countries. MLP based models are used in [11], [21], support vector machine models are used in [22] and the logistic regression is used in [12] were trained and then tested on the COVID-19 data of each country that was selected for test. In [4], the authors took into account the problem of the small sample size for the COVID-19 pandemic and trained their model on the 2003 SARS corona virus outbreak data.

III. Statement of the Forecasting problem for The Active Cases

In this section, we describe the forecasting problem for the active cases in COVID-19 outbreak. We aim to examine the generalization ability of the machine learning based forecasters for identifying their prediction performance on the COVID-19 data. To this end, we first analyze the effects of the different features on the forecasting of active cases and then select the important features that increase the forecasting accuracy. Second, we design forecasting models that perform prediction of the number of active cases. Furthermore, we analyze the performance of the forecasters for different forecasting horizons in an increasing order, and we provide the most reliable forecasting horizon for this problem by means of an empirical analysis.

A. System Design

For the forecasting of the number of active cases, we design a system that consists of the Feature Selection module and the Forecasting module. The output of the system is the predicted value of each of the number of active cases for 1- to K-day ahead forecasting. Furthermore, the detailed explanation of the methods that are used in the Feature Selection module and the Forecasting module are given in Section IV and Section V, respectively.

B. Selection of the Important Features

Since we know that there are many different features that may affect the spread of the COVID-19 outbreak, we analyze the features that we are able to access and select the feature subset. Each feature in this subset has important effects on the number of active cases. In order to improve the performance of the overall system, we perform the feature analysis combined with the forecasting module. That is, by using the feature selection methods in Section IV, we select feature subset that achieves the best forecasting performance under the considered forecasting scheme.

C. Forecasting of the Active Cases

In the forecasting problem, we aim to compute the future value of the active cases. To this end, we use machine learning models with supervised learning whose output is future value of the active cases at Kth day. According to the best of the authors' knowledge, there is no study that examines the maximum length of forecasting horizon that provides the forecasting within a reasonable accuracy for COVID-19 pandemic. In order to determine this horizon (which is the value of K), we forecast the total number of active cases for the increasing forecasting horizon length from 1-day to 30-days. We give the details of the machine learning methods that are used as the forecasting model and their input-output structures in Section V.

IV. Analysis and Selection of the Features

In this section, we describe the methods for the selection of the relevant subset of features. For each country, first, we take past values for 14 days of each of the following daily time series features: The number of total cases, the total number of deaths, and the total number of recovered patients. Then we converted those to the following three time series: the number of active cases, the number of deaths per day, and the number of recovered patients per day.

Second, we have selected additional 36 different features that might affect the spread of COVID-19 and which are online available for all countries [23], [24]. Note that none of these features is not a time series data.

In order to select the subset of these feature candidates, we apply three different feature selection methods: Iterative feature selection based on the pairwise correlation (PCorr) of each feature candidate pair (in short, correlation matrix), Recursive Feature Selection (RFS), and feature selection by using the Lasso regression (Lasso, in short). For each of the forecasting models given in Section V for each value of K, we choose one of these feature selection methods by calculating the overall performance based on the cross-validation.

A. Feature Selection based on the Pairwise Correlation (PCorr)

In this method, first, for each feature, we calculate the Pearson correlation coefficient of this feature with each of the other features. Then, we compute the indices of the feature candidate pairs each of which whose correlation value is greater than the threshold value p or less than $-p$. Third, from each pair of feature candidates, we eliminated a feature whose average correlation with other features is the greatest.

B. Recursive Feature Selection (RFS)

The aim of the RFS is to shrink the set of the selected features in a recursive way. The RFS algorithm takes the set of feature candidates as the input. First, it trains an LR model with all of the feature candidates and keeps the coefficients of this LR. Note that we selected LR as the coefficient determination model in order to increase the generalization.

C. Feature Selection by using the Lasso Regression (Lasso)

In order to select the subset of the feature candidates, we use the classical Lasso Regression with 5-fold cross-validation over the input sample set. For each fold of the cross-validation, we split data into the training and validation sets, then train the Lasso model on training set, test it on the test set and finally calculate the test score. As the best Lasso Regression model, we select the model that achieves the highest test performance over all of the models, each of which is trained in a cross-validation fold. Finally, we select the features each of whose Lasso coefficient is not equal to zero.

