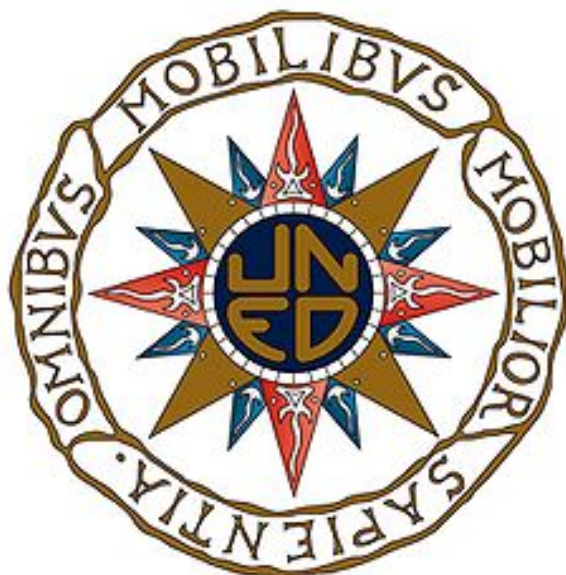


UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA.

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA
DEPARTMENT OF ARTIFICIAL INTELLIGENCE, UNED



TRABAJO FIN DE MÁSTER - MASTER THESIS.
MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL AVANZADA

General drawbacks in Deep Learning for COVID-19 Time Series Forecasting

Autor:
Luis Gutiérrez

Director:
Dr. José Luis Aznarte

September 19, 2021

Uncertainty is the only certainty there is, and knowing how to live with insecurity is the only security.

John Allen Paulos

Acknowledgements

The completion of my Masters Degree, and in particular, this dissertation would not have been possible without the support and nurturing of my wife Noelia, who accepted with stoicism my absences and changes of mood for the past years.

I would like to express my deepest appreciation and gratitude to Dr. José Luis Aznarte, who has supported me in this work, no matter how busy he had been. His comprehensive mentoring way has been crucial for bringing this project to a good end.

I would also like to extend my gratitude to PhD candidate Rodrigo de Medrano, who always was keen on sharing with me his most sincere opinion, and for his support during the difficult times.

Abstract

In the early stages of the COVID-19 breakdown, following the success of machine learning (ML) techniques, many researchers turned their efforts to predict the evolution of the global infection. In addition to classical statistical and machine learning trends, deep learning (DL) techniques are performing an important role in prediction and classifications tasks. These efforts resulted in a collection of models and applications, that were aimed to help health institutions to formulate and implement effective measures to prevent the spread of the pandemic.

Nevertheless, as it will be shown here, this emergency research activity has not always been accompanied with a minimum level of quality, affecting replicability and reproducibility. This document pretends to provide an overview about the lights and shadows on the latest trends in this specific area.

Unlike previously released literature reviews, that are providing a wide overview about any type of AI techniques applied to overall aspects of the pandemics, this document will focus specifically on the use of DL techniques applied to COVID-19 time series forecasting. The production in this field within the last months has become quite large.

After setting a group of quality criteria, related to problem definition, dataset manipulation, model identification and evaluation, 96 papers has been screened. Most of the analysed papers did not meet the common quality standards of scientific work: none of them positively scored in all of the criteria, while only about one third scored positively in at least half of the defined criteria. The emergency character of this scientific production led to getting away from some of the basic requirements for quality scientific work.

Contents

Chapters	Pages
1 Introduction	2
2 Previous works	5
3 Methodology	10
3.1 Paper Selection	10
3.2 Assessment Criteria	11
3.2.1 Criteria related with the problem description	12
3.2.2 Criteria related with the datasets	13
3.2.3 Criteria related with the model description	13
3.2.4 Criteria related with model evaluation	14
3.2.5 Proposed Scoring	15
4 Analysed models	16
4.1 Problem Description	17
4.2 Data	17
4.3 Model description	20
4.4 Evaluation	22
4.5 Final score	25
5 Discussion	27
5.1 Recommendations for Problem Description	28
5.2 Recommendations for Data	28
5.3 Recommendations for Model Description	28
5.4 Recommendations for Evaluation	30
5.5 Recommendations for Replicability and Reproducibility	31
6 Conclusions	32

Chapter 1

Introduction

World Health Organization officially named¹ the disease previously known as “2019 novel coronavirus”, as COVID-19 and the virus that causes it, as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Since WHO proclaimed the COVID-19 outbreak as a pandemic in March, 2020², the disease has followed certain patterns based on dynamic transmission of the epidemic over time and exhibited a clearly non-linear behaviour.

To try to foresee those patterns, different epidemiological models have been developed throughout the scientific community. These models can be split into two wide categories: data-driven statistical models and mechanistic models. Data-driven models use statistical regression practices and machine learning (ML) methods to predict how the disease spreads [1]. Mechanistic models are based on the fundamental laws of natural sciences, (in this case, those are epidemiological principles), constructing mathematical formulations of causal mechanisms.

This classical epidemiological approach is based on developing compartmental or Susceptible–Infected–Removed (SIR)-like models, which offer a clear epidemiological interpretation. However, predicting with them is sometimes difficult due to strong parameter value ambiguities, and the assumption that conditions for propagation will remain unchanged [2].

Conversely, Data-driven models are commonly used to predict the spread of diseases, by applying statistical regression and machine learning methods to the available data. Machine learning methods can capture complex and non-linear patterns in the data, so they are recently seen as particularly appropriate for predictions based on existing data, being considered as more accurate compared

¹[https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(COVID-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(COVID-2019)-and-the-virus-that-causes-it)

²<https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>

to common regression models.

Besides traditional machine learning models, Deep Learning (henceforth DL) models have been successfully applied to many problems, being often considered within the most powerful AI tools. These techniques consists of complex neural networks, and amongst them, we can find a wide variety of them, such like deep belief networks (DBNs), deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), auto-encoders, restricted Boltzmann machines (RBMs), and generative adversarial networks (GANs).

Forecasting is a challenging task, especially when dealing with small datasets. In any case, precise long-term predictions are often difficult to make, and the horizon for consistent predictions remains short. Prediction models must be evaluated not 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 [3]. Although these algorithms cannot diminish medical uncertainty, they can enhance the distribution of health resources [4] and can certainly improve decision making in difficult times.

In the relatively short period since the start of the pandemic, many DL applications for COVID-19 forecasting have been presented, and their performance assessed with a wide variety of metrics. However, as it will be shown here, this research race has not necessarily been in line with common quality standards in research. With a general lack of standardization, most of the works have been developed in academic environments, with little or no real-world deployment. Also, the scientific peer-review process has indeed been affected by the pandemic, sometimes lowering the quality level of the accepted papers. It came hard to find papers in which replicability and reproducibility are feasible. Most of the scientific production of DL to COVID-19 forecasting are affected by common drawbacks. This raises critical concerns about the utility of these kind of techniques for predicting and understanding these kind of emergency events.

Previous reviews already offered a very general exploration of existing common AI or ML techniques applied to drug research, genome sequencing or medical imaging. Some of the existing reviews are also covering, amongst other subjects, prediction models mainly for COVID-19 diagnosis and prognosis, but not in a very exhaustively way. However, our analysis is specifically focused on deep learning techniques applied to forecasting the number of COVID-19 infected cases. Moreover, this work is focused on the methodological shortcomings and the typical challenges that researchers are confronting when developing such models. Additionally, methodological considerations for researchers are reviewed, highlighting the possible documentation flaws in the actual literature.

To achieve this target in an objective manner, a set of quality criteria, related to problem definition, dataset manipulation, model identification and evaluation, has

been established. After an initial screening, and filtering a huge number of documents, set of 96 papers has been studied under these criteria. A quality screening has been made to highlight the methodological concerns for developing models, emphasizing on the weaknesses than can lead to issues about interpretation and replicability of the results. Even more, some specific instructions are also provided in five domains: recommendations for data handling, advices for describing the problem or object of study, suggestions for evaluating the models, proposals for describing the model and considerations for replicability and reproducibility.

This document is structured as follows: on section 2, the existing literature reviews and state-of-the-art papers about the application of DL techniques in forecasting the COVID-19 are summarized. On section 3, the methodology employed in this review is explained and a set of quality criteria covering several aspects of the scientific process is defined. On section 4, a selection of papers on which neural network approaches are applied to COVID-19 time series forecasting are reviewed under the above described criteria. Additionally, in section 5, the challenges found are examined, and the most relevant topics about the methods are discussed. And finally, in section 6, some conclusions that can be reached are shared.

Chapter 2

Previous works

The number of works published since the start of the pandemic in the field of AI applications to COVID-19, comprises a comparatively great amount of papers, given the short period of time involved. When these papers are analysed, several common shortcomings can be found in many of them. Showing results of diverse nature, with divergent metric values, in order to elucidate which of the approaches are genuinely worthwhile makes the need of a literature review utterly obvious.

When this work started, 11 state-of-the-art or review papers about AI applications in general to the different aspects of the COVID-19 (including but not limited to) forecasting were already published [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. While this analysis was being held, other 6 works were also released, namely [17, 18, 19, 20, 21, 22] and were also included into consideration here. However, most of these works had a broad-spectrum approach, making the target very general and inconclusive, reviewing only very few publications in each line of work. Other revision documents were found, such like [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33], but their primary target was mainly different (i.e. imaging and diagnostics, management, etc.).

For example, in the early work from [5], the fields of study were divided into 1) early warnings and alerts 2) tracking and prediction 3) data dashboards 4) diagnosis and prognosis 5) treatments and cures 6) social control. However, in that paper, mostly topical and opinion articles released in blogs and newspapers were cited, plus a pair of diagnosis papers and a couple of compartmental forecasting articles. In any case, authors devoted little attention to forecasting with DL techniques.

The focus of [6] was divided between blockchain and AI in general, dividing the works into estimation of coronavirus outbreak size, detection, and treatment. Within the scope of DL in forecasting, only one paper [34] was mentioned, although they emphasised the lack of unified datasets while highlighting the possibility of developing adaptive AI models for predictions. However, this work was found thus too general to answer our questions.

The study carried out by [7] covered the following fields: 1) detection and diagnosis 2) tracking and predicting the outbreak 3) infodemiology and infoveillance 4) biomedicine and pharmacotherapy. Authors propose a pair of use cases and mention challenges and solutions, especially the lack of a standard data set, forcing each model to use its own dataset, and making comparison difficult. They discuss lessons learned and give some recommendations, namely: use official datasets from health authorities, optimize algorithms, and integrate them with other methods. However, this work is not specific enough, as it covers AI in general and all aspects related to the COVID fight. Due to the colossal work that would be to cover all papers in such a wide field, it makes the selected choice arbitrary. As a result of this, some of them are just mentioned, but none was particularly analysed, resulting in that only two works [34, 35] were related to forecasting cases using deep learning approaches, both included here.

For [8], the categories were 1) quick pandemic alert 2) tracking and diagnosing cases 3) pharmacological treatment 4) public health interventions. A short set of papers were discussed, but only [36] (included here) dealt with forecasting with DL, and their final conclusions were very brief and general.

From the various models analysed in [9], only six were related to DL, while again barely two of them [37, 38] were related to forecasting the spread or cases, and were just briefly commented. Their main conclusion was very brief and open (sic): “there is a need of thorough assessment of these predictive analytic algorithm based on type of question to be answered”.

In the extensive work from [10], data sources, classical TS methods, epidemiological models, forecasting, impact and decision-making tools were analysed. In the forecasting chapter they mentioned machine learning, deep learning, ARIMA and ensemble approaches. They highlighted [39, 40] for fully connected neural networks and [36, 41, 42, 43] for recursive neural networks, while the convolutional ones were all devoted to imaging and signal processing. The main conclusions on DL approaches were about the high amount of data required, the complexity of model hyper-parametrization, and the low interpretability of the results. But as the authors themselves admit, their purpose is just to (sic) “highlight effective data-driven methodologies that have been shown to be successful in other contexts and that have potential application in the different steps of the proposed roadmap”.

Models studied by [11] were divided in 1) big data, 2) social media/other communication media data, 3) stochastic theory/mathematical models and 4) data science/machine learning techniques. In the latter category, just two works [37, 38] (also included here) were relevant to our subject and they were just concisely mentioned. The main challenges they identified were: 1) the lack of quality and quantity in data; 2) over-fitted data; 3) overly clean data with eventual integrity

loss; 4) data abundance not always improve the accuracy; 5) wrong algorithm and attribute selection that leads to misleading results; 6) model complexity that can affect the overall performance. While these questions are important, they are commonly inherent to every data-driven method. Their main conclusions were (sic): “it is important to analyse various forecasting models for COVID-19 to empower allied organizations with more appropriate information possible”. This assertion justifies by itself the existence of our paper. Moreover, the variety as well as the number of models that should be analysed must be higher in order to arrive at sound conclusions.

In the brief paper of [12], a few works were merely enumerated and categorized in 1) early detection and diagnosis of the infection 2), monitoring the treatment, 3) contact tracing of the individuals, 4) projection of cases and mortality, 5) development of drugs and vaccines, 6) reducing the workload of healthcare workers, 7) prevention of the disease. From the papers included therein, only [34] was relevant to our subject. With no identified challenge, their conclusions were both wide and general, so very little could be deduced from them.

The divisions in [13] were detection and diagnosis, virology, drug and vaccine development and epidemic. In the latter category, authors dedicated a section to outbreak detection, where a few papers [34, 36, 37, 38, 39, 44, 45, 46, 47] were just described and summarized in a table. The identified challenges were the lack of large-scale training data and the limited interaction between computer science and medicine. Still, from this paper it cannot be elucidated which DL methods could be more useful for prediction, or even more, whether DL is useful at all or not.

Deep learning, edge computing and deep transfer learning is the focus of [14], where only two of the considered papers [37, 46] were related to our scope. No conclusions could be extracted regarding DL, as its role here was only testimonial.

In a recent paper from the same authors [15], only two new citations were added compared to their previously mentioned work [5], but they were related to position articles on a blog and a website.

For [16] the main topics were 1) screening and treatment 2) contact tracing 3) prediction and forecasting 6) drugs and vaccination. Only four papers were reviewed for category 3), and only one of them was related to DL techniques. The descriptions and analysis were extensive, including the most important aspects and providing nice explanatory tables. However, the conclusions were brief (sic): “[...] deep learning algorithms [...] have more potential, robust, and advance among the other learning algorithms”, “[...] most of the models are not deployed enough to show their real-world operation [...]”. Nevertheless, the only analysed paper [41] within our scope was not enough to discard a more exhaustive further analysis.

In [17] the domains covered were 1) detection and diagnosis, 2) contact tracing, 3) forecasting, 4) vaccine development. While this paper is quite exhaustive about the role of AI in computerized tomography (CT) scans and X-Ray images, it only analyses one paper [41] in the forecasting field.

