



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Early prediction of SARS-CoV-2 reproductive number from environmental, atmospheric and mobility data: A supervised machine learning approach

Pier Francesco Caruso^{a,b}, Giovanni Angelotti^c, Massimiliano Greco^{a,b,*}, Giorgio Guzzetta^d, Danilo Cereda^e, Stefano Merler^d, Maurizio Cecconi^{a,b}

^a Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, 20072 Pieve Emanuele – Milan, Italy

^b Department of Anesthesiology and Intensive Care, IRCCS Humanitas Research Hospital, Via Manzoni 56, 20089 Rozzano – Milan, Italy

^c Artificial Intelligence Center, IRCCS Humanitas Research Hospital, Via Manzoni 56, 20089 Rozzano – Milan, Italy

^d Center for Health Emergencies, Fondazione Bruno Kessler, Trento, Italy

^e Direzione Generale Welfare Regione Lombardia, Italy

ARTICLE INFO

Keywords:

COVID-19
Machine learning
Rt prediction
Epidemiology
Mobility data
Environmental data
Data science

ABSTRACT

Introduction: SARS-CoV-2 was declared a pandemic by the WHO on March 11th, 2020. Public protective measures were enforced in every country to limit the diffusion of SARS-CoV-2. Its transmission, mainly by droplets, has been measured by the effective reproduction number (Rt) that counts the number of secondary cases caused in a population by an average infectious individual at time t. Current strategies to calculate Rt reflect the number of secondary cases after several days, due to a delay from symptoms onset to reporting. We propose a complementary Rt estimation using supervised machine learning techniques to predict short term variations with more timely results.

Material and methods: Our primary goal was to predict Rt of the current day in the twelve provinces of Lombardy with the highest possible accuracy, and with no influence of the local testing strategies. We gathered data about mobility, weather, and pollution from different public sources as a proxy of human behavior and public health measures. We built four supervised machine learning algorithms with different strategies: the outcome variable was the daily median Rt values per province obtained from officially adopted algorithms.

Results: Data from 243 days for every province were presented to our four models (from February 15th, 2020, to October 14th, 2020). Two models using differential calculation of Rt instead of the raw values showed the highest mean coefficient of determination (0.93 for both) and residuals reported the lowest mean error (-0.03 and 0.01) and standard deviation (0.13 for both) as well. The one with access to the value of Rt of the day before heavily relied on that feature for prediction, while the other one had more distributed weights.

Discussion: The model that had not access to the Rt value of the previous day and used Rt differential value as outcome (FDRt) was considered the most robust according to the metrics. Its forecasts were able to predict the trend that Rt values would have developed over different weeks, but it was not particularly accurate in predicting the precise value of Rt. A correlation among mobility, atmospheric, features, pollution and Rt values is plausible, but further testing should be performed.

1. Introduction

Human transmission of novel Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) was initially reported at the end of 2019, in Hubei province in China [1]. Italy was the second country to have a large outbreak of cases with locally transmitted cases first detected on February 20th 2020, and SARS-CoV-2 was declared a pandemic by the

World Health Organization on March 11th, 2020 [2,3]. COVID-19 disease is characterized by multi-organ involvement, with a clinical syndrome in severe cases which is dominated by bilateral interstitial pneumonia, often requiring hospitalization and intensive care admission [4]. Other organs, including kidney, heart and brain, may also be affected [5,6]. SARS-CoV-2 is transmitted by droplets and airborne transmission [7], while contact transmission has a secondary role. SARS-

* Corresponding author at: Department of Biomedical Sciences, Humanitas University. Department of Anesthesiology and Intensive Care, IRCCS Humanitas Research Hospital.

E-mail address: massimiliano.greco@hunimed.eu (M. Greco).

<https://doi.org/10.1016/j.ijmedinf.2022.104755>

Received 12 November 2021; Received in revised form 4 February 2022; Accepted 27 March 2022

Available online 1 April 2022

1386-5056/© 2022 Elsevier B.V. All rights reserved.

Table 1
Classification of models created for our analysis.

	Outcome as Rt raw	Outcome as Rt differential
Access to Rt value of the previous day	CRt	CDRt
Denied access to Rt value of the previous day	FRt	FDRt

Table 2
Descriptive analysis of data from the 12 provinces of Lombardy regions.

Full dataset (n)	(2916, 50)
Unique Lombardy provinces (n)	12
Unique days (n) - from February 15th, 2020, to October 14th, 2020	243
Mobility features:	9s
provided by Google (n)	6
provided by Apple (n)	3
Weather features provided by ARPA* (n)	16
Pollution features provided by ARPA (n)	8
Rt estimates from official algorithms (median, (InterQuartile Range))	0.99 [0.73–1.32]
Rt estimates from official algorithms (mean (Standard Deviation))	1.12 (0.59)

*ARPA, Regional Environmental Protection Agency.