V. Forecasting of the Number of Active Cases

In this section, we describe how we forecast the future values of the number of active cases and present the detailed design of the forecasting module.

In the forecasting module we use K different forecasters for 1 to K step ahead forecasting. Each of these forecasters is defined by its inputs, its output, and its internal model parameters. For each forecasting step k, we set the input of each forecasting model to the features that are selected as explained in Section IV. We set the output of this forecaster to the value of the number of active cases at the kth step in the future, denoted by $\hat{x}_1[t+k]$.

In order to forecast the future values of the number of active cases, we perform a comparative study with LR, MLP, and LSTM. We now describe the design and implementation of each of these models.

A. Linear Regression

We have selected the well-known linear regression model as a benchmark forecaster. In the implementation of this model, we use the *Linear Regression* module from *scikit-learn* library [25]. The module fits a linear model with the coefficients to minimize the residual sum of squares between the observed targets in the dataset, and the predicted values by the linear approximation.

B. Multi-Layer Perceptron

We design an MLP model, which consists of two hidden layers. We let n_l denote the number of neurons at hidden layer l. In order to find the local optimal architecture of the MLP model, we search for the values of n_l for $l \in \{1, 2\}$ within the range of $[4, 32]$ for each integral power of two. We present the resulting architecture of the MLP model and compare the performances of these models in Section VI-C. Furthermore, we set the activation function of each neuron to $tanh$. In the implementation of the MLP model, we use the *Keras* library in Python [26].

C. Long-Short Term Memory

Our implementation of the LSTM model, which is coded by using *Keras* library, consists of one lstm layer, two fully connected layers, and an output layer. We let h_{lstm} denote the number of lstm units at the lstm layer and h_e denote the number of neurons at each fully connected layer $e \in \{1, 2\}$. We exhaustively search for the local optimal values of h_{lstm} and h_e within the range of $[4, 32]$ for each integral power of two.

VI. Results

A. Dataset

In this paper, we have considered two different data domains. The first data domain is the time series data which consists of the number of active cases, the number of deaths and the number of recoveries for 71 different countries from 22th of January 2020 to 20th of July 2020. This domain contains one dataset collected from [27]. The second data

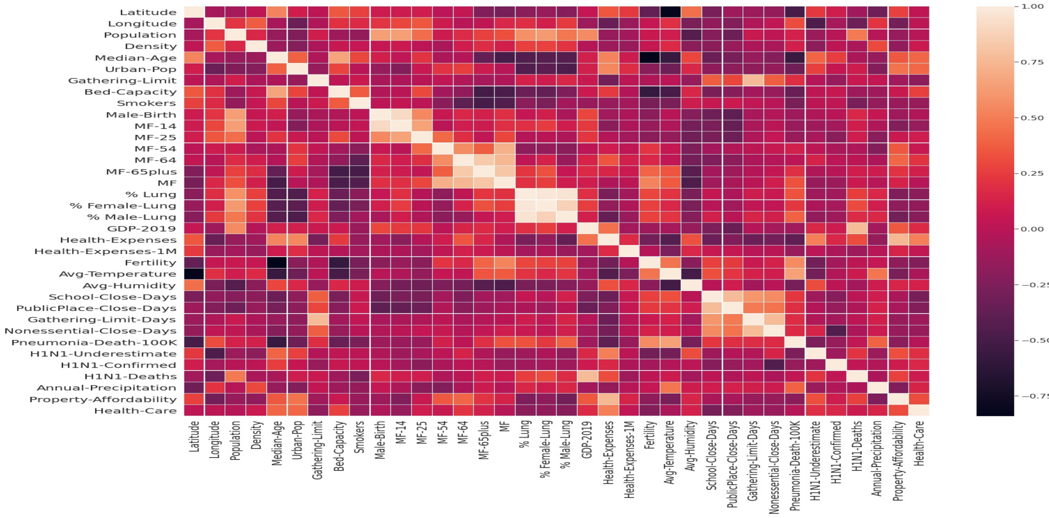

Fig. 1. Heatmap of the pairwise pearson correlation for the feature candidates that are not time-series

domain consists of two different datasets each of which includes different features that regard to each country. The first dataset in the second domain consists of 63 different features for 173 countries and is taken from [23]. The second dataset in this domain consists of 58 different features for 194 countries and is taken from [24].