The central subjects for [18] were 1) diagnosis using radiography images, 2) diagnosis using respiratory and coughing wave data, 3) severity and survival-mortality assessment, 4) outbreak forecasting models, 5) virion sequence formation and drug discovery models. In the forecasting area they provided a wide list of 27 papers, where 12 of them [35, 36, 37, 42, 44, 46, 48, 49, 50, 51, 52, 53] were related to DL, while most of the remaining ones fell in other ML areas. Unfortunately, only four of the papers were described, while the rest were just depicted in a table by their main features. The identified challenges were: model precision and reliability impacted by quickly constructed datasets and their limited real-world implementations. The final conclusions were that the utility of AI in predicting outbreak and forecasting the spread of COVID-19 is patent, but further research is needed to identify real-world uses of AI for COVID-19.

The classification chosen by the exhaustive paper from [19] was 1) diagnosis, 2) treatment and vaccines, 3) epidemiology, 4) patient outcome and 4) infodemiology. They considered 82 studies out of the 435 retrieved, from which only a few [34, 35, 37, 38, 44, 45, 50, 54] were related to forecasting with DL. They analysed the most interesting aspects of the models, like employed techniques, features of the datasets, applications, and publishing countries. Unfortunately, the models were simply summarized in tables. Authors found that papers reported AI features and results inconsistently, where approximately one third of them did not disclose the type of validation or the data size, and a few of them did not even specify the type of AI used, affecting replicability. They admitted that they missed some studies due to the search terms employed.

In [20] the considered areas were 1) clinical applications, 2) CT and X-ray image processing, 3) epidemiology, 4) pharmaceutical, 5) text processing, 6) understanding the virus, 7) dataset collection. It is in the epidemiology section where we find an exhaustive collection of papers related to forecasting [39, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80], but those were just described without any further analysis or criticism. Their main conclusions in our field of study were regarding the size of the data, the way it is collected and the variability of formats of these data, proposing global search algorithms for training the networks to avoid local optima. While those remarks were very complete and sharp, they were given from a quite broad perspective.

In [21], the considered applications were proteins and drug development, diagnosis and outcome predictions, epidemiology and infodemiology. In the latter, it can be found some modelling and forecasting papers such like [38, 44, 45, 54]. Au-

thors found that (sic) “very few of the reviewed systems have operational maturity” and identified three issues: the need of open global repositories; the creation of multidisciplinary teams; and the need for open science, so solutions can be shared globally and adapted to other contexts. While those were very precise, it is still hard to get an clear answer to the problem of this study.

In [33] an assessment on the risk of bias in the papers reviewed was made, focusing on models to predict hospital admission in the general population, diagnostic models to detect COVID-19 infection in symptomatic individuals, and prognostic models for predicting mortality risk, progression to a severe state, or length of hospital stay.

By the time of finishing this document, a systematic review, incorporating quality criteria, of the papers covering DL techniques applied to COVID-19 imaging was released [81]. Since time series forecasting and image recognition are entirely different fields, the purpose of that work might be in a similar line of the conclusions extracted here, but there is no overlap.

In conclusion, most of the analysed review papers focus on all the fields related to the fight against COVID-19, or in any of the AI disciplines available, but in particular, none is precisely focused on forecasting with DL. Even more, the trend is to describe the methods employed and highlight the overall challenges in a general manner. The comparatively low number of forecasting methods analysed, as well as the predominance of compartmental models, traditional statistical techniques, and conventional machine learning methods versus DL ones, seems to be enough to justify the existence of this work.

Chapter 3

Methodology

3.1 Paper Selection

The object of this review are works that are using artificial neural networks and DL techniques to forecast the spread of the COVID-19 pandemic.

According to the European Centre for Disease Prevention and Control (ECDC) [82], 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.

For the sake of simplicity and standardisation, the models proposed in the reviewed papers were sorted amongst one of the following categories:

- Artificial Neural Networks (ANN) [83, 84]
 - Multilayer Perceptron [85] (MLP) or Feed-Forward Multilayer Neural Network (FFNN) [86]
 - Autoregressive Networks [87]
 - Auto-encoders [88, 89]
 - Adaptive Networks [90]
- Recurrent Neural Networks (RNN) [91]
 - Long Short-Term Memory units (LSTM) [92]
 - Gated Recurrent Units (GRU) [93]

- Bidirectional RNNs (BRNN) [94]
- Multi-head attention (ATT) [95]
- Convolutional Neural Networks (CNN) [96]
- Extreme learning machines (ELM) [97]
- Ensemble methods.

Other denominations, such as Deep Neural Networks (DNN) [98], could have been ascertained into any of the previous categories, being the ‘deep’ characteristic an arbitrary boundary.

Here, only studies published in English between 1 January 2020, and 10 May 2021, including conference proceedings, dissertations, peer-reviewed articles, and preprints has been considered. Any other publications such as blogs, topical papers, opinion essays or commentaries, were discarded. No limitations regarding the origin of publication, study design, or outcomes has been contemplated. The selected ones were cross-checked with the cited bibliography from the existing reviews that were already discussed in the previous sections, resulting in the addition of a few more papers to our study. Out of the several hundred of titles retrieved through a systematic search and independent screening by titles and abstracts, initially 96 studies were retained for full text reading.

Search examples:

```
COVID forecasting "deep learning" -images -drug -sentiment -RNA -symptoms -X-Ray -CT -Radiograph
COVID forecasting Convolutional -images -drug -sentiment -RNA -symptoms -X-Ray -CT -Radiograph
COVID forecasting LSTM -images -drug -sentiment -RNA -symptoms -X-Ray -CT -Radiograph
COVID forecasting Recurrent -images -drug -sentiment -RNA -symptoms -X-Ray -CT -Radiograph
COVID forecasting RNN -images -drug -sentiment -RNA -symptoms -X-Ray -CT -Radiograph
COVID forecasting CNN -images -drug -sentiment -RNA -symptoms -X-Ray -CT -Radiograph
```

Figure 3.1: Example of some search constraints employed.

The search was performed in well-known databases like ResearchGate, Springer-Link, Elsevier, IEEE Xplore, ACM Digital Library, arXiv, medRxiv, Plos, PubMed, or Google Scholar, excluding terms like ‘sentiment’, ‘drug’, ‘X-Ray’, ‘Computer Tomography’, ‘Imaging’, ‘RNA’ etc. or any of its variants. For an example of the queries used, see **Figure 3.1**.

3.2 Assessment Criteria

In order to assess the quality of every considered paper, and following the suggestion of other meta-analysis explained in Section 2 , for the sake of fairness, the

definition of a set of criteria is mandatory. These key indicators must represent measurable features of the papers, and must be as objective as possible. In this section, the set of key quality indicators that have been chosen for comparison of the selected papers are described. These indicators, as a set of minimum criteria, assess the information that quality papers must provide to the reader, to evidence the robustness of the model, to elucidate the conditions of the study, to explain how uncertainty is managed, and to guarantee future replicability.

In relation to concerns expressed in previous works about how AI, ML or DL are applied in the field of medicine [99, 100, 101, 102, 103, 104], our work is rooted on existing paper evaluation frameworks [105, 106], which we have adapted to the specific needs of our field of study. Despite the sharp and useful recommendations from [104], their work is mainly focused on clinical trials, and thus its main purpose is to be a guideline for developing studies rather than a literature review. From the list of items described in [105], while some of them are common to any kind of AI study, and hence applicable to our problem, the majority is exclusively applicable to medical imaging. Therefore, while the medical imaging items were not considered here, the general principles were assumed in order to elaborate our list of criteria. Finally, criteria related to forecasting were also added to the list.

Below we describe the set of considered criteria:

3.2.1 Criteria related with the problem description

In any case, to be considered as a quality paper, any article must include a specific and clear description of the problem to be solved, stating the dependent and independent variables that are considered, the area of study, forecasting period, and the employed techniques (i.e., type of ANN). Authors should avoid ambiguous assertions like ‘predicting the curve’, ‘forecasting the spread’, ‘foresee the evolution’, etc., favouring clear statements about measurable variables.

1. Object of Study. The papers must clearly indicate what is the goal of the analysis, the type of predictive modelling to be performed, the target variables to be predicted, and the definition of the variables of study which are inherent to the problem description and have a direct effect on the replicability of the experiment: area of study (province, state, region, country), variables to predict (cases, deaths, recoveries), etc.
2. Model Identification. The chosen forecasting models must be properly identified and presented. In the case that the models are not novel, citations to the original works must be provided.
3. Forecast horizon. The study must specify the time lag into the future for which forecasts are to be prepared. Generally, in the time series from our

study, this may vary from short-term forecasting horizons (weeks) to long-term horizons (years) [107]. This may have a direct impact on the prediction error [108], along with the interpretation of the results.

3.2.2 Criteria related with the datasets

The papers must contain a clear description of the dataset and the data curation procedures applied, including availability and any transformations in the ETL¹ process.

4. Data Sources. The paper must clearly state the sources of the data, providing links to them, and/or depositing data used for modelling in a publicly accessible repository.
5. Features. Variables contained in the dataset (cases, deaths, recoveries, etc.) and the area where the data is circumscribed to (province, state, region, country, hospital) must be properly described in the document.
6. Dataset Interval. The paper must explicitly include the initial and final date for the considered dataset, providing a clear view of the dataset size and the period analysed.
7. Missing data handling. The paper must specify how inconsistent, missing, and/or wrong data types were handled.
8. Data pre-processing. How raw data from various sources was converted into a time series must be clearly specified, as well as the use of normalization, rescaling and/or standardization must be also indicated.

3.2.3 Criteria related with the model description

The models must be fully identified. This has a crucial impact on the replicability and thus is a strong requisite for quality science.

9. Software. The paper must specify the names, version numbers and configuration settings used in any software, libraries, frameworks, and packages used in the experiments.
10. Accessibility. The paper must state a publicly accessible repository where algorithm descriptions, full computer code used for modelling and/or data analysis should also be provided. This allow replication and a better interpretation of the study.

¹Extract, Transform and Load (<https://databricks.com/glossary/extract-transform-load>)

11. Initialization. The paper must indicate how the initial model parameters were fixed, describing the distribution from which random values were drawn for randomly initialized parameters. The source of the starting weights must be clarified, or the weights provided, if transfer learning is employed to initialize parameters. When there is a combination of random initialization and transfer learning, it must be clear which portions of the model were initialized with which strategies.
12. Network Topology. The number of layers and how they are connected must be clearly and fully specified in the paper.
13. Activation functions. The paper must specify the number and type of cells on every layer, and the type of activation function selected in every of them [109].
14. Objective function and optimizer. The paper must precisely describe the function to be optimized, also called the cost function, loss function, or error function in minimization problems [110], and the chosen optimizer and how it has been parametrized [111].

3.2.4 Criteria related with model evaluation

Cross-validation and bootstrapping are validation methods that are typically used for evaluation of the performance or for fine-tuning the model. Alternatively, hold-out validation may address the internal validity of a model but would not accurately assess its generalizability [99]. Additionally, using hold-out in small datasets may lead to biased predictions, and in that case results will be dependent on how the data is split into train and test sets. Cross-validation can provide a better indication of how well the model will perform on unseen data, as it gives the opportunity to train on multiple train-test splits [112]. An honest validation procedure should reveal the optimism that is associated with the full modelling procedure, since model uncertainty usually is more important for optimism in model performance than parameter uncertainty [113].

Despite statistical testing for calibration is not without pitfalls [114, 115, 116], when p -values are reported with sensible precision (i.e., $p = 0.023$, instead of the conventional $p < 0.05$), together with 95% confidence intervals, the consistency between the results obtained and pure chance can be measured, thus providing a better understanding of the results.

15. Validation. The papers must clearly specify how the results were validated (hold-out, cross-validation, rolling validation, etc.) and how data were assigned into training, validation, and testing partitions.

16. Error metrics. The papers must clearly describe the error metrics employed to assess the model's performance and choose appropriate and well-known metrics for forecasting problems [117].
17. Benchmark comparison. The performance of the AI model must be compared against state-of-the-art models and naïve models.
18. Statistical inference. The papers must state what kind of hypothesis tests have been applied in order to decide whether experimental results contain enough information to cast doubt on conventional wisdom.

3.2.5 Proposed Scoring

19. Final Score. Defined as a summary of the set of criteria described above, this score will be computed as the total number of criteria that are met without any reservation (with a plain YES). Only in case of a draw, this can be resolved by comparing the number of fields scoring in an implicit way (see † in the following section), and then the ones that are partially met (see ‡ in the next section), and finally, the ones that are implicit and partially met (see †‡ in the subsequent chapter).

Chapter 4

Analysed models

At the time of writing this review, several papers about DL applications to COVID-19 have been retracted [118], in what can be considered a disturbing flaw in the quality of science, albeit it might be understandable in such emergency times. However, none of them dealt with forecasting except for one, which has been indeed withdrawn on 10 Nov., 2020 [119], leaving the total amount of considered papers in 96.

	N	Y††	Y†	Y‡	Y
1. Object of Study	0	1	30	5	60
3. Forecast horizon	23	0	0	0	73
4. Data Sources	5	0	0	0	91
5. Features	4	1	14	11	66
6. Dataset Interval	13	0	0	1	82
7. Missing data handling	80	0	2	0	14
8. Data Pre-Processing	45	0	0	8	43
9. Software	43	3	1	31	18
10. Accessibility	88	0	0	0	8
11. Initialization	76	0	4	8	8
12. Topology	19	0	0	1	76
13. Activation Functions	28	0	0	29	39
14. Objective Function & Optimizer	25	0	0	33	38
15. Validation	26	0	0	57	13
16. Error Metrics	14	0	0	0	82
17. Benchmark Comparison	18	0	0	45	33
18. Statistical Inference	88	0	0	0	8

Table 4.1: Scores per field: N (no), Y††(implicit and partially yes), Y‡(partially yes), Y† (implicitly yes), Y (yes).

All the reviewed papers were evaluated against each of the criteria defined above. Papers were ranked with an “N” when they did not meet the criterion,

and with a “Y” when this was fully satisfied. Papers were awarded a “Y with reservations”, when the criteria were partially met or with certain limitations, for example in cases when the information could be only found implicitly throughout the text (†), or only partially (‡) or both (†‡).

During this survey, several authors were contacted in order to get explanations regarding some unclear information. In particular, we got clarification from [120, 73] about a typographical error that was found, [46] regarding the problem description, [62] concerning the validation function and [64] involving some data description. On the other hand, contact was made with [121] about dataset collection dates, [70, 131, 42, 164] concerning network topology description, [167] regarding the cost function employed, [41] involving the forecast horizon applied and [148] as regards of a possible typographical error in the values of the evaluation metrics, getting no answer at all.