CoV-2 may remain viable for three hours in aerosols [8,9]. To reduce the burden of the epidemic, protective measures such as social distancing and universal masking were enforced in different countries [10]. In addition, lockdowns and curfews have been applied in many countries to slow down the spread of the virus and to reduce the burden on the healthcare settings that were saturated with COVID-19 patients [11]. SARS-CoV-2 transmission is measured by the basic reproductive number index, at time zero (R_0), and subsequently by the effective reproduction number (R_t , number of secondary cases in a population at time t), which is calculated *a posteriori* from epidemic curves (number of cases per day) and it may be influenced by testing strategies [12–14]. In Italy, a delay of about two weeks from symptoms onset to reporting has been estimated [15], which may result in a delay in the implementation of social measures. To better assess relevant social distancing measures and the

resulting public health responses, economical and psychological costs [16,17], R_t calculation should be as fast and precise as possible.

We propose a novel R_t estimation algorithm using supervised machine learning techniques and based on regularly and automatically updated data from different public sources to predict short term variations of R_t estimates.

2. Material & methods

Our primary objective was to build a model able to predict R_t with the highest possible accuracy up to the current day, not influenced by the local testing strategies and the physiological delay from symptoms onset to reporting; the final objective would be to autonomously calculate and reproduce the function that links SARS-CoV-2 infectivity to the gathered data. As a secondary objective we also decided to explore the capability of machine learning techniques to forecast R_t values based on the current day data.

We gathered public data from several public sources, including mobility, pollution, and weather data. The daily value of R_t was calculated using established methods, deriving its posterior distribution from a Markov Chain Monte Carlo algorithm applied to a known likelihood function of the observed epidemic curves [15,18]. R_t values estimated in this way were considered as model reference (“ground truth”) and used to assess predictive performance. All data collected were preprocessed and analyzed through different exploratory data analyses to confirm the coherence and to interpolate missing data. The models were created using supervised machine learning techniques for R_t prediction. We divided data in a training set and a validation set. Our models were validated, cross-validated and assessed on coefficient of determination, mean squared error and explained variance.

2.1. Data source

Each dataset from public environmental data represents either an external factor influencing human behavior or a proxy of human behavior itself. Mobility data were considered as a proxy of human activity and movements and were retrieved from Google and Apple online public mobility data [19,20]. Weather and pollution data were freely available from Lombardy regional environmental agency (ARPA) [21]

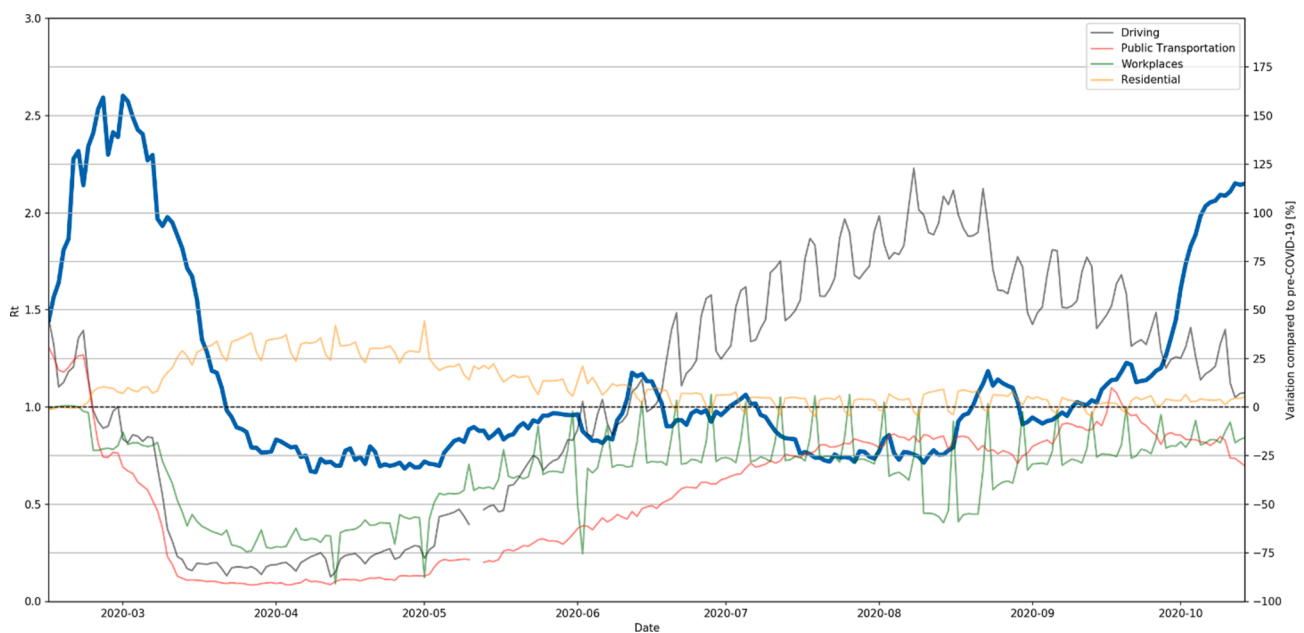


Fig. 1. Comparison between R_t values and some mobility features. On the x axis the date is reported, on the y axis R_t is reported as a bold and blue line while some macro mobility features are reported as thinner lines in other colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Coefficient of determination (R^2) performance of the four models after validation. Legend of colors: green: >0.9 , red: < 0.5 , white in between red and green. Colors shades are darker for the highest (green) and lowest (red) values.

