ADMINISTRATIVE INFORMATION

Title

Protocol for systematic review "*COVID-19 forecasting with deep learning: a distressing survey*" (Provisional title).

This protocol follows the PRISMA-P 2015 [1,2] and PRISMA 2020 [3]. statements

Registration

PROSPERO does not currently accept registrations for scoping reviews, literature reviews or mapping reviews. PROSPERO is therefore unable to accept your application or provide a registration number. This decision should not stop you from submitting your project for publication to a journal.

Authors

Luis Gutiérrez-Ríos, LGR – PhD Student at the Department of Artificial Intelligence of the UNED university.

Rodrigo de Medrano, RdM – PhD Student at the Department of Artificial Intelligence of the UNED university.

Jose Luis Aznarte, JLA- PhD and associate professor at the Department of Artificial Intelligence of the UNED university.

Contact: Jose Luis Aznarte, JLA jlaznarte@dia.uned.es. Corresponding author

Contributions: All authors contributed to draft the manuscript, the development of the selection criteria and data extraction criteria. LGR developed the search strategy and wrote the protocol. The goal is that all authors read, provide feedback and approve the final manuscript.

Amendments

Version 1 (22 July 2021).

Version 2 (14 February 2022).

- Minor corrections to "Publication Status"
- Updated publication date for studies.

Support

Sources: This systematic review is funded by the Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia – UNED

Sponsor: None.

Role of sponsor/funder: does not apply.

INTRODUCTION

Rationale

Building on the success of deep learning techniques in all sorts of classification and regression tasks, in the wake of the COVID-19 pandemic many researchers turned their tools and expertise to the task of predicting the evolution of the infection worldwide. This praiseworthy effort, based on a strong will to help, produced a panoply of models and applications aimed at helping health institutions to plan and decide on the mitigation measures that could control the spread of the pandemic, through forecasting the disease main indicators for public health.

However, forecasting is a challenging and specialized task, especially when dealing with small datasets. Predictive models must be carefully evaluated not only on their ability to capture historical events but on their exactitude in forecasting future trends, fostering a stronger appreciation of the technology's capabilities and limitations.

Objectives

The aim of this systematic review is to evaluate the methodology used to develop DL applications for COVID-19 forecasting during the pandemic, by answering the following questions:

1) Has this emergency research been prone to leave behind some of the basic requirements for quality scientific labor?
2) Can this quality be quantified for comparison purposes?

To answer the first questions, we evaluate studies that use artificial neural networks, and more precisely DL techniques to forecast the number of cases of the COVID-19 pandemic.

To answer the second question, some required 'quality criteria' are declared and assessed related to problem definition, dataset management and model identification and evaluation.

METHODS

Eligibility criteria

This review focus on, and will only include, works that are using artificial neural networks, and more precisely DL techniques, to forecast the number of cases of COVID-19. Thus, studies will be selected according to the criteria outlined below.

Cases: According to the European Centre for Disease Prevention and Control (ECDC) [4], the most accurate indicators of epidemic intensity are the absolute number of newly confirmed cases and their notification rate per 100,000 population. Hence the output of the considered models must be, at least but not limited to, the number of newly confirmed cases. This indicator is usually complemented with the number of total cases, active cases, recovered cases, deceases, and other measures.

On the other hand, the inputs will usually be the number of total recorded (confirmed) cases, but they may be accompanied by the recorded number of total cases, active cases, recovered cases, deceases etc.

Study design: Studies must use, amongst others, at least one DL technique. To stress the potential novelty of their models, certain authors tend to give imaginative or elaborate names to them, sometimes making difficult the identification. For the sake of simplicity and standardization, the models proposed in the reviewed papers were classified according to the following model taxonomy:

- Artificial neural networks (ANN) [5, 6]: multilayer perceptron [7] (MLP) or feed-forward multilayer neural network (FFNN) [8], Autoregressive Networks [9], Auto-encoders [10,11], Adaptive Networks [12].
- Recurrent Neural Networks (RNN) [13]: Long Short-Term Memory units (LSTM) [14], Gated Recurrent Units (GRU) [15], Bidirectional RNNs (BRNN) [16], Multihead attention (ATT) [17].
- Convolutional Neural Networks (CNN) [18].
- Extreme learning machines (ELM) [19].
- Ensemble methods.

Other denominations, such as Deep Neural Networks (DNN) [20], could have been ascertained into any of the previous categories, being the 'deep' characteristic an arbitrary boundary.

Timing: We consider studies published in English between 1 January 2020 and 14 February 2022. As in any other scientific discipline, urgencies do not usually favor quality, so this emergency research endeavor has not always been in line with common quality standards in research. Including early papers makes it possible to compare the results of early studies to most recent ones.

Language: Only studies published in English are considered.