4.1 Problem Description

To highlight the potential novelty of their models, certain authors tend to give imaginative or fancy elaborated names to them, sometimes making the identification more difficult. Nevertheless, all the models found in the considered papers were classified according to the model taxonomy detailed in **Section 3.1**. Amongst the 96 analysed papers, a total amount of 143 models were employed. As can be seen in **Figure 4.1**, the most popular model was LSTM, followed by FFN, while GRU and CNN ranked in 3rd and 4th position.

All the considered papers explicitly state the object of study, albeit with different fortune. For example, 12 of them [37, 41, 47, 48, 59, 62, 69, 122, 123, 124, 125, 126] did this only in an implicit way, by distributing the information throughout the text. Only [38, 127] did this in a partial manner, not mentioning the variables to predict. The information provided by the former was implicit.

From all the analysed papers, 23 of them did not explicitly state the forecast horizon employed [44, 52, 53, 54, 55, 56, 60, 69, 70, 73, 80, 122, 123, 124, 127, 128, 129, 130, 131, 132, 133, 134, 135], while the rest were found to do so in one way or another.

4.2 Data

From all the reviewed papers, only 5 failed to state the source of the datasets used [54, 77, 122, 120, 136]. The other 89 mentioned up to 110 data sources in total. As can be seen in **Figure 4.2**, 21 of those employed governmental data, while twelve more used data from local or regional health agencies, such like Centers for Disease

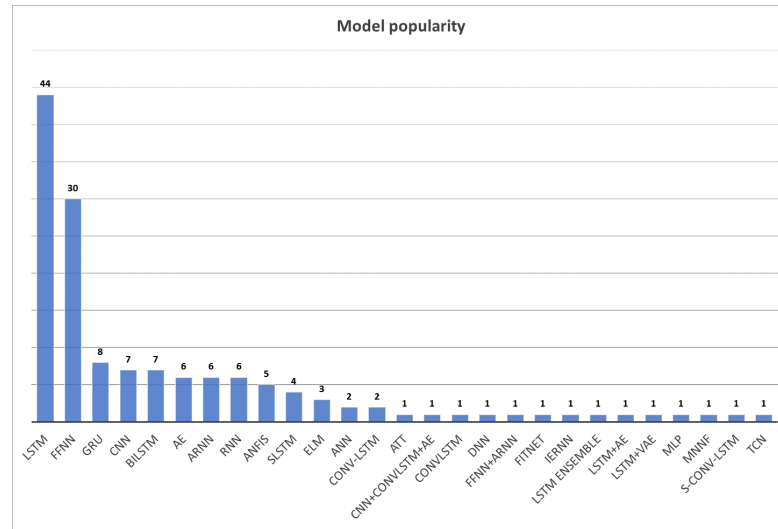


Figure 4.1: Model popularity.

Control and Prevention (CDC), European Center for Disease Control and Prevention (ECDC), Chinese Center for Disease Control and Prevention CCDC, etc. The most popular data source was the repository of the Johns Hopkins University (JHU), mentioned 37 times, whereas The Center for Systems Science and Engineering (CSSE) at JHU was specifically mentioned in only 24 of them. The main international organisation mentioned was World Health Organisation (WHO) (30 times), while publicly accessible data repositories were relatively popular: Kaggle was mentioned 5 times, Worldometers 4 times and OurWorldInData 3 times. Only two private repositories were found to be considered: an API with authorised access from [79] and hospital data from [80]

When describing the features present in the dataset, the results are more heterogeneous: 4 papers failed to report any detail at all [77, 120, 137, 135], while 10 of them only described the features partially [39, 41, 50, 52, 54, 57, 128, 133, 138, 139], while [140] did it only in an implicit manner. Other 14 provided this information implicitly and distributed throughout the whole document [42, 53, 59, 61, 63, 64, 75, 124, 126, 134, 141, 142, 143, 144]. The most common reason for this is that the variables are not specified (new cases, accumulated deaths, etc.) or even the area where the data belongs to is not declared.

The considered time interval (and thus the length of the dataset) was not stated in up to 13 of the analysed papers [58, 61, 72, 77, 123, 130, 132, 134, 137, 138, 140, 145, 135]. As can be seen in **Appendix ??**, this size varies from only 14 days used by [38] up to two hundred and eighty-four days from [146]. The average size of the dataset was 100.36 days with a standard deviation of 56.19.

With respect to missing data, only 14 studies stated how they dealt with this

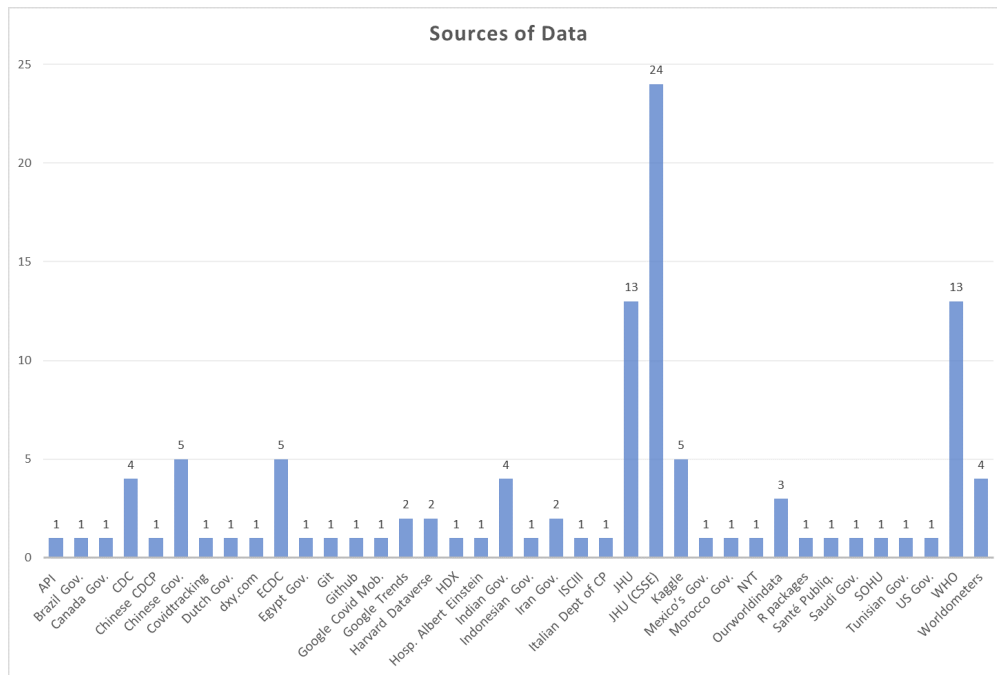


Figure 4.2: Data source popularity.

problem [43, 53, 61, 68, 71, 125, 132, 141, 147, 148, 149, 150, 151, 152], while 2 others did this implicitly [22, 136]. The rest of the papers did not mention anything about this aspect, which does not mean that they failed to approach the issue. The lack of missing data might be behind this, but it is always a good practice to explicitly state it. Amongst the papers which dealt with missing data, the approaches are heterogeneous. For example, missing data was just left blank [53], simply eliminated [71, 132, 147], or no missing data found [151]. Others replaced the missing data by the average of five previous and posterior data points [61], or by an average of one week of data [141] or by reversed order values from the sequence [125] or by using linear weighted moving average [150].

Up to 45 of the papers did not mention anything regarding what kind of data pre-processing was applied [35, 39, 42, 44, 46, 49, 50, 51, 52, 54, 55, 56, 57, 61, 62, 63, 64, 65, 67, 69, 71, 77, 80, 123, 126, 132, 134, 120, 136, 139, 143, 145, 146, 147, 149, 150, 153, 154, 121, 155, 156, 157, 158, 135], while 8 only acknowledged this partially [53, 72, 75, 124, 125, 144, 159, 160]. While this does not necessarily mean that data was not pre-processed, these kind of inscrutabilities obviously hinder replicability. The most widespread practice was minimax normalization.

As an example, an Autoregressive RNN called DeepAR was briefly analysed by [155], together with five other forecasting methods, namely ARIMA, Holt winter's exponential smoothing (HWES), Trigonometric seasonal formulation Box-Cox transformation ARMA errors and trend component (TBAT), Prophet and N-Beats

(ReLu based Fully Connected network) to predict active COVID-19 cases in the ten countries with the greatest number of total confirmed cases (70% of the confirmed cases globally). Data was retrieved from Kaggle and split into 72 instances for training, 25 for validation and 7 for test. They concluded that the tested statistical methods prevail over the two deep learning counterparts, both in Friedman statistical test rank and RMSE, due to the lack of large amounts of data. Being not a popular conclusion, this knowledge is highly valuable, as it confirms the known fact that small data can lead to overfitting. But even more interesting is the lack of fear in publishing such a negative result.

On the other hand, in [41], LSTM was chosen for confirmed cases in Canada. A small dataset was collected from JHU and Canadian Health authority until, applied a wavelet transformation filter to mitigate random noise and split in 80/20 for training/test sets. Training and testing on Canadian dataset; for short term predictions the RMSE error was about 34.8 with an accuracy of 93.4% and a RMSE of 45.70 with an accuracy of 92.67% for long term predictions. Training on Italian dataset, to predict short-term and long-term infections in Canada, had a worse performance, as expected, but still within reasonable values. These results at least prove that it is possible to make predictions for COVID-19 using DL counting on very few data (They claim to have data until March 31, 2020, but the starting date is unveiled) thanks to transfer learning techniques. It is also remarkable the impact of an intensive curation of the available data in their positive result. However, nothing was said about network architecture or hyperparameters.

Similarly, in [138] a Clustered MAE was benchmarked with three LSTM models with no prior clustering to forecast time-series data for the number of daily confirmed cases of COVID-19 in Brazil. These baseline models were one LSTM, one Deep LSTM (DLSTM), and one LSTM-SAE, being the latter the best performer. For the MAE, Uniform Manifold Approximation (UMAP) clustering algorithm was applied to all regions, to determine the similarity in their distributions. This was to avoid training the model based on data from other region with totally different dynamics. Then, a Modified Auto-Encoders (AE) was trained with the clustered data, resulting in the best performance, even when compared with SIR models. This findings result quite interesting, as a clustering approach may also help to solve the problem of the short datasets available.

4.3 Model description

Only 18 works fully documented the software packages and libraries employed, including the versions [37, 38, 39, 46, 52, 55, 73, 74, 77, 130, 120, 143, 152, 121, 156, 159, 161, 162]. While [80] did not made it explicit, it was possible to infer it from the source code. Only the name of the software could be implicitly extracted

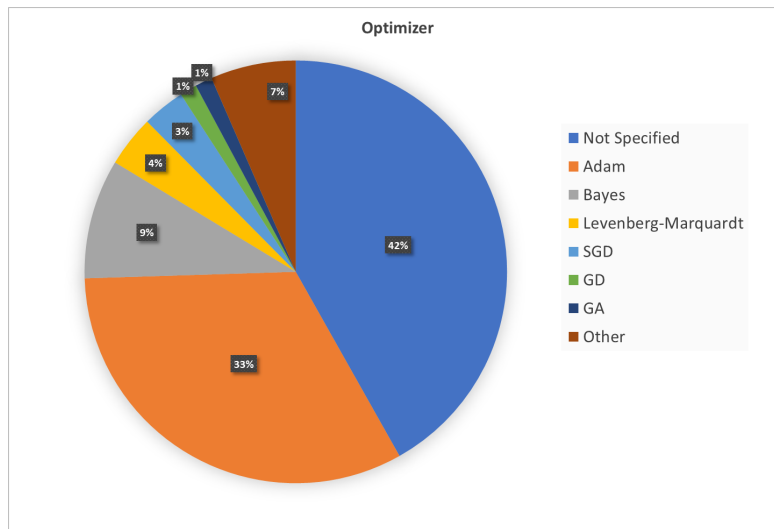


Figure 4.3: Optimizer function popularity.

from the repository in [62, 72, 155], which is not a recommended practice. From the rest, 31 papers only revealed the software name [22, 40, 42, 47, 48, 58, 61, 64, 68, 69, 70, 71, 75, 79, 124, 125, 126, 127, 132, 133, 134, 148, 149, 150, 157, 160, 163, 164, 165, 166, 135], while the others did not include any mention at all. This practice leads to difficulties in reproducibility.

Only 8 papers decided to provide a repository where the full experimental protocol could be accessed [48, 55, 62, 80, 122, 155, 156, 162]. This is an opaque practice that does not favour replicability.

Concerning the initialization of the model, 8 of the articles undisclosed the chosen way for initializing the weights in an unambiguous manner [38, 62, 79, 125, 130, 133, 120, 152], while 4 more mentioned this just in an implicit way [55, 80, 155, 162]. Other 8 decided to provide this information just in a partial way, only declaring that this was done randomly, but without specifying which distribution was used [22, 43, 51, 63, 122, 151, 156, 135]. The rest did not make any mention at all about this aspect.

Regarding network topology, up to 19 of the works failed to mention the number and type of layers from which the network was made up [38, 41, 42, 56, 57, 64, 67, 123, 130, 132, 139, 141, 143, 149, 153, 154, 155, 164, 135]. Only [122] did this task in a partial way, while the rest clearly stated this fact.

From all of the analysed papers, up to 27 failed to explain the number of units employed in each layer and their activation functions [13, 38, 41, 42, 45, 54, 56, 57, 60, 61, 64, 69, 77, 123, 126, 130, 132, 141, 143, 149, 153, 154, 121, 155, 161, 164, 135]. Other 29 did it only in a partial way [34, 36, 37, 39, 44, 46, 47, 49, 52, 53, 62, 63, 74, 76, 79, 122, 124, 127, 128, 133, 134, 137, 139, 145, 156, 157, 159, 162, 166], while

the rest made this information explicit and complete. The most popular was ReL followed by Sigmoid + Tanh (due to the popularity of LSTMs) and standalone Tanh.

Up to 25 of the works failed to describe the selected objective function, and/or the optimizer applied to minimise it [42, 44, 52, 53, 54, 57, 61, 63, 67, 70, 74, 78, 130, 131, 132, 120, 139, 141, 143, 144, 145, 121, 155, 156, 166]. Another 33 succeeded in this task only partially [37, 38, 41, 45, 46, 47, 48, 49, 55, 56, 59, 62, 64, 69, 71, 73, 76, 122, 123, 124, 125, 126, 127, 129, 133, 137, 140, 146, 149, 153, 154, 164, 167]. As can be seen in Figure 4.3, the most frequently chosen optimizer was Adam, followed by far by Bayesian optimizer.

For instance, in [76] Artificial Neural Networks (ANN) and Auto-Regressive Integrated Moving Average (ARIM1) models were applied for prediction of Iranian new cases. The data from Iran Ministry of Health reports and JHU, was normalized through Box-Cox transformation, and divided into 35 days for training and 6 days for testing. ANN was trained with three hidden layers, each of them containing 10 neurons. An exhaustive statistical analysis performed: results were given with 95% confidence interval, significance tests, residuals versus observation order plots, histogram and ACF-PACF plots of the residuals, Shapiro-Wilk test for residual normality and Box-Lung test for residual stationarity. However, both models were over-optimistic, showing huge values for MSE and MAE. Very little was explained regarding the hyperparameters, which may lead to think that the model was not fully optimized.