Name	Abbreviation	CRT	FRT	CDRt	FDRt
Milano	MI	0.954	0.163	0.983	0.988
Como	CO	0.917	0.024	0.963	0.967
Monza and Brianza	MB	0.871	-0.58	0.955	0.949
Varese	VA	0.865	-1.123	0.952	0.951
Brescia	BS	0.845	-0.158	0.954	0.939
Sondrio	SO	0.879	0.071	0.937	0.937
Lecco	LC	0.915	-0.333	0.938	0.927
Lodi	LO	0.794	-0.138	0.93	0.93
Pavia	PV	0.745	-1.338	0.887	0.895
Mantova	MN	0.854	-0.723	0.915	0.906
Bergamo	BG	0.729	-1.1	0.893	0.864
Cremona	CR	0.777	-1.426	0.851	0.852
TOT	TOT	0.845	-0.555	0.93	0.925

and were included as a proxy of human activity. Both datasets were processed to achieve daily time series of average values for each province and variable of interest. Weather features and temperature describe variables which are independent from human activities, but that strongly influence human behavior itself. Pollution is a direct consequence of human behavior: higher levels of particulate in the air represent increased vehicle circulation and pollution from industrial activity.

We separated every province to consider geographical and demographical variations (i.e., city inhabitants may act differently from people from mountainous provinces, even when considering similar weather and pollution conditions) to provide the models with an improved representation of Lombardy region.

Google mobility data report aggregated anonymized movement trends. Data include six different categories of places (groceries and pharmacies, parks, retail and recreation shops, transit stations, workplaces, and residential areas) daily updated with provincial granularity. The baseline is calculated using the median value from a timespan between January 3, 2020, and February 6, 2020. Every weekday is compared to the baseline of the same weekday, reporting day-by-day percent changes.

Apple similarly report aggregated, anonymized movement trends compared to a baseline day (January 13th, 2020). Data are updated daily and are divided in three main categories: walking, driving and public transportation. Apple dataset, differently from google, provides data only at regional level. We kept both sets in models, although highly correlated, and let a purely data driven approach for feature selection derive which best suits the task.

We created high-level features representing known national holidays (i.e., June 2nd), regional lockdowns, indication of national restrictions. All these features were represented as categorical features.

In the end, the dataset contained mobility, weather, pollution, and lockdown covariates. The outcome variable was defined as daily median Rt values for each province up to October 16th, 2020. Apart from Rt, all other covariates were obtained by implementing an Extraction-Transform-Load algorithm which updates data automatically (see [Supplementary Materials](#)).

2.2. Modelling

We built a machine learning regression algorithm leveraging interpretability under a supervised framework, using the Gradient Boosting technique. We excluded support vector machine and deep learning techniques as interpretability was fundamental to preserve the decision-making process of the models. Among decision trees models, we trained with a random forest, but it showed significantly lower performances under any circumstances than the gradient boosting technique. Moreover, we favored Gradient Boosting since it provided us with quantile loss, which in turn allowed us to have lower and higher bound estimates with one training pass. The model predicts the 80% confidence interval of Rt by means of quantile loss training. We combined two different strategies according to how heavily Rt[T] depends on past Rt values and to whether differential or raw Rt[T] values are used as outcome.

The four models are as follows (Table 1):

1. CRT: has access to the Rt value of the previous day and the outcome is an Rt raw value.
2. CDRt: has access to the Rt value of the previous day and the outcome is an Rt differential value.
3. FRT: has not access to the Rt value of the previous day and the outcome is an Rt raw value.
4. FDRt: has not access to the Rt value of the previous day and the outcome is an Rt differential value.

Training was performed by means of 10-fold cross validation and hyperparameters were optimized by means of randomized selection over 50.000 different models, for a total of 500.000 models trained. Computations were performed on a cloud VM with 64CPUs and 58 GB of RAM. The optimal model was selected as the one maximizing the coefficient of determination (R^2).

The earliest 75% of the available Rt samples for each province were used as training sample (from February 15th to August 14th, 2020), while the remaining timestamps available after that date were used for validation (from August 15th to October 14th, 2020).

Python environment (Python Software Foundation. Python Language Reference, version 3.8, available at <https://www.python.org>) was