We first got the intersection of all of the datasets with respect to the countries. The number of samples in the resulting dataset varies between 7440 and 9470 for different values of K. Then in the resulting dataset, we eliminated the features that are not available for all of the countries. Furthermore, we chose the subset of the country specific features, and we got 36 features. As a result, our dataset consists of 78 features, where 42 of them are the time series features during the COVID-19 pandemic, and 36 of them are the general country related features. For these feature the correlation matrix is displayed in Fig. 1.

B. Performance Evaluation by using 10-Fold Cross-Validation

For each of the LR, MLP, and LSTM models, in order to measure the generalization ability of the model, we perform 10-fold Cross-Validation (CV).

In each fold of the CV, we train the model on the training set and test it on the test set for the current fold. Then, we measure both of the training and test performance of the model by using the r^2 metric [28].

In the training of the MLP and LSTM models, we use the ADAM algorithm as optimizer with the loss selected as the mean squared error (MSE). We set the parameters of the ADAM algorithm [29] as follows: the initial learning rate to 0.001, beta1 to 0.9, beta2 to 0.999. Furthermore, we set the batch size to 200. During the training of MLP and LSTM models, we set the maximum number of epochs to 600 for the early stopping that executes the training at the epoch where the training loss has not been decreasing for the last 30 successive epochs.

C. Forecasting Results for the Active Cases

In this subsection, we discuss the predictability of the number of active cases for the COVID-19 pandemic. We also compare the forecasting performances of the LR, MLP and LSTM models.

During our experiments, we have observed: 1) LR performs the best under Lasso. 2) MLP achieves its best r^2 performance under RFS up to $K = 5$, and its performances under No FS and under Lasso are comparable after $K = 5$. 3) RFS outperforms all of the other feature selection methods for the LSTM model.

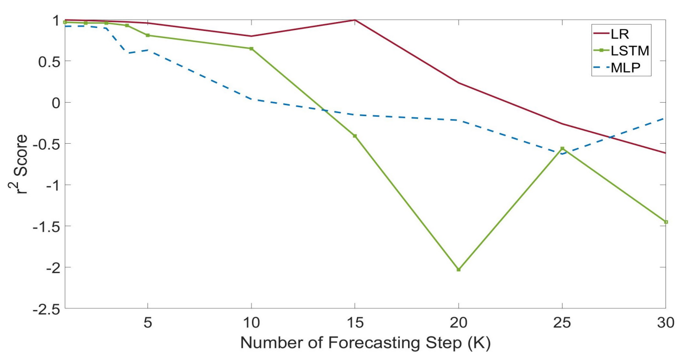

Fig. 2. Comparison of the forecasting performance of LR, MLP, and LSTM under the best feature selection method for each value of K with respect to the mean of the CV test scores.

In Fig. 2, we give the comparison of LR, MLP, and LSTM models each of which is applied together with the best performing feature selection method for each value of K. First, we see that only up to $K = 3$, the r^2 performances of all models are higher than 0.9. However, the MLP model significantly decreases at $K = 3$, where this point is $K = 4$ for the LSTM and $K = 15$ for the LR. Second, we see that after $K = 20$, there are no forecaster that achieves the r^2 score which is higher than 0. It is concluded that since the number of samples for the forecasting problem of the number of active cases during COVID-19 pandemic is quite small to represent

