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

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

Enrico Longato, Mario Luca Morieri, Giovanni Sparacino, Barbara Di Camillo, Annamaria Cattelan, Sara Lo Menzo, Marco Trevenzoli, Andrea Vianello, Gabriella Guarnieri, Federico Lionello, Angelo Avogaro, Paola Fioretto, Roberto Vettor, Gian Paolo Fadini

**A R T I C L E   I N F O**

**Article history:**
Received 16 December 2021
Revised 21 April 2022
Accepted 9 May 2022

**Keywords:**
COVID-19
Mortality
Dynamic Bayesian network
Dynamic time warping
Graphical model

**A B S T R A C T**

**Background and Objective:** COVID-19 severity spans an entire clinical spectrum from asymptomatic to fatal. Most patients who require in-hospital care are admitted to non-intensive wards, but their clinical conditions can deteriorate suddenly and some eventually die. Clinical data from patients’ case series have identified pre-hospital and in-hospital risk factors for adverse COVID-19 outcomes. However, most prior studies used static variables or dynamic changes of a few selected variables of interest. In this study, we aimed at integrating the analysis of time-varying multidimensional clinical-laboratory data to describe the pathways leading to COVID-19 outcomes among patients initially hospitalised in a non-intensive care setting.

**Methods:** We collected the longitudinal retrospective data of 394 patients admitted to non-intensive care units at the University Hospital of Padova (Padova, Italy) due to COVID-19. We trained a dynamic Bayesian network (DBN) to encode the conditional probability relationships over time between death and all available demographics, pre-existing conditions, and clinical laboratory variables. We applied resampling, dynamic time warping, and prototyping to describe the typical trajectories of patients who died vs. those who survived.

**Results:** The DBN revealed that the trajectory linking demographics and pre-existing clinical conditions to death progressed directly through kidney dysfunction or, more indirectly, through cardiac damage. As expected, admission to the intensive care unit was linked to markers of respiratory function. Notably, death was linked to elevation in procalcitonin and D-dimer levels. Death was associated with persistently high levels of procalcitonin from admission and throughout the hospital stay, likely reflecting bacterial superinfection. A sudden raise in D-dimer levels 3–6 days after admission was also associated with subsequent death, possibly reflecting a worsening thrombotic microangiopathy.

**Conclusions:** This innovative application of DBNs and prototyping to integrated data analysis enables visualising the patient’s trajectories to COVID-19 outcomes and may instruct timely and appropriate clinical decisions.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Coronavirus disease 2019 (COVID-19) is a heterogeneous clinical condition caused by infection with the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). The clinical spectrum of the disease ranges from asymptomatic to mildly symptomatic to severe forms requiring hospitalisation. Hospitalised patients with COVID-19 can require intensive care support and mechanical ventilation to maintain adequate respiratory gas exchange. The overall mortality is 2.2%, but it increases substantially in those hospitalised (14%) and further so in those admitted to the intensive care unit (37%) [1]. The most common causes of death in COVID-19 patients are multiple organ dysfunction syndrome, secondary infections, refractory hypoxemia, and ischemic events [2, 3]. Pulmonary artery thrombosis and thrombotic microangiopathy are typically associ-
ated with diffuse alveolar damage, leading to multi-organ dysfunction [4,5]. It is still unclear what drives the transition through the various COVID-19 stages of severity. Although a few treatments have been approved for COVID-19, the exact timing of their use is debated [6]. Studying the trajectories of patients with COVID-19 from hospitalisation to recovery/death can unravel valuable information for monitoring the disease course and choosing the appropriate treatment at each stage.

To date, the detailed description of COVID-19 clinical course has been hampered by large patient heterogeneity, differences in data collection, and by the wide variety of variables being recorded during hospital stay. Integration of data time-series with clinical stage and outcomes requires complex analytical approaches and generally leads to complex input-output relationships. In this sense, following the paradigms of explainable AI, graphical methods are particularly suitable for providing a human-interpretable description of phenomena [7].

Here, we used dynamic Bayesian networks (DBNs) to model the trajectories of patients hospitalised for COVID-19 in a non-intensive care setting.

2. Methods

2.1. Database

This was a single-centre, retrospective, observational study. We collected data of 394 patients who were admitted to the non-intensive care units of the University Hospital of Padova between February 21st, 2020 and April 14th, 2020 due to COVID-19. Patients were followed over time until discharge, death, or May 25th, 2020 (median follow-up time: 18 days; IQR: 6 to 30 days). As shown in Table S1, we collected a wide range of variables, including demographics, pre-existing conditions, and clinical laboratory data. Such variables were logically grouped to inform on inflammation, metabolism, coagulation, tissue injury, as well as respiratory, cardiac renal, hepatic and pancreas function. We recorded information on ICU entry and death. Updated values of time-dependent variables and their acquisition times, as well as dates of death or hospital discharge, were also available.