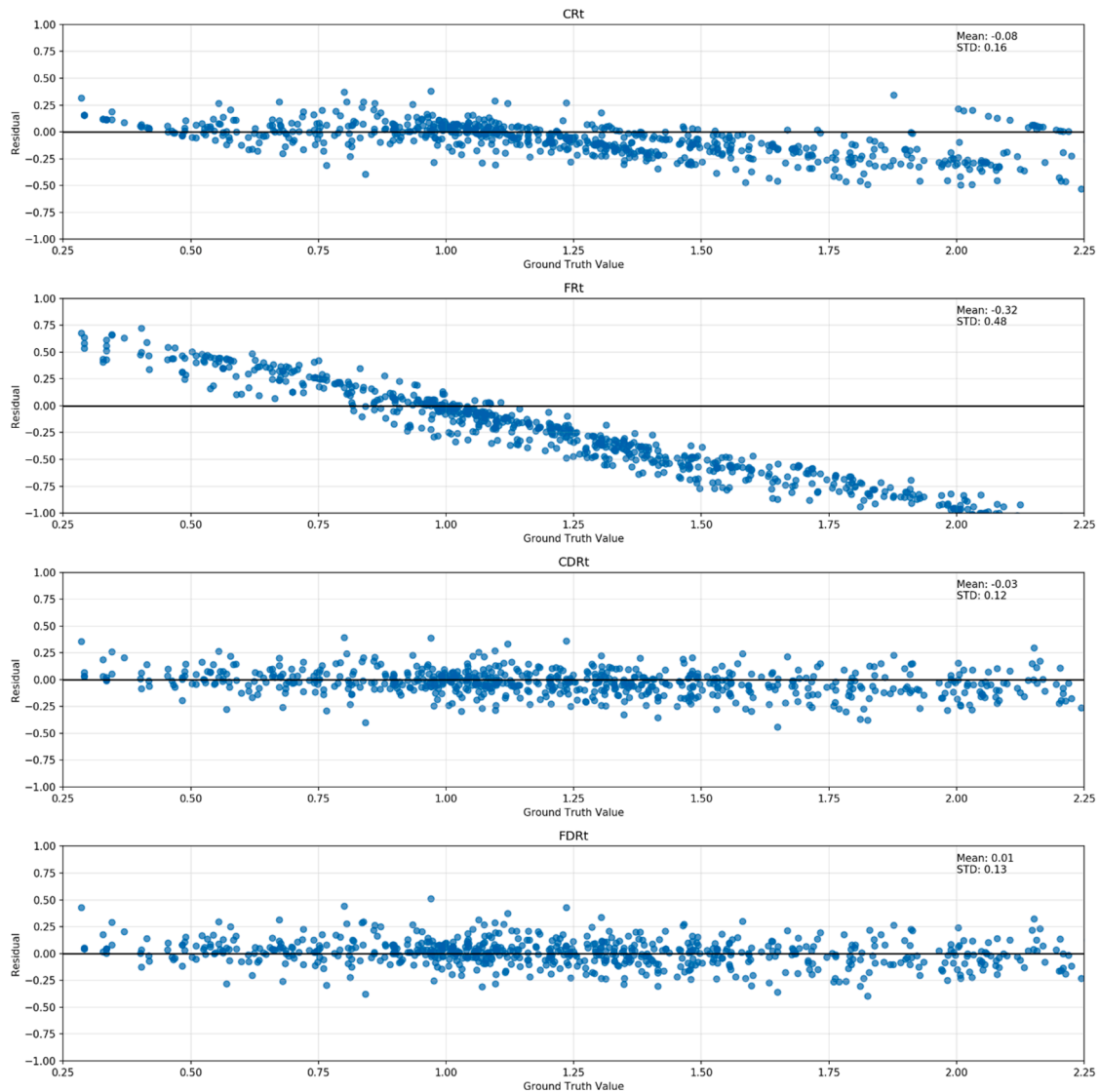


Fig. 2. Residual distribution. These graphs show how accurate the models are comparing the Rt ground truth value to the estimated ones. On the x axis the value of the estimated Rt from official algorithms (considered as ground truth) is reported while on the y axis the difference between that value and the estimated value of our models is represented. Differential based models (CDRt and FDRt) consistently handle all the considered values of Rt while the performance of CRT and FRT decreases inversely to the ground truth value of Rt reported.

used for preprocessing, data analysis and visualization, model training and validation.

More extensive methods are reported in supplemental material.

3. Results

Descriptive analysis of the data from the twelve provinces of Lombardy regions are reported in Table 2: data from 243 days (from February 15th, 2020, to October 14th, 2020) were structured and presented to all the models with an average Rt of 1.12 (Standard Deviation: 0.59) and a median value of 0.99 [InterQuartile Range? 0.73–1.32]. All median values and interquartile range for major features can be found in the supplemental Table 1.

Graphic representation of some mobility covariates together with Rt over time are reported in Fig. 1. Mobility features are expressed relatively to a pre-COVID-19 period, thus often resulting in a negative median and interquartile range. In contrast, atmospheric data are always expressed as positive real numbers.

Coefficient of determination (R^2) is computed as $1 - (\text{residual error} /$

horizontal line error). When the residual error is greater than horizontal line error, the equation yields a negative value for R^2 . The fit of the model was worse than the fit of a horizontal line. Thus, the sum-of-squares from the model is larger than the sum-of-squares from the horizontal line.

The coefficient of determination (R^2) of model validation results after training are presented in Table 3. R^2 was computed as one minus the ratio between the residual error and an horizontal line error: when the residual error is greater than an horizontal line error, the equation yields a negative value for R^2 .

CDRt and FDRt models have a mean coefficient of determination (R^2) across provinces of 0.93 and 0.925 respectively, which are the highest across the models. Analyzing the single provinces, it is possible to see that the lowest R^2 is in the Cremona province (shortened as CR) with a value of 0.825 for both models.