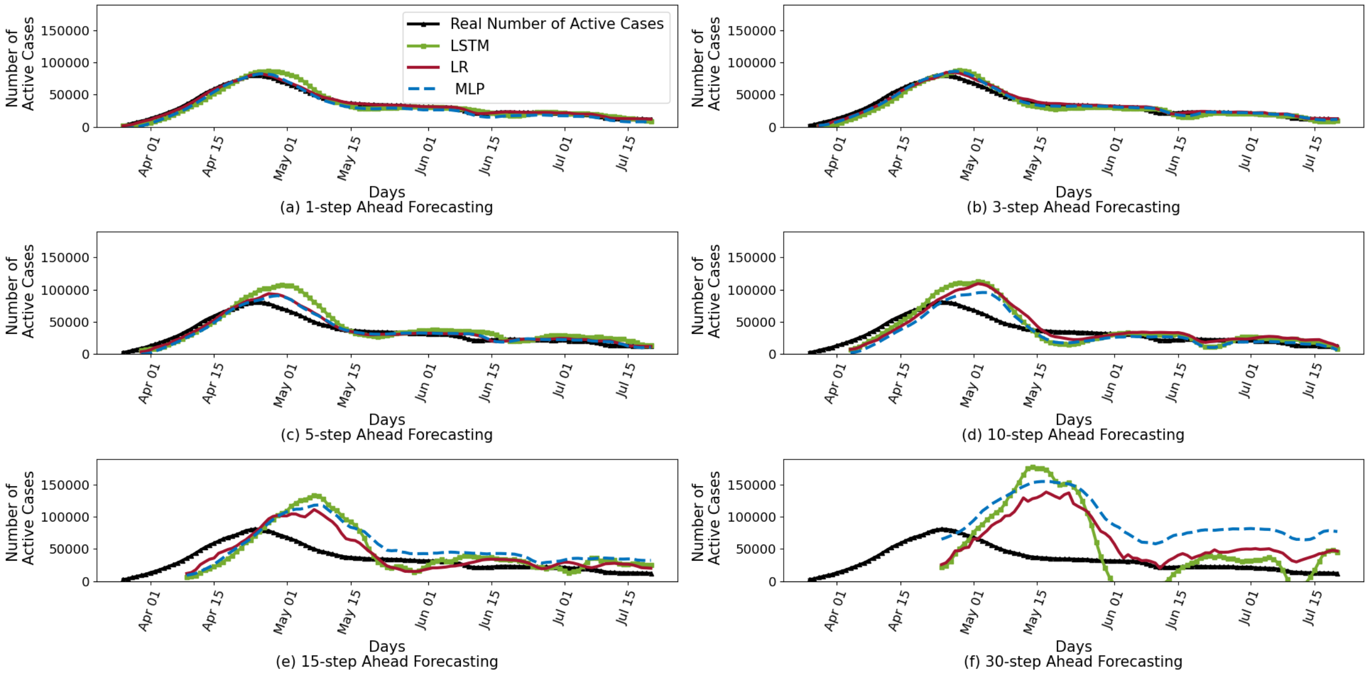

Fig. 3. Comparison of the LR, MLP, and LSTM with respect to each of 1, 3, 5, 10, 15 and 30-step ahead forecasting of the number of active cases in Turkey between 26th of March 2020 and 20th of July 2020.

the feature space, we see that LR outperforms the other two models for all values of K, except $K = 30$.

Furthermore, in Fig. 2, we see that due to small sample size, it is hard to forecast the number of active cases in COVID-19 outbreak with high generalization ability after 3 days, except the 15th day for which LR produces high prediction accuracy that might be due to the linear relation caused by the 14-day quarantine period applied to suspected persons.

1) Forecasting of Active Cases on Extended Dataset: In order to see the performance improvement with increasing sample size, we extended the dataset (which was collected from January 22 to July 20, 2020) with the number of active cases, the number of deaths and the number of recoveries for 71 different countries in COVID-19 pandemic until July 20, 2020. For this extended dataset, we repeat the methodology (in Section III) to generate the results in the rest of this section.

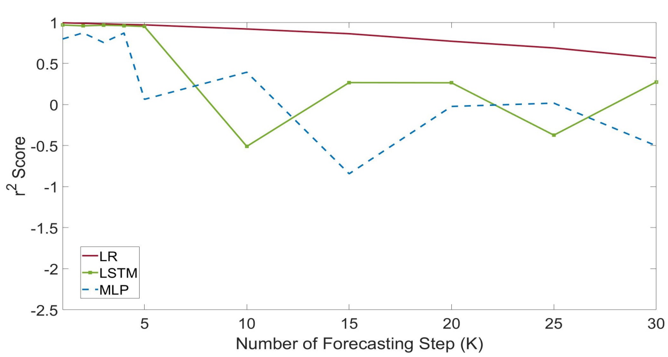

Fig. 4. Comparison of the forecasting performance of LR, MLP, and LSTM under the best feature selection method for each value of K with respect to the mean of the CV test scores on the extended dataset.