On the other hand, in [167] some Evolutionary Algorithms (EAs), such like GA, PSO and GWO to predict the COVID-19 outbreak. Data were collected from Worldometers for total cases over 30 days in Italy, Germany, Iran, USA, and China. The MLP was trained using 8, 12, and 16 inner neurons. Each input of the ANFIS was handled by two membership function (MF), which were tested with the Tri shape, Trap shape, and Gauss shapes. The output MF type was selected to be linear with a hybrid optimizer type. GWO was the best EA, but far below the ML approaches. MLP outperformed in all the countries, followed by the ANFIS. They stated that MSE was the chosen cost function for GWO, the presented formula corresponded to RMSE. The EAs employed showed issues with local search and tend to get trapped in local optimum in high-dimensional space, while requires low knowledge about the problem.

4.4 Evaluation

Concerning evaluation, only 13 of the analyzed papers informed about a full cross-validation method [37, 43, 46, 55, 63, 71, 80, 126, 127, 141, 152, 159, 166], while other 25 did not mention if any kind of validation was performed at all [34, 35, 36,

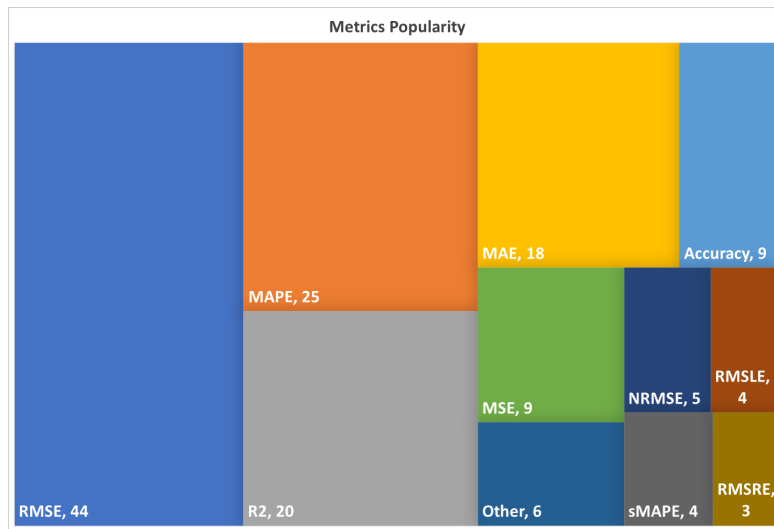


Figure 4.4: Metrics popularity.

38, 47, 53, 54, 57, 58, 59, 61, 72, 77, 123, 134, 138, 139, 143, 145, 147, 148, 149, 158, 165, 135]. This worrisome fact of course makes the interpretation of the results an exercise of faith. The 57 remaining papers only performed a 1-split hold-out validation, which of course can introduce some bias in their conclusions, especially when the size of the dataset is small. The most common split rate was 80/20 for test and validation, respectively, followed by 90/10 and 75/25. Only [154, 156] applied a 50/50 split, and [73] even went for 40/60. This practice is not advisable, specially with such small datasets, as models will find more difficulties to learn the general principles and will show a poor validation and test performance.

Unfortunately, from all the reviewed papers, up to 14 failed to provide any error metric at all [36, 50, 56, 57, 62, 64, 67, 77, 123, 124, 128, 131, 143, 145]. Furthermore, while the problem at hand is clearly a regression one, 8 of the studies employed only metrics for classification, making difficult to understand how far the predictions were from the actual values [80, 137, 140, 141, 142, 149, 121, 157]. Particularly, [137, 140] used an own formula in an effort to 'adapt' accuracy metric to prediction problems. Another five articles used a mix of classification and regression metrics, leaving some room for comparisons [41, 58, 127, 130, 146]. The rest provided only regression metrics. From Figure 4.4 it is clear that the most common metric was RMSE, followed by MAPE, R^2 and MAE, while up to 6% of the times, accuracy was chosen.

As it was mention here before, comparison with naïve and state-of-the-art models is key to prove the goodness of any forecast attempt. Of all the analyzed papers, 18 papers did not include any kind of benchmark comparison against any other model [34, 41, 42, 51, 53, 55, 56, 57, 62, 66, 77, 131, 132, 120, 142, 158, 163, 164].

Another 31 only compared their proposals against complex algorithms, assuming that all of them are thus better than basic persistence or random approaches, which may lead to problematic conclusions [22, 37, 38, 46, 47, 59, 61, 63, 64, 67, 70, 75, 76, 78, 128, 129, 130, 133, 134, 137, 139, 140, 143, 144, 146, 152, 153, 154, 121, 155, 157].

Less than 10% of the papers reported the application of some kind of statistical inference to their results, and thus, for the rest, it is difficult to assess that the true gain of the model is not due to simply chance [22, 46, 65, 76, 129, 141, 155, 135].

Regarding confidence intervals, only 18 papers [57, 67, 76, 141, 38, 153, 143, 145, 69, 146, 36, 42, 41, 122, 22, 65, 62, 125] employed them to communicate their results, while [156] mentioned this during the training phase only. From those ones mentioned, solely a few of them [57, 145, 65, 125] employed the intervals for accompanying the numerical results, while the rest only applied them for the charts. As an example, [42, 38] only used them for only one out of the several curves provided.

In particular, only 11 of those [76, 141, 153, 143, 145, 146, 36, 42, 65, 62, 125] provided a 95% confidence interval, while [57, 22] employed a threshold of 80% for their uncertainty intervals, but not in the article, but in a website that supports their paper.

Some singular practices were found, For example [149], where predictions were made with $\pm 50\%$ of the predicted value, and some of the charts depicted an interval which level of confidence was not defined. On the other hand, [134] that provided the metric values with its mean and its variance, which at least provide some additional information about the fitness of the model. In an attempt to capture uncertainty, in [137, 140] metrics were delivered for different error margins (from 0.05 to 0.5, in steps of 0.05).

Finally, [67, 69, 41, 122, 42, 38] did not mention any numerical indication for the confidence threshold. This practice in particular, together with the disuse of any confidence intervals at all, makes more difficult to interpret the uncertainty in their predictions, as the estimated probability of capturing the truth is ambiguous. The rest did not employed any kind of confidence interval, or at least, failed to mention it.

As an example, in [55], authors use a JHU (CSSE) dataset, containing information on infected, recovered, and deceased patients in 406 locations over 51 days and split in 80/20 randomly selected data points for training and testing, respectively. This dataset was used in training a regressive MLP to achieve a worldwide model of the maximal number of cases across all locations per time unit. Exhaustive hyper-parameter combinations were tested, finding that the best architecture was 4-4-4-4, with ReLu (chosen over identity, logistic, tanh) with an LBFGS (chosen over Adam) solver, keeping the best settings of the other parameters for each variable. R2 scores were of 0.98599 for the confirmed patient model, 0.97941 for

the recovered patient model, and 0.99429 for the deceased patient model. Models also showed poor tracking of sudden and unexpected changes, and a drop in R2 scores with cross-validation. They concluded that this suggests low robustness of the model, especially for the recovered patients.

4.5 Final score

As can be seen in Figure 4.5, only 35 of the 96 studied papers scores in at least half of the fields. The best score was 16 over 17, obtained by [152], only failing in the statistical inference and the accessibility fields. It is followed by [22, 43], both with a positive score in 12 of the fields, and both failing in accessibility. But [22] provides information about how missing data was handled in an implicit way, while [43] totally fails in managing statistical inference and in providing information about the software.

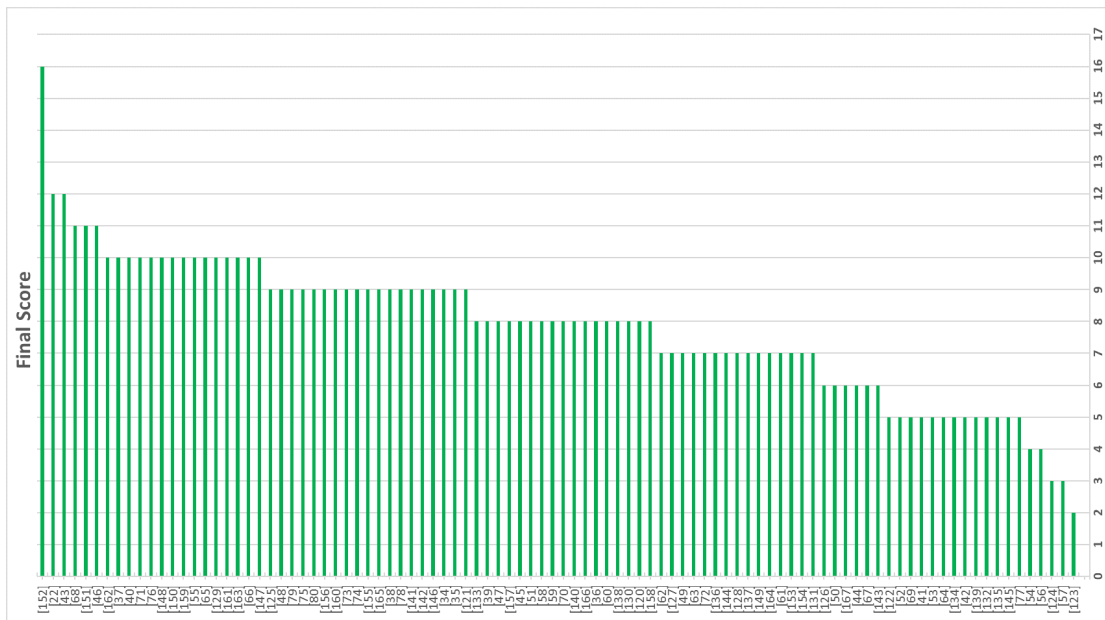


Figure 4.5: Final Scores.

On the other side of the coin, [123] is scoring only in the data source and features description, while the object of study is only available implicitly throughout the whole document. Scores partially only in benchmark comparison and in objective function and optimizer, while failing in the rest of the fields.