Fig. 2 reports model residuals, defined as the difference between the observed value and the predicted one, where differential models (CDRt and FDRt) show lower mean error and lower standard deviation compared to CRT and FRT. In addition, the models based on differential

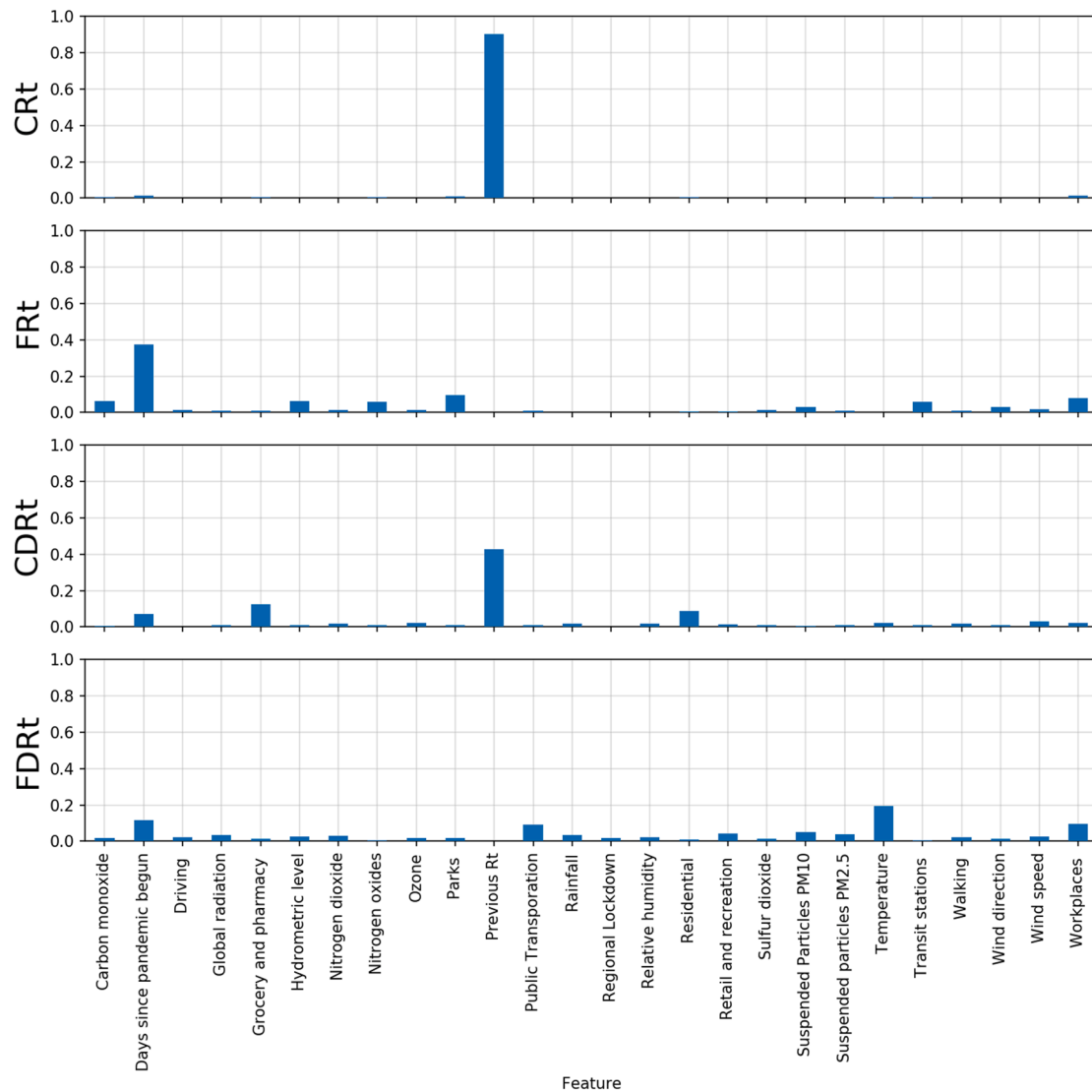


Fig. 3. Feature importance bar. Every graph represents a different model: on the x axis are reported all the features analyzed, while on the y axis is reported the weight given to that specific feature by every model. Rt calculated the day before (Previous Rt) is the most important feature when available, while the models (FDRt and FRt) spread the features importance much more when previous Rt is not available.

Rt values can more consistently handle all the considered values of Rt while the performance of CRT and FRt decrease inversely to the ground truth value of Rt reported.

Fig. 3 shows feature importance bars of all the four models. Models keeping the value of Rt of the previous day as an input heavily rely on it. The relative weight of previous Rt for CRT is 0.9. For CDrt, previous Rt accounts for around 0.4 of the predicted value; features importance is more sparsely distributed in FRt and FDRt.

Derivative models (FDrt and CDrt) achieve higher scores across provinces compared to direct Rt predictors (CRT and FRt). Accuracy in 60 days forecasts of some provinces by derivative best models are reported in Fig. 4. All the provinces' forecasts can be found in the supplemental material, Supplemental Fig. 1.