Geographical range: Worldwide studies are included, with any geographical constraint.

Publication status: Published and pre-prints jobs are included. Retracted papers are excluded.

Information sources

The search was performed in well-known databases like ResearchGate, SpringerLink, Elsevier, IEEE Xplore, ACM Digital Library, arXiv, medRxiv, and Google Scholar.

To ensure exhaustively cover the literature, reference lists of the studies identified through the search and existing systematic reviews will be also scanned.

Search strategy

We scan the academic databases mentioned above, using a systematic search methodology based on the following concepts to be included and excluded:

INCLUDE	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Keywords	**Covid-19**	**Forecasting**	**Deep Learning**	**Artificial Neural Network**	**Convolutional Neural Network**	**Recurrent Neural Network**
Synonyms	Covid	Predict		ANN	CNN	RNN
	Corona	Prediction		Multilayer Perceptron		LSTM
	Coronavirus	Estimate		MLP		Long Short-Term Memory
	Sars-Cov-2	Estimation		Autoencoders		Gated Recurrent Units
		Prognosis		AE		GRU
		Foresee				

EXCLUDE	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Keywords	**X-Ray**	**Computer Tomography'**	**Sentiment**	**Drug**	**RNA**	**symptoms**
Synonyms	Radiograph	CT	-	-	Ribonucleic Acid	-
	Radiography					
	Images				Messenger RNA	
	Imaging					

Let's illustrate the search strings in metalanguage:

a) COVID search string:

OR('covid*', 'corona*', 'sars-cov-2')

b) Purpose search string:

OR('forecasting', 'predict*', 'foresee', 'estimat*', 'prognos*')

c) DL search string:

OR('Artificial Neural Network', 'Convolutional Neural Network', 'Recurrent Neural Network', 'ANN', 'CNN', 'RNN', 'Multilayer Perceptron', 'LSTM', 'MLP', 'Long Short-Term Memory', 'Autoencoders, 'Gated Recurrent Units', 'AE', 'GRU')

d) Exclude search string:

NOT('sentiment', 'drug', 'X-Ray', 'Radiograph*', 'Computer Tomography', 'CT, 'Imag*', '*RNA', 'symptoms')

Search strings are combined to create a wide range of search terms. So the final search string would be like:

AND(COVID search string, Purpose search string, DL search string, Exclude search string)

Example:

(covid* OR corona* OR sars-cov-2) AND (forecast* OR predict* OR foresee OR estimat* OR prognos*) AND ("deep learning" OR "Neural Network" OR "Artificial Neural Network" OR "Convolutional Neural Network" OR "Recurrent Neural Network" OR "Long Short-Term Memory" OR "Gated Recurrent Unit*" OR "Multilayer Perceptron" OR ANN OR CNN OR RNN OR LSTM OR MLP OR AE OR GRU OR Autoencoders) AND NOT (sentiment OR drug OR *RNA OR symptoms OR cell* OR X-Ray OR Radiograph* OR Computer OR Tomography OR CT OR Imag*)

Based on combined search terms, papers are identified and then screened manually: First, papers are sorted out based on the title, then the remaining papers are read, focusing on the abstract and methodology sections. Any other publications such as blogs, topical papers, opinion essays or commentaries, were discarded.

Study records

Data management: Literature search results will be stored in a shared Zotero database, where a first screening based on title is done and references are de-duplicated, keeping track on all existing versions.

Selection process: A second part of the screening is done based on abstract as well as the data and methodology sections of the studies. Schematic records will be held in Excel-sheets where a short description describing how every 'required quality criteria' has been resolved by the authors.

Data collection process: Data will be extracted from the identified relevant studies. Data abstracted is listed in the following section. Reviewers will go through all studies, resolving disagreements by discussion, with JLA having veto on any disagreements. Study authors will be contacted to resolve any uncertainties.

Data items: The following data will be extracted from the studies:

A. PROBLEM DESCRIPTION
   1. Object of Study.
   2. Model Identification.
   3. Forecast horizon.

B. DATASETS
   4. Data Sources.
   5. Features description.
   6. Dataset Interval.
   7. Missing data handling.
   8. Data pre-processing.

C. MODEL
   9. Software description.
   10. Accessibility.
   11. Initialization method.
   12. Network Topology description.
   13. Activation functions.
   14. Objective function and optimizer.

D. EVALUATION
   15. Validation methods.
   16. Error metrics.
   17. Benchmark comparison.
   18. Statistical inference.

Outcomes and prioritization

The primary outcome will be to summarize the DL models employed on predicting the Covid-19. The secondary output will be to try to evaluate if the research done using the proposed models follows a minimum quality.

Risk of bias in individual studies

We do not attempt to correct for biases, but we will explore any biases due to methodology, managing data, selecting evaluation metrics, and selective outcome reporting.