In Fig. 4, we display the r^2 performance of each forecasting scheme LR, LSTM, and MLP. We see that for each value of K, LR outperforms to both MLP and LSTM forecasters; however, even the performance of LR is around the 0.5 for $K = 15$. In addition, the r^2 performance of LSTM decreases after $K = 5$, and that of MLP decreases significantly after $K = 4$. Furthermore, due to the increased sample size of the dataset from Fig. 2 to Fig. 4, we see that the performances of all forecasting schemes increase significantly for all value of K.

2) Forecasting of Active Cases for Turkey: Now, in Fig. 3, we present the forecasting results for the number of active cases for the increasing time step ahead forecasting in Turkey between 26th of March 2020 and 20th of July 2020. From Fig. 3(a) to Fig. 3(f), we respectively set the value of K equal to 1, 3, 5, 10, 15 and 30. In Fig. 3, for each value of K, we concatenate the Kth-step ahead forecasting over the sliding windows with 1-day sliding at each step.

In Fig. 3(a), we see that the LR and MLP perform better than the LSTM forecaster, where LSTM is not able to forecast the number of active cases at around peak day. In this figure, except the days between 20th April and 10th May, all of the LR, MLP and LSTM models perform forecasting, which is close to real number of active cases. From Fig. 3(a) to Fig. 3(f), as the value of K increases, we see that forecasting performances of all forecasting schemes decreases, and LR performs the closest forecasting to the real value of the number of active cases. Thus, in Fig. 3(f), although the MLP forecasts close to real until 1st of May and LR forecasts close to real between 1st of June and 15th of June, we see that none of the forecasting models are able capture the general trend of the number of active cases and forecast the number of active cases for the peak day correctly for Turkey when $K = 30$.

VII. Conclusions

In this paper, we perform a study to determine the accurate forecasting horizon for the number of active cases in COVID-19 pandemic. To this end, we compare the performance of the Linear Regression (LR), Multi-Layer Perceptron (MLP), and Long-Short Term Memory (LSTM) for a variety of forecasting horizon lengths. Herein, the linear static model LR is chosen for its potentially high generalization ability. The most widely used static nonlinear neural network model MLP is preferred due to its powerful approximation property. The recurrent neural network LSTM is taken as the third benchmark model since it is the state-of-the-art model that is highly successful in capturing temporal relations in time series data. Considering the existence of limited number of samples only for COVID-19 pandemic, in order to achieve acceptable generalization ability for each of the three forecasters, we perform a feature selection to the input of the forecaster for reducing the model complexities. The forecaster under no feature selection (No FS) are then compared to the forecasters with the feature selection based on the Pairwise Correlation (PCorr), Recursive Feature Selection (RFS), and feature selection by using Lasso regression (Lasso), respectively.

Our main conclusion is that the long term forecasting (in other words, prediction) of the number of active cases in COVID-19 pandemic is not possible with high test accuracy at least for the considered three benchmark models as a consequence of their poor generalization abilities under the very limited number of samples available, up to now, for the COVID-19 pandemic. This study is not conclusive. The other machine learning models such as 1-dimensional or multi-dimensional Convolutional Neural Networks may be applied for forecasting COVID-19 features such as active cases. However, all of these forecasting models will suffer from the small sample size problem.

The study presented in this paper shows that the forecasting problem of the active cases might be solved by achieving the high performance and generalization ability up to 3-days ahead only. In addition, this statement is also valid for the 15th day ahead but only by using a linear model. Furthermore, even the best performing model is not able to perform better than fitting the mean of the data (which corresponds r^2 value equals to 0) after 20-days.

References

[1] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the arima model on the COVID-2019 epidemic dataset," *Data in brief*, p. 105340, 2020.