4. Discussion

Among the four built models, FDRt outperformed the others, with a very high mean coefficient of determination (R^2) across every province (0.925): ranging from 0.852 in the Cremona province (CR) to 0.988 in the Milano province (M). It is capable to handle any value of Rt due to its

derivative nature as showed by the residuals, and it relies principally on four different features (Days since the pandemic begun, public transportation, temperature, workplaces) although it was modulated by many others. Its forecasts on the test set were able to predict, in most cases and nearly in every province, the trends that Rt would have developed over different weeks, especially the upcoming surge of Rt value. Although trends were correctly detected, even our best model was not able to calculate precisely the Rt values.

FRt model, the non-derivative version that does not carry past Rt values, has very low performances, probably due to the dataset that is not well suited for this kind of modeling.

Residuals are markedly skewed for FRt, highlighting the poor performances, and marginally skewed in CRT. In the latter, residual skewness is most pronounced for Rt values above 1.5, signaling that the model is less precise with values that are above that threshold. This limitation was not found in the CDrt and FDRt models as they are more stable due to the lower influence of high and low values of Rt on models that are based on differential values.

Despite quite promising performances, the CDrt model demonstrated to be too much conservative in the test phase, as most of

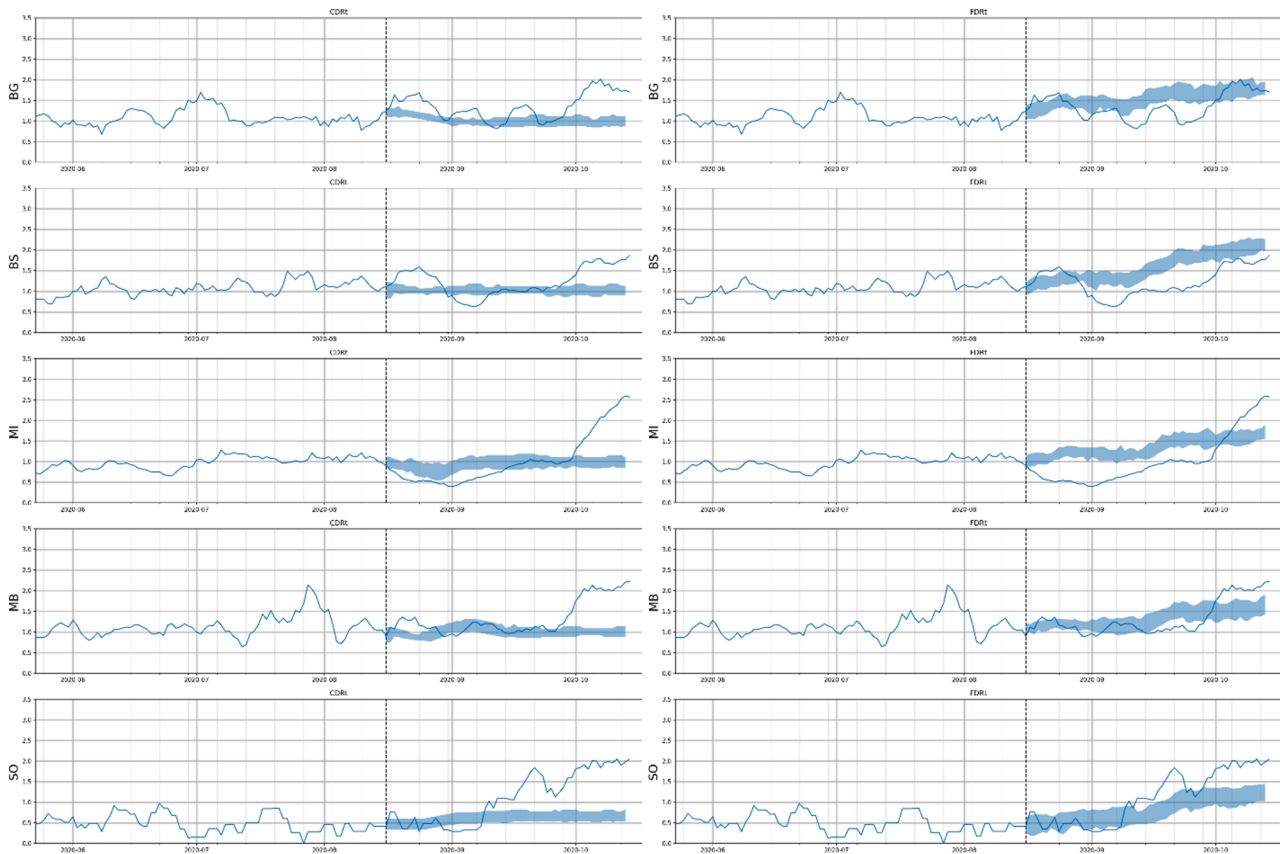


Fig. 4. Forecasts of some provinces* made by the models with the highest coefficient of determination (CDRt and FDRt) from 15th August to 14th October 2020 (60 days). On the x axis the date is reported, while on the y axis the values of Rt and the abbreviation of the name of the province. The first column visualized the forecasts made by CDRt while the second one is for FDRt. The area in blue indicates the forecast made by the model, while the blue line represents the Rt value computed using traditional methods. *BG: ‘Bergamo’, BS: ‘Brescia’, MI: ‘Milan’, MB: ‘Monza and Brianza’, SO: ‘Sondrio’. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predictions are linear and thus unable to predict variations and surges. This might be due to the excessive weight given to the value of Rt of the day before (Previous Rt) that does not let the model predict variations in the trends of the Rt values and anchors the forecasts to the values of the day before, leaving too narrow space to correctly modulate Rt.