Ref	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Final Score
152	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	16
222	Y	Y	Y	Y	Y	Y+	Y	Y+	N	Y+	Y	Y	Y	Y+	Y	Y	Y	12
43	Y	Y	Y	Y	Y	Y	Y	N	N	Y+	Y	Y	Y	Y	Y	Y+	N	12
68	Y	Y	Y	Y	Y	Y	Y	Y+	N	N	Y	Y	Y	Y+	Y	Y+	N	11
151	Y	Y	Y	Y	Y	Y	Y	N	N	Y+	Y	Y	Y	Y+	Y	Y+	N	11
46	Y	Y	Y	Y	Y	N	N	Y	N	N	Y	Y+	Y+	Y	Y	Y	Y	11
162	Y+	Y	Y	Y	Y	N	Y	Y	Y	Y+	Y	Y+	Y	Y+	Y	Y	Y	10
37	Y+	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y+	Y+	Y	Y	Y	N	10
40	Y	Y	Y	Y	Y	N	Y	Y+	N	N	Y	Y	Y	Y+	Y	Y+	N	10
71	Y	Y	Y	Y	Y	Y	N	Y+	N	N	Y	Y	Y+	Y	Y	Y+	N	10
76	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	Y+	Y+	Y+	Y	Y	Y	10
148	Y+	Y	Y	Y	Y	Y	Y	Y+	N	N	Y	Y	Y	N	Y	Y+	N	10
150	Y	Y	Y	Y	Y	N	Y	Y+	N	N	Y	Y	Y	Y+	Y	Y+	N	10
159	Y	Y	Y	Y	Y	N	Y+	Y	N	N	Y	Y+	Y	Y	Y	Y+	N	10
55	Y	N	Y	Y	Y	N	N	Y	Y	Y+	Y	Y	Y+	Y	Y	N	N	10
65	Y	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y+	Y	Y+	Y	10
129	Y	N	Y	Y	Y	N	Y	N	N	N	Y	Y	Y+	Y+	Y	Y	Y	10
161	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	N	Y	Y+	Y	Y+	N	10
163	Y	Y	Y	Y	Y	N	Y	Y+	N	N	Y	Y	Y	Y+	Y	Y	N	10
66	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y+	Y	N	N	10
147	Y	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	N	Y	Y	N	10
125	Y+	Y	Y	Y	Y	Y	Y+	Y+	N	Y	Y	Y	Y+	Y+	Y	Y+	N	9
48	Y+	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y+	Y+	Y	Y+	N	9
79	Y+	Y	Y	Y	Y	N	Y	Y+	N	Y	Y	Y+	Y	Y+	Y	Y+	N	9
75	Y	Y	Y	Y+	Y	N	Y+	Y+	N	N	Y	Y	Y	Y+	Y	Y	N	9
80	Y+	N	Y	Y	Y	N	N	Y+	Y	Y+	Y	Y	Y	Y	Y	Y+	N	9
156	Y	Y	Y	Y	Y	N	N	Y	Y	Y+	Y	Y+	N	Y+	Y	Y	N	9
160	Y	Y	Y	Y	Y	N	Y+	Y+	N	N	Y	Y	Y	Y+	Y	Y+	N	9
73	Y	N	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y+	Y+	Y	Y	N	9
74	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y+	N	Y+	Y	Y+	N	9
155	Y	Y	Y	Y	Y	N	N	Y	Y+	N	N	N	N	Y+	Y	Y	Y	9
165	Y+	Y	Y	Y	Y	N	Y	Y+	N	N	Y	Y	Y	N	Y	Y+	N	9
38	Y+	Y	Y	Y	Y	N	Y	N	N	Y	N	N	Y+	N	Y	Y	N	9
78	Y+	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	N	Y+	Y	Y	N	9
141	Y	Y	Y	Y+	Y	Y	Y	N	N	N	N	N	N	Y	Y	Y	Y	9
142	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	Y	Y	Y+	Y	N	N	9
146	Y	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y+	Y+	Y	Y	N	9
34	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	Y+	Y	N	Y	N	N	9
35	Y	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	N	Y	Y+	N	9
121	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	N	N	Y+	Y	Y	N	8
133	Y	N	Y	Y+	Y	N	Y	Y+	N	Y	Y	Y+	Y+	Y+	Y	Y	N	8
39	Y	Y	Y	Y+	Y	N	Y	N	N	N	Y	Y+	Y	Y+	Y	Y+	N	8
47	Y+	Y	Y	Y	Y	N	Y	Y+	N	N	Y	Y+	Y+	N	Y	Y	N	8
157	Y+	Y	Y	Y	Y	N	Y	Y+	N	N	Y	Y+	Y	Y+	Y	Y	N	8
45	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	N	Y+	Y+	Y	Y	N	8
51	Y+	Y	Y	Y	Y	N	N	N	N	Y+	Y	Y	Y	Y+	Y	Y	N	8
58	Y+	Y	Y	Y	N	N	Y	Y+	N	N	Y	Y	Y	N	Y	Y	N	8
59	Y+	Y	Y	Y+	Y	N	Y	N	N	N	Y	Y	Y+	N	Y	Y	N	8
70	Y+	N	Y	Y	Y	N	Y	Y+	N	N	Y	Y	N	Y+	Y	Y	N	8
140	Y	Y	Y	Y+	N	N	Y	N	N	N	Y	Y	Y+	Y+	Y	Y	N	8
166	Y	Y	Y	Y	Y	N	N	Y	N	N	Y	Y+	N	Y	Y	Y	N	8
36	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	Y+	Y	N	N	Y	N	8
60	Y	N	Y	Y	Y	N	Y	N	N	N	Y	N	Y	Y+	Y	Y	N	8
138	Y	Y	Y	Y+	N	N	Y	N	N	N	Y	Y	Y	N	Y	Y+	N	8
130	Y	N	Y	Y	N	N	Y	Y	N	N	N	N	N	Y+	Y	Y	N	8
120	Y	Y	N	Y	Y	N	N	Y	N	Y	Y	Y	N	Y+	Y	N	N	8
158	Y+	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	N	Y	N	N	8
62	Y+	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y+	Y+	Y+	N	N	N	7
127	Y+	N	Y	Y	Y	N	Y	Y+	N	N	Y	Y+	Y+	Y	Y	Y+	N	7
49	Y	Y	Y	Y	Y	N	N	N	N	N	Y	Y+	Y+	Y+	Y	Y+	N	7
63	Y+	Y	Y	Y+	Y	N	N	N	N	Y+	N	Y+	N	Y	Y	Y	N	7
72	Y+	Y	Y	Y	N	N	Y	Y+	Y+	N	N	Y	Y	N	Y	Y+	N	7
136	Y+	Y	N	Y	Y	Y+	N	N	N	N	Y	Y	Y	Y	Y	Y+	N	7
144	Y+	Y	Y	Y+	Y	N	Y	N	N	N	Y	Y	N	Y+	Y	Y	N	7
128	Y	N	Y	Y+	Y	N	Y	N	N	N	Y	Y+	Y	Y+	N	Y	N	7
137	Y	Y	Y	N	N	N	Y	N	N	N	Y	Y+	Y+	Y+	Y	Y	N	7
149	Y	Y	Y	Y	Y	Y	Y	Y+	N	N	N	N	Y	Y+	N	Y	Y	7
164	Y	Y	Y	Y	Y	N	Y	Y+	N	N	N	N	Y+	Y+	Y	N	N	7
61	Y	Y	Y	Y+	N	Y	N	Y	N	N	Y	N	N	N	Y	Y	N	7
153	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	Y	Y+	Y+	Y	N	7
154	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	Y+	Y+	Y	Y	7
131	Y	N	Y	Y	Y	N	Y	N	N	N	Y	Y	N	Y+	N	N	N	7
126	Y+	Y	Y	Y	Y	N	N	Y	N	N	Y	N	Y	Y	Y	Y	N	6
50	Y+	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y	Y	Y	N	6
167	Y+	Y	Y	Y	Y+	Y	N	Y	N	N	N	N	Y	N	Y	Y	N	6
44	Y	N	Y	Y	Y	N	N	N	N	N	Y	Y+	N	Y+	Y	Y+	N	6
67	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	Y+	N	Y	6
143	Y	Y	Y	Y+	Y	N	Y	N	N	N	N	N	N	N	N	Y	N	6
122	Y+	N	N	Y	Y	N	Y	N	Y	Y+	Y	Y	Y	Y	Y	Y	N	5
52	Y+	N	Y	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	Y	Y	N	5
69	Y+	N	Y	Y	Y	N	N	Y	N	N	Y	N	Y	Y	Y	Y	N	5
41	Y+	Y	Y	Y	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y	N	5
53	Y+	N	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	N	N	Y	N	5
64	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N	Y	Y	N	Y	N	5
134	Y+	N	Y	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	N	Y	N	5
42	Y	Y	Y	Y	Y	N	Y	Y	N	N	N	N	N	Y	Y	Y	N	5
139	Y+	Y	Y	Y	Y	N	N	N	N	N	N	Y	Y	N	Y	Y	N	5
132	Y	N	Y	Y	Y	N	Y	Y	N	N	N	N	N	Y	Y	Y	N	5
135	Y	N	Y	N	N	N	N	Y	N	Y	N	N	N	N	Y	Y	N	5
145	Y	Y	Y	Y	N	N	N	N	N	N	Y	Y	N	N	N	Y	N	5
77	Y	Y	N	N	N	N	N	Y	N	N	Y	N	N	N	N	N	N	5
54	Y	N	N	Y	Y	N	N	N	N	N	Y	N	N	N	Y	Y	N	4
56	Y	N	Y	Y	Y	N	N	N	N	N	N	N	Y	Y	N	N	N	4
124	Y+	N	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	N	Y	N	3
57	Y+	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	3
123	Y+	N	Y	Y	N	N	N	N	N	N	N	N	Y	N	N	Y	N	2

Figure 4.6: Summary of scores that papers received in each criteria (the column titles corresponds to the item numbers used in Section 3.2).

Chapter 5

Discussion

It must be admitted that a significant part of the papers mentioned here are pre-prints and they were not subject yet to a formal peer review. However, due to the fast evolution of the outbreak, and the huge impact in the world's population, it has been challenging to get all of them reviewed and approved on time. So, in order to not miss any potential valid work, it was decided to include them in our work.

The outstanding efforts seen in the reviewed literature to model and forecast the COVID-19 pandemic using deep learning techniques result obvious. Nevertheless, it has been also patent the large number of methodological and reporting deficiencies, even flaws, ending in none of the papers fully scoring in all of the proposed criteria. The study with more criteria fulfilled, without reservations, is [152] with one single fail. It's followed by [22] with one single fail but with four criteria partial or implicitly met, and then [125, 152, 162] failing in only two of them. Despite [43] reaches a score of 12, it is totally failing in 3 of the criteria. With only 35 of the reviewed literature meeting at least 50% of the criteria, and only 3 of those [152, 22, 43] scoring above 70%, it can be asserted that the overall robustness and replicability of the evaluated papers is very poor.

As can be seen in Table 4.1, and with more detail in Figure 4.6, the most common weaknesses are scarce application of statistical inference in 88 articles; missing or poor definition of the experiments, again in 88 of the papers; lack of missing-data handling in 80 of the works; no model initialization details in 76 studies; no data pre-processing in 45 of them; no software information revealed in 43 of the articles or just a poor one in 35 more. These issues may lead to excessively enthusiastic performance estimations and reduced replicability.

5.1 Recommendations for Problem Description

Sometimes it is not clear what the target of the paper, due to the use of ambiguous terms or incomplete assertions, such like "predict COVID-19 infection" or "forecast COVID-19 outbreak", avoiding to be more specific in the objectives.

The enunciation of a forecasting problem, in contrast with other type of AI endeavours, should not be necessarily a difficult task. It might be enough to state the variables to predict (i.e., 'number of COVID-19 confirmed cases'), the region, area or community where the predictions are being made (i.e., 'China', 'Emilia-Romagna' or 'Hospital Albert Einstein'), the forecasting horizon (i.e., 'in the next ten days') and the model employed (i.e., 'An stacked LSTM model'). So, a simple and clear statements explicitly justifying these factors should suffice.

5.2 Recommendations for Data

The small volumes of COVID-19 data, the different dataset sizes, the diverse kind of variables collected within the datasets and the way this data is collected by the different organizations and governments remains a huge challenge for an accurate model comparison. We agree with [7] in suggesting that the use of big collaboratively and high-quality datasets provided by governments and healthcare organizations (i.e., WHO, CDC, JHU, etc.) may help to overcome this issue. The surveillance on the quality of the aggregated data by renowned organizations can help to avoid 'retrodden' datasets and may reduce over-fitting, derived from the fact that the community is focused on outperforming benchmarks on a single public dataset.

However, not explicitly stating the sources of the data or making them accessible, not providing a clear description of the variables contained in the dataset, as well as the data intervals considered, or the decisions taken about missing data or the pre-processing stages must be considered a serious flaw.

5.3 Recommendations for Model Description

Nowadays, in research environments in which open science is becoming more and more encouraged, and for the sake of interpretability and replicability, it is strongly recommended to reveal as much details from the model as possible, so the experiments can be reproduced, and models can be compared to future research.

The disclosure of the software packages, frameworks and libraries employed, as well as its versions, can improve the understanding of the performance and conclusions derived from the experiment, while it is enabling the replicability.

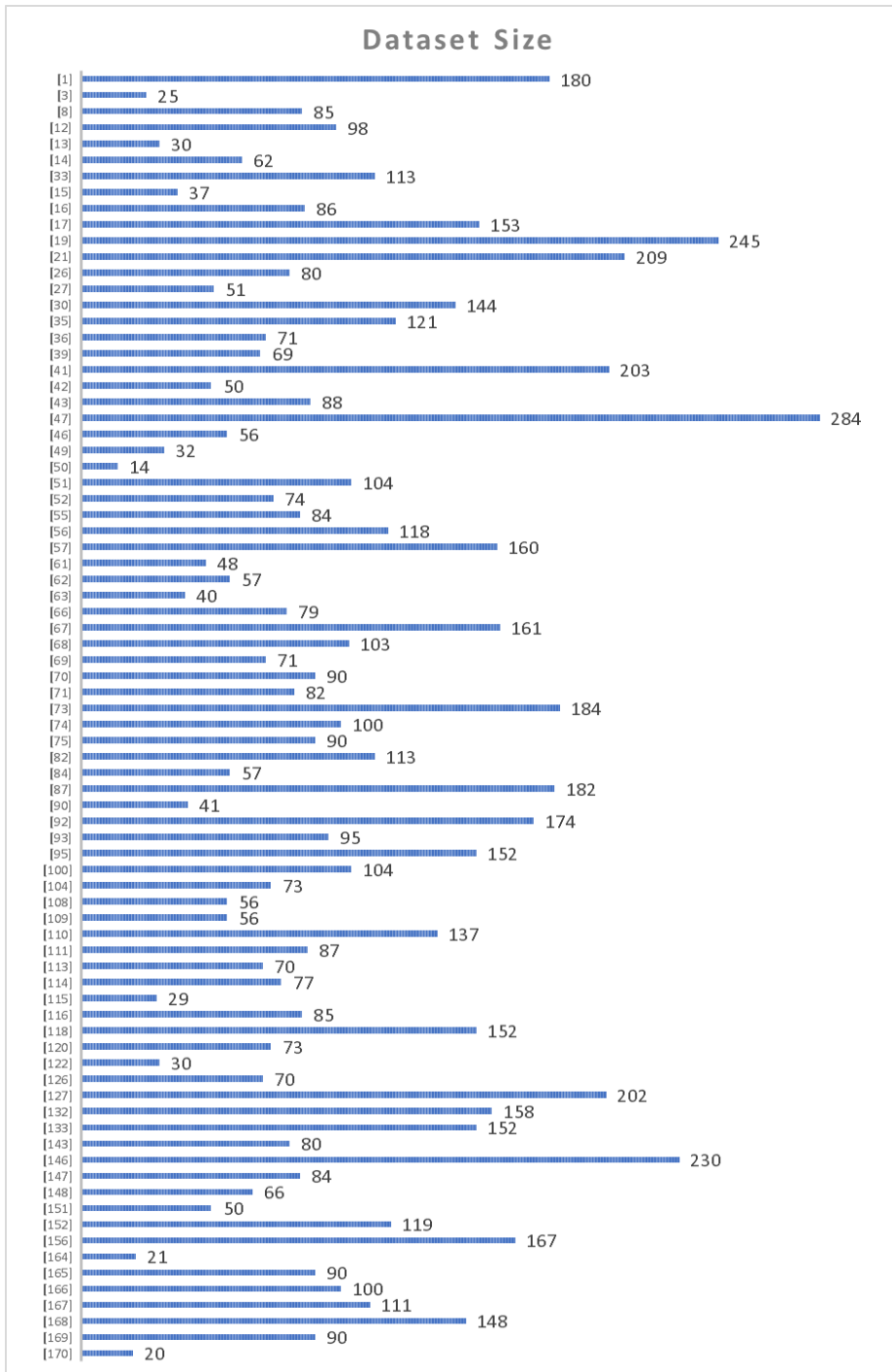


Figure 5.1: Dataset size.

In a similar way, revealing the number of layers, number of units and the activation functions, together with the objective and optimizer function, becomes essential to understand the developed model and its eventual advantages and drawbacks.

Another potential source of error is the randomization of the weights [138], being one of the main sources of stochasticity of the model. Unveiling the distribution from which the weights are being initialised, as well as the employed seeds, becomes essential for reproducing the investigation.

Counting with the source code and the original dataset employed enhances the comprehension of the model itself. In such sense, making the complete study available in a public repository is a practice that boost the progress of science, especially in these challenging times.

5.4 Recommendations for Evaluation

There is no unique appropriate metric for model errors. Using RMSE, leads to large errors having a relatively greater influence on the total compared to the smaller ones [170]. This makes MAE better for discriminating among models. Despite its robustness against outliers, MAE is more sensitive to variance, fluctuating its value between several errors sets with the same RMSE [171].

RMSE might be selected to minimize cost function, because it helps to calculate the gradient of absolute errors. With a low number of samples (i.e. 100), giving the values of the errors themselves is probably better than any statistics. Otherwise large outliers might be excluded from the RMSE calculation [171]. But when having more samples, RMSE can reconstruct error distribution, with a standard deviation lower than 5 %. Inconsistency in comparing RMSEs from different studies is not due to error-scale variance alone [170].

Choosing one single metric, removes a lot of information, so an error distribution should be provided. MAE is suitable for uniformly distributed errors, while RMSE is better when errors follow a normal distribution, which is the most common case. For other kinds of distributions, more statistical, such as mean, variance, skewness, and flatness, should be provided [171]. So to better depict the model behaviour, the best recommendation might be to provide several metrics on the studies.

When reporting results, including a statistical significance test with the p-value obtained (rather than just simply passing or not the famous 0.05 value) and/or confidence intervals to reflect the uncertainty in the forecast is strongly recommended for avoiding bias. But also, when benchmarking, better use simple, but manageable models as a baseline, such like naïve, instead of fancy sophisticated models that are hard to optimize. It is very common to see how the interest that

has been put in developing the desired model is inversely proportional to the effort invested in the complex benchmark one. This may lead to overoptimistic interpretations of the results, as well as an unrealistic idea of the real capabilities of the developed model.

Monte-Carlo stochastic simulation seems to be a suitable practice for modelling infectious outbreaks that changes across geographical areas and through time [37]. Also hyper-parameter search and sensitivity tests are strongly recommended.