[2] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos, "Data-based analysis, modelling and forecasting of the COVID-19 outbreak," *PloS one*, vol. 15, no. 3, p. e0230405, 2020.

[3] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus COVID-19," *PloS one*, vol. 15, no. 3, p. e0231236, 2020.

[4] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai *et al.*, "Modified seir and ai prediction of the epidemics trend of COVID-19 in china under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, p. 165, 2020.

[5] X. Yan and X. Su, *Linear regression analysis: theory and computing*. World Scientific, 2009.

[6] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[7] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.

[8] R. Muthukrishnan and R. Rohini, "Lasso: A feature selection technique in predictive modeling for machine learning," in *2016 IEEE international conference on advances in computer applications (ICACA)*. IEEE, 2016, pp. 18–20.

[9] B. M. Ndiaye, L. Tendeng, and D. Seck, "Analysis of the COVID-19 pandemic by sir model and machine learning technics for forecasting," *arXiv preprint arXiv:2004.01574*, 2020.

[10] I. G. Pereira, J. M. Guerin, A. G. S. Junior, C. Distante, G. S. Garcia, and L. M. Goncalves, "Forecasting COVID-19 dynamics in brazil: a data driven approach," *arXiv preprint arXiv:2005.09475*, 2020.

[11] R. M. Rizk-Allah and A. E. Hassanien, "COVID-19 forecasting based on an improved interior search algorithm and multi-layer feed forward neural network," *arXiv preprint arXiv:2004.05960*, 2020.

[12] M. Villalobos-Arias, "Using generalized logistics regression to forecast population infected by COVID-19," *arXiv preprint arXiv:2004.02406*, 2020.

[13] W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, "Why is it difficult to accurately predict the COVID-19 epidemic?" *Infectious Disease Modelling*, vol. 5, pp. 271–281, 2020.

[14] Z. Ceylan, "Estimation of COVID-19 prevalence in Italy, Spain, and France," *Science of The Total Environment*, p. 138817, 2020.

[15] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," *Chaos, Solitons & Fractals*, p. 109853, 2020.

[16] P. Kumar, H. Kalita, S. Patairiya, Y. D. Sharma, C. Nanda, M. Rani, J. Rahmani, and A. S. Bhagavathula, "Forecasting the dynamics of COVID-19 pandemic in top 15 countries in April 2020: Arima model with machine learning approach," *medRxiv*, 2020.

[17] N. Chintalapudi, G. Battineni, and F. Amenta, "COVID-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach," *Journal of Microbiology, Immunology and Infection*, 2020.

[18] S. M. Zandavi, T. H. Rashidi, and F. Vafaee, "Forecasting the spread of COVID-19 under control scenarios using lstm and dynamic behavioral models," *arXiv preprint arXiv:2005.12270*, 2020.

[19] R. Pal, A. A. Sekh, S. Kar, and D. K. Prasad, "Neural network based country wise risk prediction of COVID-19," *arXiv preprint arXiv:2004.00959*, 2020.

[20] S. R. Vadyala, S. N. Betgeri, E. A. Sherer, and A. Amritphale, "Prediction of the number of COVID-19 confirmed cases based on k-means-lstm," *arXiv preprint arXiv:2006.14752*, 2020.

[21] S. Tamang, P. Singh, and B. Datta, "Forecasting of COVID-19 cases based on prediction using artificial neural network curve fitting technique," *Global Journal of Environmental Science and Management*, vol. 6, no. Special Issue (COVID-19), pp. 53–64, 2020.

[22] M. Yadav, M. Perumal, and M. Srinivas, "Analysis on novel coronavirus (COVID-19) using machine learning methods," *Chaos, Solitons & Fractals*, p. 110050, 2020.

[23] Patrick, "COVID-19 country data," *Kaggle*. [Online]. Available: https://www.kaggle.com/bitsnpieces/covid19-country-data

[24] M. Koryto, "countryinfo," *Kaggle*. [Online]. Available: https://www.kaggle.com/koryto/countryinfo

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[26] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[27] S. Rajkumar, "Novel corona virus 2019 dataset," *Kaggle*. [Online]. Available: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

[28] S. Glantz, "A., slinker, bk, 1990. primer of applied regression and analysis of variance."

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.