The FDRt model automatically selected temperature as the most important variable to make decisions and this variable finds its importance confirmed in the literature as Wang et al. [22] described a negative correlation among temperature, relative humidity and the reproduction number of SARS-CoV-2 both in USA and China. A similar correlation has been described by Xiaohan et al. [23] that analyzed the city of Wuhan between January and March 2020, and by other studies [24,25], but causation is far from being defined and further studies are required [26].

Another interesting feature selected by the FDRt model was the counter of the number of days since the first detected case was discovered. We added this variable because time might be a general proxy of adaptation to the pandemic situation (instauration of correct preventive measures, higher awareness in the general population and a bigger number of research conducted on the virus that revealed new ways to address the problem).

Overall, FDRt provides satisfactory results without full loss of interpretability. Deep learning models may yield to better raw predicting performances, but at cost of more complex and non-interpretable models.

5. Limitations

Despite the promising performances reported in this paper, several

limitations must be pointed out. First, these models have been trained and tested on a limited time interval (until the beginning of the second wave in Lombardy) and the relation between our variables and the Rt function might be different now. Second, our models are tailored to the Lombardy region, thus generalization cannot be done on other locations without a correct retraining of the models to recalculate their weights; but if the same data are available, a new set of models can be easily trained on other geographical locations.

The results of our models not meant to be considered for epidemiological purposes but mostly as an evaluation of new techniques for the rapid estimation of Rt.

6. Conclusion

We trained different machine learning algorithms with the goal of estimating timelier Rt values based solely on mobility, weather, and pollution data. We used interpretable machine learning algorithms to derive predictions and identified trends associated with Rt fluctuations in a data-driven manner. FDRt, the differential model that did not rely on previous Rt values is the most promising one: it showed a high coefficient of determination, low residual values and a plausible distribution of the most important selected features.

In addition, the forecasts made with FDRt model were promising as well, showing a capability to predict future trends of Rt in the following weeks, even if it was not able to predict its value with precision. For future development, we plan to develop an automatic calibration routine to retrain these models periodically; to adjust for modifications in the relations between human behavior (e.g., vaccinations) and virus

diffusion; and as such identify the major drivers correlated with brisk changes in R_t as the pandemic evolves.

Summary Table:

What was already known on the topic?

- Prediction of SARS-CoV-2 diffusion is pivotal and very hard to achieve.
- Reproduction number (R_t) was used as proxy of virus diffusion in most developed countries, but its intrinsic limitation is a fourteen-days delay.
- Human behaviors and environmental conditions could play a role in the virus diffusion and could be considered to predict R_t variations.

What this study added to our knowledge?

- Trends of our forecasts could give insights on narrowed ranges of R_t with reduced delay compared to classic R_t calculation.
- Analysis of human behaviors and environmental conditions might help in earlier and better decision-making.
- Machine learning models perform better when calculating the variation in the slope of R_t more than in raw estimations.
- Further tests in other geographical regions/countries should be carried to validate our models.

Ethics approval and consent to participate

The institutional ethics board of Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, approved this study and waived the need for informed consent from individual patients owing to the retrospective nature of the study. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

CRedit authorship contribution statement

Pier Francesco Caruso: Conceptualization, Formal analysis, Methodology, Writing - original draft. **Giovanni Angelotti:** Conceptualization, Formal Analysis, Methodology, Writing - original draft. **Massimiliano Greco:** Conceptualization, Formal analysis, Methodology, Writing - original draft. **Giorgio Guzzetta:** Data Curation, Writing - review & editing. **Danilo Cereda:** Data Curation, Writing - review & editing, Supervision. **Stefano Merler:** Data Curation, Writing - review & editing, Supervision. **Maurizio Cecconi:** Conceptualization, Writing - review & editing, Supervision.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijmedinf.2022.104755>.