Data synthesis

Study data will be quantitative and qualitatively synthesized using the 'quality criteria', as summarized measures, listed in the previous section. When there are missing data, we will attempt to contact the original authors of the study to obtain the relevant missing data.

If studies are sufficiently homogeneous in terms of design and output, meta-analyses will be conducted. But diversity in study populations, validation methods or outcomes, may lead to impossibility of conducting meta-analyses at all. Sensitivity analyses, subgroup analysis or meta-regression may be not possible to be held.

A systematic narrative synthesis throughout all the summarized 'quality criteria' might be provided with information presented in the text and tables to summarize and explain the findings. Scores also will be assigned to all 'quality criteria' on each of the studies in an attempt to make comparisons between studies.

Meta-bias(es)

Detecting or correcting for publication bias in a systematic review is difficult. Therefore, publication or dissemination bias is pout of the scope of this study and will not be assessed. Selective reporting of outcomes (outcome reporting bias) will be evaluated where possible, by evaluating the Benchmark comparison held and comparing the outcomes reported in the methods and results sections of the published report.

Confidence in cumulative evidence

As the quality of evidence often varies between outcomes [21]. Each of the analyzed studies focus on a different population, on a different period of time, using different metrics for validation. In such situation, the assessments on each important outcome across studies would be difficult.

On the other hand, there is by necessity a considerable amount of subjectivity in each decision. Two persons evaluating the same body of evidence might reasonably come to different conclusions about its certainty.

We strongly recommend that assessments across outcomes for each study must be included by the authors of the reviewed studies. So this will be measured indirectly in our review through the analysis of the proposed 'quality criteria' included in the EVALUATION part (Validation methods, Error metrics, Benchmark comparison, Statistical inference). As it was stated before, those aspects will be summarized in tables and scored for comparison.

References

[1] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation BMJ 2015; 349 :g7647 doi:10.1136/bmj.g7647

[2] Moher, D., Shamseer, L., Clarke, M. et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 4, 1 (2015). https://doi.org/10.1186/2046-4053-4-1

[3] Page M J, McKenzie J E, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews BMJ 2021; 372 :n71 doi:10.1136/bmj.n71

[4] [1] Strategies for the surveillance of covid-19, en, Available at https://www.ecdc.europa.eu/en/publicationsdata/strategies-surveillance-covid-19 [Online; accessed 2021-02-28], Apr. 9, 2020.

[5] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," en, The bulletin of mathematical biophysics, vol. 5, no. 4, pp. 115–133, Dec. 1, 1943, Company: Springer Distributor: Springer Institution: Springer Label: Springer number: 4 publisher: Kluwer Academic Publishers.

[6] J. Patterson and A. Gibson, Deep Learning, en.O'Reilly Media, Inc., Aug. 2017.

[7] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," eng, Psychological Review, vol. 65, no. 6, pp. 386–408, Nov. 1958, PMID: 13602029.

[8] ——, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms," The American Journal of Psychology, vol. 76, no. 4, pp. 705–707, 1963.

[9] B. Curry and P. H. Morgan, "Neural networks, linear functions and neglected non-linearity," en, Computational Management Science, vol. 1, no. 1, pp. 15–29, Dec. 1, 2003.

[10] Y. Lecun, PhD thesis: Modeles connexionnistes de l'apprentissage (connectionist learning models), English (US). Universite P. et M. Curie (Paris 6), Jun. 1987.

[11] D. Ballard, "Modular learning in neural networks, "AAAI 87, pp. 279–284, 1987.

[12] P. J. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," PhD thesis, Harvard University, 1974.

[13] D. E. Rumelhart et al., "Learning representations by back-propagating errors," en, Nature, vol. 323, no. 6088, pp. 533–536, Oct. 1986, number: 6088 publisher: Nature Publishing Group.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, pp. 1735–80, Dec. 1, 1997.

[15] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv:1406.1078 [cs, stat], Sep. 2, 2014, arXiv: 1406.1078.

[16] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, event: IEEE Transactions on Signal Processing.

[17] A. Vaswani et al., "Attention is all you need," arXiv:1706.03762 [cs], Dec. 5, 2017, arXiv: 1706.03762 version: 5.

[18] Y. Lecun, "Generalization and network design strategies," English (US), Connectionism in perspective, 1989, publisher: Elsevier.

[19] G.-B. Huang et al., "Extreme learning machine: A new learning scheme of feedforward neural networks," in 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), vol. 2, 2004, 985–990 vol.2.

[20] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," en, Science, vol. 313, no. 5786, pp. 504–507, Jul. 28, 2006, publisher: American Association for the Advancement of Science section: Report PMID: 16873662.

[21] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of clinical epidemiology. 2011;64(4):401-6