5.5 Recommendations for Replicability and Reproducibility

According to the American Statistical Association (ASA)[173], following concepts must be distinguished:

- **Reproducibility:** A study is reproducible if you can take the original data and the computer code used to analyse the data and reproduce all of the numerical findings from the study.
- **Replicability:** This is the act of repeating an entire study, independently of the original investigator without the use of original data (but generally using the same methods).

Although it must be admitted that full replicability is theoretically not achievable, a clear description of the methods, models, materials, procedures, metrics, and other variables involved in the study would facilitate it. A clear description of the dataset, data pre-processing and missing data handling is essential. A description of the statistical inference decisions made and whether the study is exploratory or confirmatory, as well as discussion of the expected constraints for generality, uncertainty of the measurements, results, and inferences are definitely helpful.

The easiest way to replicate an experiment is to count with the source code and the original dataset employed. But a potential opacity might occur when publicly available datasets or code are being updated. Therefore, it is advisable to keep track of specific cached versions of datasets and code, so those ones can be correctly referenced. Many public repository sites are providing tools to make this task much easier. These practices are also enabling scientific reproducibility, speeding up future discoveries in any discipline.

Chapter 6

Conclusions

In this systematic review, current deep learning literature for COVID-19 prediction and forecast has been considered. The focus has been put on evaluating a set of papers, underlining the quality flaws of the methods employed and the reproducibility and replicability issues.

After establishing a set of minimum quality indicators, it has been observed that no papers in the reviewed literature currently have documented satisfactorily the methodologies employed for the entire process, failing to follow good practices for developing a reproducible deep learning model.

A robust cross-validation methodology is a pending subject throughout the literature, as well as a deeper and standardized usage of the metrics. There is a lot of room for improvement in model comparison against naïve baselines, as well as the extended use of any kind of statistical inference, to minimally discard any possibility of changes in the results.

Also, some specific recommendations to the researchers for better practices in all analysed criteria has been provided. The different kinds of error metrics delivered by the researchers in the analyzed papers, the variety of forecast periods and the different kinds of variables to predict, makes any comparison a difficult task.

We agree with [19] that it is vital to develop a standardized reporting protocol and checklists to reduce the poorly conducted COVID-19 studies in favour of more properly conducted studies, and to improve replicability.

The pandemic highlighted the importance of communicating openly with other researchers and members of the public. This was also seen as a key step in reducing the chances of data being misappropriated. The community can help address these challenges by developing policies and initiatives that support the dissemination of data, improve data access, and increase diversity in the field of research.

Scholarly journals provide a set of control rights to the authors while enabling the knowledge transfer across researchers. Non-peer reviewed recognized com-

munity pre-print servers also allow the authors to select from several distribution/reuse options for making manuscripts available. Due to this, and together with the spread use of open source licenses nowadays, this excessive secrecy is not justified at all.

Bibliography

- [1] A. Adiga, D. Dubhashi, B. Lewis, M. Marathe, S. Venkatramanan, and A. Vullikanti, “Mathematical models for covid-19 pandemic: A comparative analysis,” *Journal of the Indian Institute of Science*, vol. 100, pp. 793–807, 10 2020. Company: Springer Distributor: Springer Institution: Springer Label: Springer number: 4 publisher: Springer India.
- [2] M. Castro, S. Ares, J. A. Cuesta, and S. Manrubia, “The turning point and end of an expanding epidemic cannot be precisely forecast,” *Proceedings of the National Academy of Sciences*, vol. 117, pp. 26190–26196, 10 2020. publisher: National Academy of Sciences section: Biological Sciences PMID: 33004629.
- [3] D. Lazer, R. Kennedy, G. King, and A. Vespignani, “The parable of google flu: Traps in big data analysis,” *Science*, vol. 343, pp. 1203–1205, 3 2014.
- [4] J. H. Chen and S. M. Asch, “Machine learning and prediction in medicine — beyond the peak of inflated expectations,” *The New England journal of medicine*, vol. 376, pp. 2507–2509, 6 2017. PMID: 28657867 PMCID: PMC5953825.
- [5] W. Naudé, “Artificial intelligence against covid-19: An early review,” tech. rep., Social Science Research Network, Rochester, NY, 4 2020. [Online; accessed 2021-01-24].
- [6] D. C. Nguyen, M. Dinh, P. N. Pathirana, and A. Seneviratne, “Blockchain and ai-based solutions to combat coronavirus (covid-19)-like epidemics: A survey,” *Preprint*, 4 2020. publisher: Preprints.
- [7] Q. Pham, D. C. Nguyen, T. Huynh-The, W. Hwang, and P. N. Pathirana, “Artificial intelligence (ai) and big data for coronavirus (covid-19) pandemic: A survey on the state-of-the-arts,” *IEEE Access*, vol. 8, pp. 130820–130839, 4 2020. event: IEEE Access.
- [8] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini, and J. Wu, “How big data and artificial intelligence can help better manage the covid-19 pandemic,” *International Journal of Environmental Research and Public Health*, vol. 17, 5 2020. PMID: 32370204 PMCID: PMC7246824.
- [9] P. N. Mahalle, N. P. Sable, N. P. Mahalle, and G. R. Shinde, “Data analytics: Covid-19 prediction using multimodal data,” *Preprint*, 5 2020. publisher: Preprints.
- [10] T. Alamo, D. G. Reina, and P. Millán, “Data-driven methods to monitor, model, forecast and control covid-19 pandemic: Leveraging data science, epidemiology and control theory,” *arXiv:2006.01731 [physics, q-bio]*, 6 2020. arXiv: 2006.01731.
- [11] G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, and A. E. Hassanien, “Forecasting models for coronavirus disease (covid-19): A survey of the state-of-the-art,” *SN Computer Science*, vol. 1, p. 197, 6 2020.

- [12] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial intelligence (ai) applications for covid-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research Reviews*, vol. 14, pp. 337–339, 7 2020.
- [13] J. Chen, K. Li, Z. Zhang, K. Li, and P. S. Yu, "A survey on applications of artificial intelligence in fighting against covid-19," *arXiv:2007.02202 [cs, q-bio]*, 7 2020. arXiv: 2007.02202.
- [14] A. Sufian, A. Ghosh, A. S. Sadiq, and F. Smarandache, "A survey on deep transfer learning to edge computing for mitigating the covid-19 pandemic," *Journal of Systems Architecture*, vol. 108, p. 101830, 9 2020. PMID: null PMID: PMC7326453.
- [15] W. Naudé, "Artificial intelligence vs covid-19: limitations, constraints and pitfalls," *AISO-CIETY*, vol. 35, pp. 761–765, 9 2020.
- [16] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: A review," *Chaos, Solitons & Fractals*, vol. 139, p. 110059, 10 2020.
- [17] P. Sarosh, S. A. Parah, R. F. Mansur, and G. M. Bhat, "Artificial intelligence for covid-19 detection – a state-of-the-art review," *arXiv:2012.06310 [cs]*, 11 2020. arXiv: 2012.06310.
- [18] J. Rasheed, A. Jamil, A. A. Hameed, U. Aftab, J. Aftab, S. A. Shah, and D. Draheim, "A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the covid-19 pandemic," *Chaos, Solitons & Fractals*, vol. 141, p. 110337, 12 2020.
- [19] A. Abd-Alrazaq, M. Alajlani, D. Alhuwail, J. Schneider, S. Al-Kuwari, Z. Shah, M. Hamdi, and M. Househ, "Artificial intelligence in the fight against covid-19: Scoping review," *Journal of Medical Internet Research*, vol. 22, p. e20756, 12 2020. PMID: 33284779 PMID: PMC7744141.
- [20] M.-H. Tayarani N., "Applications of artificial intelligence in battling against covid-19: A literature review," *Chaos, Solitons & Fractals*, vol. 142, p. 110338, 1 2021.
- [21] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against covid-19," *arXiv:2003.11336 [cs]*, 1 2021. arXiv: 2003.11336.
- [22] J. Devaraj, R. Madurai Elavarasan, R. Pugazhendhi, G. M. Shafullah, S. Ganesan, A. Kaarthic Jeysree, I. Ahmad Khan, and E. Hossain, "Forecasting of covid-19 cases using deep learning models: Is it reliable and practically significant?," *Results in Physics*, p. 103817, 1 2021.
- [23] A. Alimadadi, S. Aryal, I. Manandhar, P. B. Munroe, B. Joe, and X. Cheng, "Artificial intelligence and machine learning to fight covid-19," *Physiological Genomics*, vol. 52, pp. 200–202, 3 2020. publisher: American Physiological Society.
- [24] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE Reviews in Biomedical Engineering*, 2020. publisher: Institute of Electrical and Electronics Engineers Inc.
- [25] Y. Mohamadou, A. Halidou, and P. T. Kapen, "A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of covid-19," *Appl Intell*, 2020. [Online; accessed 2021-01-30].

- [26] A. Kumar, P. K. Gupta, and A. Srivastava, "A review of modern technologies for tackling covid-19 pandemic," *DiabetesMetabolic Syndrome*, vol. 14, pp. 569–573, 8 2020. PMID: 32413821 PMCID: PMC7204706.
- [27] M. Tsikala Vafea, E. Atalla, J. Georgakas, F. Shehadeh, E. K. Mylona, M. Kalligeros, and E. Mylonakis, "Emerging technologies for use in the study, diagnosis, and treatment of patients with covid-19," *Cellular and Molecular Bioengineering*, vol. 13, pp. 249–257, 8 2020.
- [28] A. S. Ahuja, V. P. Reddy, and O. Marques, "Artificial intelligence and covid-19: A multi-disciplinary approach," *Integrative Medicine Research*, vol. 9, p. 100434, 9 2020.
- [29] A. Shoeibi, M. Khodatars, R. Alizadehsani, N. Ghassemi, M. Jafari, P. Moridian, A. Khadem, D. Sadeghi, S. Hussain, A. Zare, Z. A. Sani, J. Bazeli, F. Khozeimeh, A. Khosravi, S. Nahavandi, U. R. Acharya, and P. Shi, "Automated detection and forecasting of covid-19 using deep learning techniques: A review," *arXiv:2007.10785 [cs, eess]*, 7 2020. arXiv: 2007.10785.
- [30] R. Madurai Elavarasan and R. Pugazhendhi, "Restructured society and environment: A review on potential technological strategies to control the covid-19 pandemic," *Science of The Total Environment*, vol. 725, p. 138858, 7 2020.
- [31] A. Waleed Salehi, P. Baglat, and G. Gupta, "Review on machine and deep learning models for the detection and prediction of coronavirus," *Materials Today. Proceedings*, vol. 33, pp. 3896–3901, 2020. PMID: 32837918 PMCID: PMC7309744.
- [32] S. Latif, M. Usman, S. Manzoor, W. Iqbal, J. Qadir, G. Tyson, I. Castro, A. Razi, M. N. K. Boulos, A. Weller, and J. Crowcroft, "Leveraging data science to combat covid-19: A comprehensive review," *IEEE Transactions on Artificial Intelligence*, vol. 1, pp. 85–103, 8 2020. event: IEEE Transactions on Artificial Intelligence.
- [33] L. Wynants, B. Van Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, M. M. J. Bonten, D. L. Dahly, J. A. Damen, T. P. A. Debray, V. M. T. de Jong, M. De Vos, P. Dhimman, M. C. Haller, M. O. Harhay, L. Henckaerts, P. Heus, M. Kammer, N. Kreuzberger, A. Lohmann, K. Luijken, J. Ma, G. P. Martin, D. J. McLernon, C. L. Andaur Navarro, J. B. Reitsma, J. C. Sergeant, C. Shi, N. Skoetz, L. J. M. Smits, K. I. E. Snell, M. Sperrin, R. Spijker, E. W. Steyerberg, T. Takada, I. Tzoulaki, S. M. J. van Kuijk, B. C. T. van Bussel, I. C. C. van der Horst, F. S. van Royen, J. Y. Verbakel, C. Wallisch, J. Wilkinson, R. Wolff, L. Hooft, K. G. M. Moons, and M. van Smeden, "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *BMJ*, vol. 369, 2020. [Online; accessed 2021-01-30].
- [34] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, "Artificial intelligence forecasting of covid-19 in china," *Preprint*, 2 2020. [Online; accessed 2020-12-15].
- [35] Z. Hu, Q. Ge, S. Li, E. Boerwinkle, L. Jin, and M. Xiong, "Forecasting and evaluating intervention of covid-19 in the world," *arXiv:2003.09800 [q-bio]*, 3 2020. arXiv: 2003.09800.
- [36] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Luo, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, and J. He, "Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions," *Journal of Thoracic Disease*, vol. 12, pp. 165–174, 3 2020. PMID: 32274081 PMCID: PMC7139011.

- [37] S. J. Fong, G. Li, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction," *Applied Soft Computing*, vol. 93, p. 106282, 8 2020.
- [38] S. J. Fong, G. Li, N. Dey, R. González-Crespo, and E. Herrera-Viedma, "Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. Special Issue on Soft Computing, 2020. [Online; accessed 2020-12-15].
- [39] 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:2004.05960 [cs, eess]*, 4 2020. arXiv: 2004.05960.
- [40] O. Torrealba-Rodriguez, R. A. Conde-Gutiérrez, and A. L. Hernández-Javier, "Modeling and prediction of covid-19 in mexico applying mathematical and computational models," *Chaos, Solitons & Fractals*, vol. 138, p. 109946, 9 2020.
- [41] V. K. R. Chimmula and L. Zhang, "Time series forecasting of covid-19 transmission in canada using lstm networks," *Chaos, Solitons & Fractals*, vol. 135, p. 109864, 6 2020.
- [42] A. Tomar and N. Gupta, "Prediction for the spread of covid-19 in india and effectiveness of preventive measures," *Science of The Total Environment*, vol. 728, p. 138762, 8 2020.
- [43] N. Zheng, S. Du, J. Wang, H. Zhang, W. Cui, Z. Kang, T. Yang, B. Lou, Y. Chi, H. Long, M. Ma, Q. Yuan, S. Zhang, D. Zhang, F. Ye, and J. Xin, "Predicting covid-19 in china using hybrid ai model," *IEEE Transactions on Cybernetics*, 2020.
- [44] C.-J. Huang, Y.-H. Chen, Y. Ma, and P.-H. Kuo, "Multiple-input deep convolutional neural network model for covid-19 forecasting in china," *medRxiv*, p. 2020.03.23.20041608, 3 2020. publisher: Cold Spring Harbor Laboratory Press.
- [45] M. A. A. Al-qaness, A. A. Ewees, H. Fan, and M. Abd El Aziz, "Optimization method for forecasting confirmed cases of covid-19 in china," *Journal of Clinical Medicine*, vol. 9, p. 674, 3 2020. number: 3 publisher: Multidisciplinary Digital Publishing Institute.
- [46] S. Ayyoubzadeh, S. M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, and S. Rostam Nikan Kalhori, "Predicting covid-19 incidence through analysis of google trends data in iran: Data mining and deep learning pilot study," *JMIR Public Health and Surveillance*, vol. 6, 3 2020.
- [47] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Covid-19 epidemic analysis using machine learning and deep learning algorithms," *medRxiv*, p. 2020.04.08.20057679, 6 2020. publisher: Cold Spring Harbor Laboratory Press.
- [48] S. k. Paul, S. Jana, and P. Bhaumik, "A multivariate spatiotemporal spread model of covid-19 using ensemble of convlstm networks," *medRxiv*, 2020.
- [49] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting covid-19 time-series data: A comparative study," *Chaos, Solitons & Fractals*, vol. 140, p. 110121, 11 2020.
- [50] R. Dandekar and G. Barbastathis, "Neural network aided quarantine control model estimation of global covid-19 spread," *arXiv:2004.02752 [physics, q-bio]*, 4 2020. arXiv: 2004.02752.