References

- [1] N. Chen, M. Zhou, X. Dong, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet*, Published online 2020, doi: 10.1016/S0140-6736(20)30211-7.
- [2] G. Grasselli, A. Pesenti, M. Cecconi, Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response, Published online, *JAMA* (2020), <https://doi.org/10.1001/jama.2020.4031>.
- [3] WHO announces COVID-19 outbreak a pandemic, Published online March 2020.
- [4] G. Grasselli, M. Greco, A. Zanella, et al., Risk factors associated with mortality among patients with COVID-19 in intensive care units in Lombardy, Italy, Published online, *JAMA Int. Med.* (2020), <https://doi.org/10.1001/jamaintermmed.2020.3539>.
- [5] V.G. Puelles, M. Lütgehetmann, M.T. Lindenmeyer, et al., Multiorgan and renal tropism of SARS-CoV-2, *N. Engl. J. Med.* 383 (6) (2020) 590–592, <https://doi.org/10.1056/NEJMc2011400>.
- [6] W.O. Vasquez-Bonilla, R. Orozco, V. Argueta, et al., A review of the main histopathological findings in the Coronavirus Disease 2019 (COVID-19), *Human Pathology*, Published online August 2020, doi: 10.1016/j.humpath.2020.07.023.
- [7] T. Greenhalgh, J.L. Jimenez, K.A. Prather, Z. Tufekci, D. Fisman, R. Schooley, Ten scientific reasons in support of airborne transmission of SARS-CoV-2, *Lancet* 397 (10285) (2021) 1603–1605, [https://doi.org/10.1016/S0140-6736\(21\)00869-2](https://doi.org/10.1016/S0140-6736(21)00869-2).
- [8] Y. Shen, C. Li, H. Dong, et al., Community outbreak investigation of SARS-CoV-2 transmission among bus riders in Eastern China, *JAMA Int. Med.* Published online September 2020, doi: 10.1001/jamaintermmed.2020.5225.
- [9] N. Van Doremalen, T. Bushmaker, D.H. Morris, et al., Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1, *N. Engl. J. Med.* 382 (16) (2020) 1564–1567, <https://doi.org/10.1056/NEJMc2004973>.
- [10] J. Walker, M.E. Fleece, R.L. Griffin, et al., Decreasing high risk exposures for healthcare-workers through universal masking and universal SARS-CoV-2 testing upon entry to a tertiary care facility, *Clin. Infect. Dis.* Published online September 2020, doi: 10.1093/cid/ciaa1358.
- [11] S. Alrashed, N. Min-Allah, A. Saxena, I. Ali, R. Mehmood, Impact of lockdowns on the spread of COVID-19 in Saudi Arabia, *Inf. Med. Unlocked* 20 (2020), <https://doi.org/10.1016/j.imu.2020.100420>.
- [12] P.L. Delamater, E.J. Street, T.F. Leslie, Y.T. Yang, K.H. Jacobsen, Complexity of the basic reproduction number (R_0), *Emerg. Infect. Dis.* 25 (1) (2019) 1–4, <https://doi.org/10.3201/eid2501.171901>.
- [13] T.V. Inglesby, Public health measures and the reproduction number of SARS-CoV-2, *JAMA* 323 (21) (2020) 2186–2187, <https://doi.org/10.1001/jama.2020.7878>.
- [14] A. Pan, L. Liu, C. Wang, et al., Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China, *JAMA* 323 (19) (2020) 1915–1923, <https://doi.org/10.1001/jama.2020.6130>.
- [15] D. Cereda, M. Tirani, F. Rovida, et al., The early phase of the COVID-19 outbreak in Lombardy, Italy. Published online 2020. Available from: <<http://arxiv.org/abs/2003.09320>>.
- [16] A.K. Singh, A. Misra, Impact of COVID-19 and comorbidities on health and economics: focus on developing countries and India, *Diabetes Metab. Syndr.: Clin. Res. Rev.* 14 (6) (2020) 1625–1630, <https://doi.org/10.1016/j.dsx.2020.08.032>.
- [17] H. Liu, A. Manzoor, C. Wang, L. Zhang, Z. Manzoor, The COVID-19 outbreak and affected countries stock markets response, *Int. J. Environ. Res. Public Health* 17 (8) (2020), <https://doi.org/10.3390/ijerph17082800>.
- [18] A. Cori, N.M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics, *Am. J. Epidemiol.* 178 (9) (2013) 1505–1512, <https://doi.org/10.1093/aje/kwt133>.
- [19] Google COVID-19 mobility data. Available from: <<https://www.google.com/covid19/mobility/>>.
- [20] Apple COVID-19 mobility data. Available from: <<https://covid19.apple.com/mobility>>.
- [21] ARPA Lombardia - Regional environmental agency. Available from: <<https://www.arpalombardia.it/>>.
- [22] J. Wang, K. Tang, K. Feng, et al., Impact of temperature and relative humidity on the transmission of COVID-19: a modelling study in China and the United States, *BMJ Open*. 11 (2) (2021) 1–16, <https://doi.org/10.1136/bmjopen-2020-043863>.
- [23] X. Si, H. Bambrick, Y. Zhang, et al., Weather variability and transmissibility of COVID-19: a time series analysis based on effective reproductive number, *Exp. Res.* 2 (2021) 1–10, <https://doi.org/10.1017/exp.2021.4>.
- [24] R. Tosepu, J. Gunawan, D. Savitri, L. Ode, A. Imran, H. Lestari, Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia Ramadhan, *Sci. Total Environ.* (2020).
- [25] J. Xie, Y. Zhu, Association between ambient temperature and COVID-19 infection in 122 cities from China, 2020.
- [26] T. Amnuaylojaroen, N. Parasin, The association between COVID-19, air pollution, and climate change, *Front. Public Health* 9 (2021), <https://doi.org/10.3389/fpubh.2021.662499>.