- [51] M. R. Ibrahim, J. Haworth, A. Lipani, N. Aslam, T. Cheng, and N. Christie, "Variational-lstm autoencoder to forecast the spread of coronavirus across the globe," *medRxiv*, p. 2020.04.20.20070938, 4 2020. publisher: Cold Spring Harbor Laboratory Press.
- [52] R. Kafieh, R. Arian, N. Saeedizadeh, S. Minaee, Z. Amini, S. K. Yadav, A. Vaezi, N. Rezaei, and S. H. Javanmard, "Covid-19 in iran: A deeper look into the future," *medRxiv*, p. 2020.04.24.20078477, 4 2020. publisher: Cold Spring Harbor Laboratory Press.
- [53] L. R. Kolozsvari, T. Berczes, A. Hajdu, R. Gesztelyi, A. Tiba, I. Varga, A. B. Al-Tammemi, G. J. Szollosi, S. Harsanyi, S. Garboczy, and J. Zsuga, "Predicting the epidemic curve of the coronavirus (sars-cov-2) disease (covid-19) using artificial intelligence," *medRxiv*, p. 2020.04.17.20069666, 1 2021. publisher: Cold Spring Harbor Laboratory Press.
- [54] S. Dutta and S. K. Bandyopadhyay, "Machine learning approach for confirmation of covid-19 cases: Positive, negative, death and release," *medRxiv*, p. 2020.03.25.20043505, 3 2020. publisher: Cold Spring Harbor Laboratory Press.
- [55] Z. Car, S. B. Šegota, N. Anđelić, I. Lorencin, and V. Mrzljak, "Modeling the spread of covid-19 infection using a multilayer perceptron," *Computational and Mathematical Methods in Medicine*, vol. 2020, 1 2020. [Online; accessed 2021-02-01].
- [56] P. Hartono, "Similarity maps and pairwise predictions for transmission dynamics of covid-19 with neural networks," *Informatics in Medicine Unlocked*, vol. 20, p. 100386, 1 2020.
- [57] S. Uhlig, K. Nichani, C. Uhlig, and K. Simon, "Modeling projections for covid-19 pandemic by combining epidemiological, statistical, and neural network approaches," *medRxiv*, p. 2020.04.17.20059535, 4 2020. publisher: Cold Spring Harbor Laboratory Press.
- [58] H. Dutta, "Neural network model for prediction of covid-19 confirmed cases and fatalities," *Preprint*, 5 2020.
- [59] B. Yan, X. Tang, J. Wang, Y. Zhou, G. Zheng, Q. Zou, Y. Lu, B. Liu, W. Tu, and N. Xiong, "An improved method for the fitting and prediction of the number of covid-19 confirmed cases based on lstm," *Computers, Materials & Continua*, vol. 64, pp. 1473–1490, 6 2020.
- [60] M. A. A. Al-qaness, A. A. Ewees, H. Fan, L. Abualigah, and M. Abd Elaziz, "Marine predators algorithm for forecasting confirmed cases of covid-19 in italy, usa, iran and korea," *International Journal of Environmental Research and Public Health*, vol. 17, 5 2020. PMID: 32443476 PMID: PMC7277148.
- [61] M. Karimuzzaman, S. Afroz, M. M. Hossain, and A. Rahman, "Forecasting the covid-19 pandemic with climate variables for top five burdening and three south asian countries," *medRxiv*, p. 2020.05.12.20099044, 5 2020. publisher: Cold Spring Harbor Laboratory Press.
- [62] S. Cabras, "A bayesian - deep learning model for estimating covid-19 evolution in spain," *arXiv:2005.10335 [stat]*, 5 2020. arXiv: 2005.10335.
- [63] A. M. Javid, X. Liang, A. Venkitaraman, and S. Chatterjee, "Predictive analysis of covid-19 time-series data from johns hopkins university," *arXiv:2005.05060 [cs, eess]*, 5 2020. arXiv: 2005.05060.
- [64] M. Azarafza, M. Azarafza, and J. Tanha, "Covid-19 infection forecasting based on deep learning in iran," *medRxiv*, p. 2020.05.16.20104182, 5 2020. publisher: Cold Spring Harbor Laboratory Press.

- [65] S. M. Zandavi, T. H. Rashidi, and F. Vafaei, “Forecasting the spread of covid-19 under control scenarios using lstm and dynamic behavioral models,” *arXiv:2005.12270 [physics]*, 5 2020. arXiv: 2005.12270.
- [66] C. Direkoglu and M. Sah, “Worldwide and regional forecasting of coronavirus (covid-19) spread using a deep learning model,” *medRxiv*, p. 2020.05.23.20111039, 5 2020. publisher: Cold Spring Harbor Laboratory Press.
- [67] L.-P. Chen, “Analysis and prediction of covid-19 data in taiwan,” tech. rep., Social Science Research Network, Rochester, NY, 5 2020. DOI: 10.2139/ssrn.3611761.
- [68] A. Chatterjee, M. W. Gerdes, and S. G. Martinez, “Statistical explorations and univariate timeseries analysis on covid-19 datasets to understand the trend of disease spreading and death,” *Sensors (Basel, Switzerland)*, vol. 20, 5 2020. PMID: 32486055 PMCID: PMC7308840.
- [69] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, “Covid-19 pandemic prediction for hungary; a hybrid machine learning approach,” *Mathematics*, vol. 8, no. 6, 2020.
- [70] T. H. H. Aldhyani, M. Alrasheed, A. A. Alqarni, M. Y. Alzahrani, and A. H. Alahmadi, “Deep learning and holt-trend algorithms for predicting covid-19 pandemic,” *medRxiv*, p. 2020.06.03.20121590, 6 2020. publisher: Cold Spring Harbor Laboratory Press.
- [71] R. S. Pontoh, T. Toharudin, S. Zahroh, and E. Supartini, “Effectiveness of the public health measures to prevent the spread of covid-19,” *Commun. Math. Biol. Neurosci.*, vol. 2020, p. Article ID 31, 6 2020. number: 0.
- [72] P. Melin, J. C. Monica, D. Sanchez, and O. Castillo, “Multiple ensemble neural network models with fuzzy response aggregation for predicting covid-19 time series: The case of mexico,” *Healthcare (Basel, Switzerland)*, vol. 8, 6 2020. PMID: 32575622 PMCID: PMC7349072.
- [73] 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:2006.14752 [physics, q-bio]*, 6 2020. arXiv: 2006.14752.
- [74] Y. Tian, I. Luthra, and X. Zhang, “Forecasting covid-19 cases using machine learning models,” *medRxiv*, p. 2020.07.02.20145474, 7 2020. publisher: Cold Spring Harbor Laboratory Press.
- [75] A. Kapoor, X. Ben, L. Liu, B. Perozzi, M. Barnes, M. Blais, and S. O’Banion, “Examining covid-19 forecasting using spatio-temporal graph neural networks,” *arXiv:2007.03113 [cs]*, 7 2020. arXiv: 2007.03113.
- [76] L. Moftakhar, M. Seif, and M. Safe, “Exponentially increasing trend of infected patients with covid-19 in iran: A comparison of neural network and arima forecasting models,” *Iranian Journal of Public Health*, vol. 49, 7 2020.
- [77] S. K. Tamang, P. D. 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, pp. 53–64, 8 2020.
- [78] N. Hasan, “A methodological approach for predicting covid-19 epidemic using eemd-ann hybrid model,” *Internet of Things*, vol. 11, p. 100228, 9 2020.

- [79] R. G. da Silva, M. H. D. M. Ribeiro, V. C. Mariani, and L. d. S. Coelho, “Forecasting brazilian and american covid-19 cases based on artificial intelligence coupled with climatic exogenous variables,” *Chaos, Solitons, and Fractals*, vol. 139, p. 110027, 10 2020. PMID: 32834591 PMCID: PMC7324930.
- [80] T. B. Alakus and I. Turkoglu, “Comparison of deep learning approaches to predict covid-19 infection,” *Chaos, Solitons & Fractals*, vol. 140, p. 110120, 11 2020.
- [81] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J. R. Weir-McCall, Z. Teng, E. Gkrania-Klotsas, J. H. F. Rudd, E. Sala, and C.-B. Schönlieb, “Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans,” *Nature Machine Intelligence*, vol. 3, pp. 199–217, 3 2021. number: 3 publisher: Nature Publishing Group.
- [82] “Strategies for the surveillance of covid-19,” 4 2020. Available at <https://www.ecdc.europa.eu/en/publications-data/strategies-surveillance-covid-19> [Online; accessed 2021-02-28].
- [83] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 12 1943. Company: Springer Distributor: Springer Institution: Springer Label: Springer number: 4 publisher: Kluwer Academic Publishers.
- [84] J. Patterson and A. Gibson, *Deep Learning*. O’Reilly Media, Inc., 8 2017.
- [85] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain,” *Psychological Review*, vol. 65, pp. 386–408, 11 1958. PMID: 13602029.
- [86] F. Rosenblatt, “Principles of neurodynamics. perceptrons and the theory of brain mechanisms,” *The American Journal of Psychology*, vol. 76, no. 4, pp. 705–707, 1963.
- [87] B. Curry and P. H. Morgan, “Neural networks, linear functions and neglected non-linearity,” *Computational Management Science*, vol. 1, pp. 15–29, 12 2003.
- [88] Y. Lecun, *PhD thesis: Modeles connexionnistes de l’apprentissage (connectionist learning models)*. Universite P. et M. Curie (Paris 6), June 1987.
- [89] D. Ballard, “Modular learning in neural networks,” *AAAI 87*, pp. 279–284, 1987.
- [90] P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, 1974.
- [91] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, pp. 533–536, 10 1986. number: 6088 publisher: Nature Publishing Group.
- [92] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, pp. 1735–80, 12 1997.
- [93] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” *arXiv:1406.1078 [cs, stat]*, 9 2014. arXiv: 1406.1078.
- [94] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol. 45, pp. 2673–2681, 11 1997. event: IEEE Transactions on Signal Processing.

- [95] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv:1706.03762 [cs]*, 12 2017. arXiv: 1706.03762 version: 5.
- [96] Y. Lecun, "Generalization and network design strategies," *Connectionism in perspective*, 1989. publisher: Elsevier.
- [97] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "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, pp. 985–990 vol.2, 2004.
- [98] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 7 2006. publisher: American Association for the Advancement of Science section: Report PMID: 16873662.
- [99] S. H. Park and K. Han, "Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction," *Radiology*, vol. 286, pp. 800–809, 1 2018. publisher: Radiological Society of North America.
- [100] S. H. Park and H. Y. Kressel, "Connecting technological innovation in artificial intelligence to real-world medical practice through rigorous clinical validation: What peer-reviewed medical journals could do," *Journal of Korean Medical Science*, vol. 33, p. e152, 5 2018. PMID: 29805337 PMID: PMC5966371.
- [101] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, S. Huang, M. Brooks, M. J. Lee, and H. Asadi, "Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods," *American Journal of Roentgenology*, vol. 212, pp. 38–43, 10 2018. publisher: American Roentgen Ray Society.
- [102] D. A. Bluemke, L. Moy, M. A. Bredella, B. B. Ertl-Wagner, K. J. Fowler, V. J. Goh, E. F. Halpern, C. P. Hess, M. L. Schiebler, and C. R. Weiss, "Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and readers—from the radiology editorial board," *Radiology*, vol. 294, pp. 487–489, 12 2019. publisher: Radiological Society of North America.
- [103] D. G. Altman, I. Simera, J. Hoey, D. Moher, and K. Schulz, "Equator: reporting guidelines for health research," *The Lancet*, vol. 371, pp. 1149–1150, 4 2008. publisher: Elsevier PMID: 18395566.
- [104] J. M. Provenzale and R. J. Stanley, "A systematic guide to reviewing a manuscript," *AJR. American journal of roentgenology*, vol. 185, pp. 848–854, 10 2005. PMID: 16177399.
- [105] W. Luo, D. Phung, T. Tran, S. Gupta, S. Rana, C. Karmakar, A. Shilton, J. Yearwood, N. Dimitrova, T. B. Ho, S. Venkatesh, and M. Berk, "Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view," *Journal of Medical Internet Research*, vol. 18, p. e5870, 12 2016. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research publisher: JMIR Publications Inc., Toronto, Canada.
- [106] J. Mongan, L. Moy, and C. E. Kahn, "Checklist for artificial intelligence in medical imaging (claim): A guide for authors and reviewers," *Radiology: Artificial Intelligence*, vol. 2, p. e200029, 3 2020. publisher: Radiological Society of North America.
- [107] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice (3rd ed)*. Melbourne, Australia: OTexts, 3 ed., 2021. [Online; accessed 2021-04-14].

- [108] S. K. Smith and T. Sincich, “An empirical analysis of the effect of length of forecast horizon on population forecast errors,” *Demography*, vol. 28, no. 2, pp. 261–274, 1991. publisher: Springer.
- [109] Y. Wang, Y. Li, Y. Song, and X. Rong, “The influence of the activation function in a convolution neural network model of facial expression recognition,” *Applied Sciences*, vol. 10, p. 1897, 1 2020. number: 5 publisher: Multidisciplinary Digital Publishing Institute.
- [110] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.
- [111] G. Perin and S. Picek, “On the influence of optimizers in deep learning-based side-channel analysis,” *IACR Cryptol. ePrint Arch.*, 2020.
- [112] C. Bergmeir, R. J. Hyndman, and B. Koo, “A note on the validity of cross-validation for evaluating autoregressive time series prediction,” *Computational Statistics & Data Analysis*, vol. 120, pp. 70–83, 2018.
- [113] E. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Statistics for Biology and Health, Springer International Publishing, 2 ed., 2019. DOI: 10.1007/978-3-030-16399-0.
- [114] V. Amrhein, S. Greenland, and B. McShane, “Scientists rise up against statistical significance,” *Nature*, vol. 567, pp. 305–307, 3 2019. number: 7748 publisher: Nature Publishing Group.
- [115] D. W. Hosmer, T. Hosmer, S. L. Cessie, and S. Lemeshow, “A comparison of goodness-of-fit tests for the logistic regression model,” *Statistics in Medicine*, vol. 16, no. 9, pp. 965–980, 1997.
- [116] R. Nuzzo, “Scientific method: Statistical errors,” *Nature News*, vol. 506, p. 150, 2 2014. section: News Feature.
- [117] A. Davydenko and R. Fildes, “Forecast error measures: Critical review and practical recommendations,” 01 2016.
- [118] “Retracted coronavirus (covid-19) papers,” 4 2020. Available at <https://retractionwatch.com/retracted-coronavirus-covid-19-papers> [Online; accessed 2021-04-27].
- [119] N. B. Yahia, M. D. Kandara, and N. B. B. Saoud, “Deep ensemble learning method to forecast covid-19 outbreak,” tech. rep., In Review, 5 2020. DOI: 10.21203/rs.3.rs-27216/v1.
- [120] S. Chakraborty, A. K. Choudhary, M. Sarma, and M. K. Hazarika, “Reaction order and neural network approaches for the simulation of covid-19 spreading kinetic in india,” *Infectious Disease Modelling*, vol. 5, pp. 737–747, 1 2020.
- [121] K. T. Ly, “A covid-19 forecasting system using adaptive neuro-fuzzy inference,” *Finance Research Letters*, p. 101844, 11 2020.
- [122] N. Yudistira, “Covid-19 growth prediction using multivariate long short term memory,” *Preprint*, 5 2020. [Online; accessed 2020-12-15].
- [123] A. Chatterjee and S. Roy, “An analytics overview & lstm-based predictive modeling of covid-19: A hardheaded look across india,” in *Machine Intelligence and Soft Computing* (D. Bhattacharyya and N. Thirupathi Rao, eds.), (Singapore), pp. 289–307, Springer Singapore, 2021.

- [124] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm," *Chaos, Solitons & Fractals*, vol. 140, p. 110212, 11 2020.
- [125] L. Mohimont, A. Chemchem, F. Alin, M. Krajecki, and L. Steffemel, "Convolutional neural networks and temporal cnns for covid-19 forecasting in france," *Applied Intelligence*, 2021. [Online; accessed 2021-01-08].
- [126] H. Abbasimehr and R. Paki, "Prediction of covid-19 confirmed cases combining deep learning methods and bayesian optimization," *Chaos, Solitons & Fractals*, vol. 142, p. 110511, 1 2021.
- [127] V. Bharadi, "Random net implementation of mlp and lstms using averaging ensembles of deep learning models," *2020 International Conference on Decision Aid Sciences and Application (DASA)*, pp. 1197–1204, 2020.
- [128] A. Prakash, P. Sharma, I. K. Sinha, and U. P. Singh, "Spread peak prediction of covid-19 using ann and regression (workshop paper)," in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pp. 356–365, 2020.
- [129] A. Mollalo, K. M. Rivera, and B. Vahedi, "Artificial neural network modeling of novel coronavirus (covid-19) incidence rates across the continental united states," *International Journal of Environmental Research and Public Health*, vol. 17, p. 4204, 1 2020. number: 12 publisher: Multidisciplinary Digital Publishing Institute.
- [130] M. Saqib, "Forecasting covid-19 outbreak progression using hybrid polynomial-bayesian ridge regression model," *Applied Intelligence*, 10 2020. [Online; accessed 2021-01-22].
- [131] R. Moulay Taj, Z. Ait El Mouden, A. Jakimi, and M. Hajar, "Towards using recurrent neural networks for predicting influenza-like illness: Case study of covid-19 in morocco," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 7945–7950, 10 2020.
- [132] K. Shyam Sunder Reddy, Y. C. A. Padmanabha Reddy, and C. Mallikarjuna Rao, "Recurrent neural network based prediction of number of covid-19 cases in india," *Materials Today: Proceedings*, 11 2020. [Online; accessed 2021-01-12].
- [133] M. A. M. Arceda, P. C. L. Laura, and V. E. M. Arceda, "Forecasting time series with multiplicative trend exponential smoothing and lstm: Covid-19 case study," in *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 2* (K. Arai, S. Kapoor, and R. Bhatia, eds.), (Cham), pp. 568–582, Springer International Publishing, 2021.
- [134] S. Chander, V. Padmanabha, and J. Mani, "Jaya spider monkey optimization-driven deep convolutional lstm for the prediction of covid'19," *Bio-Algorithms and Med-Systems*, vol. 16, 11 2020.
- [135] A. S. Fokas, N. Dikaios, and G. A. Kastis, "Mathematical models and deep learning for predicting the number of individuals reported to be infected with sars-cov-2," *Journal of The Royal Society Interface*, vol. 17, no. 169, p. 20200494, 2020.
- [136] Z. Li, Y. Zheng, J. Xin, and G. Zhou, "A recurrent neural network and differential equation based spatiotemporal infectious disease model with application to covid-19," *arXiv:2007.10929 [cs, q-bio, stat]*, 9 2020. arXiv: 2007.10929.
- [137] M. Wiczorek, J. Siłka, and M. Woźniak, "Neural network powered covid-19 spread forecasting model," *Chaos, Solitons & Fractals*, vol. 140, p. 110203, 11 2020.

- [138] I. Pereira, J. Guérin, A. Junior, G. Garcia, P. Piscitelli, A. Miani, C. Distanto, and L. Gonçalves, “Forecasting covid-19 dynamics in brazil: A data driven approach,” *International Journal of Environmental Research and Public Health*, vol. 17, p. 5115, 7 2020.
- [139] s. Kirbaş, A. Sözen, A. D. Tuncer, and F. i. Kazancıoğlu, “Comparative analysis and forecasting of covid-19 cases in various european countries with arima, narnn and lstm approaches,” *Chaos, Solitons & Fractals*, vol. 138, p. 110015, 9 2020.
- [140] M. Wiecek, J. Siłka, D. Połap, M. Woźniak, and R. Damaševičius, “Real-time neural network based predictor for cov19 virus spread,” *PLOS ONE*, vol. 15, no. 12, p. e0243189, 2020. publisher: Public Library of Science.
- [141] A. Rodriguez, A. Tabassum, J. Cui, J. Xie, J. Ho, P. Agarwal, B. Adhikari, and B. A. Prakash, “Deepcovid: An operational deep learning-driven framework for explainable real-time covid-19 forecasting,” *medRxiv*, p. 2020.09.28.20203109, 9 2020. publisher: Cold Spring Harbor Laboratory Press.
- [142] A. Ramchandani, C. Fan, and A. Mostafavi, “Deepcovidnet: An interpretable deep learning model for predictive surveillance of covid-19 using heterogeneous features and their interactions,” *IEEE Access*, vol. 8, pp. 159915–159930, 2020. event: IEEE Access.
- [143] R. Sujath, J. M. Chatterjee, and A. E. Hassanien, “A machine learning forecasting model for covid-19 pandemic in india,” *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 959–972, 7 2020. Company: Springer Distributor: Springer Institution: Springer Label: Springer number: 7 publisher: Springer Berlin Heidelberg.
- [144] S. D. Khan, L. Alarabi, and S. Basalamah, “Toward smart lockdown: A novel approach for covid-19 hotspots prediction using a deep hybrid neural network,” *Computers*, vol. 9, p. 99, 12 2020.
- [145] C. Distanto, I. G. Pereira, L. M. G. Gonçalves, P. Piscitelli, and A. Miani, “Forecasting covid-19 outbreak progression in italian regions: A model based on neural network training from chinese data,” *medRxiv*, p. 2020.04.09.20059055, 4 2020. publisher: Cold Spring Harbor Laboratory Press.
- [146] A. H. Elsheikh, A. I. Saba, M. A. Elaziz, S. Lu, S. Shanmugan, T. Muthuramalingam, R. Kumar, A. O. Mosleh, F. A. Essa, and T. A. Shehabeldeen, “Deep learning-based forecasting model for covid-19 outbreak in saudi arabia,” *Process Safety and Environmental Protection*, vol. 149, pp. 223–233, 5 2021.
- [147] S. Dhamodharavadhani, R. Rathipriya, and J. M. Chatterjee, “Covid-19 mortality rate prediction for india using statistical neural network models,” *Frontiers in Public Health*, vol. 8, 2020. publisher: Frontiers.
- [148] S. Prasanth, U. Singh, A. Kumar, V. A. Tikkiwal, and P. H. J. Chong, “Forecasting spread of covid-19 using google trends: A hybrid gwo-deep learning approach,” *Chaos, Solitons & Fractals*, vol. 142, p. 110336, 1 2021.
- [149] H. T. Rauf, M. I. U. Lali, M. A. Khan, S. Kadry, H. Alolaiyan, A. Razaq, and R. Irfan, “Time series forecasting of covid-19 transmission in asia pacific countries using deep neural networks,” *Personal and Ubiquitous Computing*, 1 2021. [Online; accessed 2021-01-22].
- [150] P. Arora, H. Kumar, and B. K. Panigrahi, “Prediction and analysis of covid-19 positive cases using deep learning models: A descriptive case study of india,” *Chaos, Solitons & Fractals*, vol. 139, p. 110017, 10 2020.

- [151] O. Istaiteh, T. Owais, N. Al-Madi, and S. Abu-Soud, "Machine learning approaches for covid-19 forecasting," in *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pp. 50–57, 2020.
- [152] Y. Karadayi, M. N. Aydin, and A. S. Öğrenci, "Unsupervised anomaly detection in multivariate spatio-temporal data using deep learning: Early detection of covid-19 outbreak in italy," *IEEE Access*, vol. 8, pp. 164155–164177, 2020. event: IEEE Access.
- [153] A. I. Saba and A. H. Elsheikh, "Forecasting the prevalence of covid-19 outbreak in egypt using nonlinear autoregressive artificial neural networks," *Process Safety and Environmental Protection*, vol. 141, pp. 1–8, 9 2020.
- [154] S. Balli, "Data analysis of covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods," *Chaos, Solitons & Fractals*, vol. 142, p. 110512, 1 2021.
- [155] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, "Covid-19: A comparison of time series methods to forecast percentage of active cases per population," *Applied Sciences*, vol. 10, p. 3880, 1 2020. number: 11 publisher: Multidisciplinary Digital Publishing Institute.
- [156] M. Hawas, "Generated time-series prediction data of covid-19's daily infections in brazil by using recurrent neural networks," *Data in Brief*, vol. 32, p. 106175, 10 2020.
- [157] M. A. Achterberg, B. Prasse, L. Ma, S. Trajanovski, M. Kitsak, and P. Van Mieghem, "Comparing the accuracy of several network-based covid-19 prediction algorithms," *International Journal of Forecasting*, 10 2020. [Online; accessed 2021-01-20].
- [158] P. Wang, X. Zheng, G. Ai, D. Liu, and B. Zhu, "Time series prediction for the epidemic trends of covid-19 using the improved lstm deep learning method: Case studies in russia, peru and iran," *Chaos, Solitons & Fractals*, vol. 140, p. 110214, 11 2020.
- [159] S. Thakur, D. K. Patel, B. Soni, M. Raval, and S. Chaudhary, "Prediction for the second wave of covid-19 in india," in *Big Data Analytics* (L. Bellatreche, V. Goyal, H. Fujita, A. Mondal, and P. K. Reddy, eds.), (Cham), pp. 134–150, Springer International Publishing, 2020.
- [160] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of covid-19 using deep learning models: India-usa comparative case study," *Chaos, Solitons & Fractals*, vol. 140, p. 110227, 11 2020.
- [161] S. Saif, P. Das, and S. Biswas, "A hybrid model based on mba-anfis for covid-19 confirmed cases prediction and forecast," *Journal of The Institution of Engineers (India): Series B*, 1 2021. [Online; accessed 2021-01-22].
- [162] Z. Zhao, K. Nehil-Puleo, and Y. Zhao, "How well can we forecast the covid-19 pandemic with curve fitting and recurrent neural networks?," *Preprint*, 5 2020. DOI: 10.1101/2020.05.14.20102541.
- [163] N. M. Ghazaly, M. A. Abdel-Fattah, and A. A. A. El-Aziz, "Novel coronavirus forecasting model using nonlinear autoregressive artificial neural network," *International Journal of Advanced Science and Technology*, vol. 29, pp. 1831–1849, 4 2020. number: 5s.
- [164] S. Bahri, M. Kdayem, and N. Zoghلامي, "Deep learning for covid-19 prediction," in *2020 4th International Conference on Advanced Systems and Emergent Technologies*, pp. 406–411, 2020.

- [165] Y. Gautam, "Transfer learning for covid-19 cases and deaths forecast using lstm network," *ISA Transactions*, 1 2021. [Online; accessed 2021-01-20].
- [166] J. A. L. Marques, F. N. B. Gois, J. Xavier-Neto, and S. J. Fong, "Artificial intelligence prediction for the covid-19 data based on lstm neural networks and h2o automl," in *Predictive Models for Decision Support in the COVID-19 Crisis* (J. A. L. Marques, F. N. B. Gois, J. Xavier-Neto, and S. J. Fong, eds.), SpringerBriefs in Applied Sciences and Technology, pp. 69–87, Cham: Springer International Publishing, 2021.
- [167] S. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. Varkonyi-Koczy, U. Reuter, T. Rabczuk, and P. Atkinson, "Covid-19 outbreak prediction with machine learning," *Algorithms*, vol. 13, 9 2020.
- [168] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, "Chapter 10 - deep learning," in *Data Mining (Fourth Edition)* (I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, eds.), pp. 417–466, Morgan Kaufmann, 1 2017. DOI: 10.1016/B978-0-12-804291-5.00010-6.
- [169] J. Brownlee, *Deep Learning for Time Series Forecasting*. Machine Learning Mastery, 2019.
- [170] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005. publisher: Inter-Research Science Center.
- [171] T. Chai and R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)?," *Geosci. Model Dev.*, vol. 7, 1 2014.
- [172] "Artifact review and badging - current." Available at <https://www.acm.org/publications/policies/artifact-review-and-badging-current> [Online; accessed 2021-05-09].
- [173] K. Broman, M. Cetinkaya-Rundel, A. Nussbaum, C. Paciorek, R. Peng, D. Turek, and H. Wickham, "Recommendations to funding agencies for supporting reproducible research."
- [174] S. L. Wu, A. N. Mertens, Y. S. Crider, A. Nguyen, N. N. Pokpongkiat, S. Djajadi, A. Seth, M. S. Hsiang, J. M. Colford, A. Reingold, B. F. Arnold, A. Hubbard, and J. Benjamin-Chung, "Substantial underestimation of sars-cov-2 infection in the united states," *Nature Communications*, vol. 11, p. 4507, 9 2020. number: 1 publisher: Nature Publishing Group.