2.2. Dynamic Bayesian networks

Bayesian networks (BNs) are graphical models that encode joint probabilistic relationships in the form of a directed acyclic graph (DAG), i.e., as a set of nodes (representing variables) connected by directed edges (signalling the existence of a conditional probability link) without self-feed or feedback loops [8,9]. The structure of a trained BN encapsulates the statistical independence properties of the underlying joint probability distribution that is factorised by the DAG. A set of conditional probability tables (CPTs) completes the representation by quantifying the strength of the association between each node and its parents, i.e., the nodes from which all edges pointing to the node start.

Standard BNs are implicitly static constructs because the mandated absence of cycles prevents variables from influencing their own future values. They can, however, be extended to encode conditional dependency relationships over time, defining a graphical model known as a dynamic Bayesian network (DBN) [10], whose fundamental structural unit is the two-timeslice BN (2TBN). Briefly, a 2TBN is a constrained DAG between a set of variables at a certain point in time (the t slice) and a subset of those that were also measured at the immediately preceding time point (the t − 1 slice). Following a first-order Markovianity assumption, edges can only exist between variables in the t slice, or they can follow an ideal arrow of time and go from the t − 1 slice to the t slice; edges among variables in the t − 1 slice are forbidden, and so are those from the t slice to the t − 1 slice. Furthermore, it is also assumed that network topology and CPTs do not change over time, i.e., between sets of t / t − 1 slices. In practice, this means that static or quasi-static variables (i.e., variables that are expected to change or be updated on a much larger time scale than the inter-slice period, if at all) belong to the t slice, while dynamic variables are replicated in both slices.

2.3. Data pre-processing

The raw data, collected during routine care, were inherently sparse and sparsely sampled. They also comprised a mixture of continuous and categorical variables, a problem currently intractable with state-of-the-art DBN software. To address these shortcomings, we applied a cascade of three pre-processing steps: 1) aggregation into multiple-day slices, 2) intra-subject forward and backward fill of sparse measurements, and 3) quantisation of continuous variables.

Variable update followed a real-life schedule, with new tests being requested by clinicians on an “as-needed” basis. To decrease the degree of missingness, and to avoid the interpretation difficulties associated with a direct modelling of the passage of time within the DBN, we divided each patient’s follow-up time into contiguous 3-day time slices, and aggregated all measurements collected within each 3-day interval. If more than one measurement was available within the same slice, we retained only the worst value (i.e. the maximum value, except for Antithrombin III, albumin, arterial pH, and oxygen saturation, of which we selected the minimum). This distinction was based on emphasising pathological vs. physiological states that might have been apparent during a slice. Due to the constraints imposed by the 2TBN structure, we excluded the 27 patients who did not have at least two consecutive slices. We selected a 3-day slice width to minimise the number of excluded subjects due to lack of contiguous visits: the next best alternatives, 2 and 4 days, would have resulted in >10% exclusion rates vs. the obtained 6.85%.

After aggregation, we further addressed intra-subject missingness due to lack of updates by progressively forward filling the most updated value of each missing variable. We also backward filled any remaining missing measurements at the beginning of follow-up. This is consistent with the assumption that clinicians would request a new batch of tests only if there were a suspicion of a meaningful change having happened since the least measurement. As for the remaining missing variables, we imputed their values via K-nearest neighbours with K = 10 after quantisation (see below), but before entering the DBN training algorithm.

Finally, we converted all continuous variables to categorical variables via quantisation. As shown in Table S2, we followed two distinct strategies: either we adhered to commonly-used, clinically significant thresholds (rule type = TH); or we defined three categories based on the terciles calculated on the original data, before forward and backward filling (rule type = Q). We preferred the latter approach when thresholds were undefined or outright meaningless in the context of hospitalisation for COVID-19 (e.g., as evidenced by the terciles reported in Table S2, C-reactive protein was consistently outside the reference range). In all but two cases, quantisation criteria were based on a single variable. The exceptions were HDL cholesterol, which has different thresholds in male and female subjects (40 vs. 50 mg/dL), and oxygen saturation, for which we designed four categories, three based on the 90% and 94% clinically significant thresholds, and a fourth one corresponding to absence of pneumonia without a recorded saturation measurement.
2.4. Structure and training of the DBN

Being completely data-driven and based on conditional probability refactorisation, DBNs might converge to a probabilistically sound, but difficult-to-interpretable structure. Hence, we imposed several constraints on network topology based on the distinction between static (or quasi-static) and dynamic variables, and on basic domain knowledge. This process is known as layering. Specifically, we treated all variables with repeated measurements as dynamic, except brain natriuretic peptide, total cholesterol, amylase, and creatine phosphokinase, which did not exhibit sufficient variability after quantisation. We also considered death as a quasi-static variable, because, by definition, it was always equal to 0 except, possibly, once at the end of follow-up. All other variables were inherently static. Additionally, we forbade age and sex from having parent nodes to avoid unrealistic edge directions (i.e., age or sex being determined by other variables). The final layering was as follows: age and sex could influence all other static/quasi-static variables and dynamic variables at time t, static/quasi-static variables could influence each other and dynamic variables at time t, dynamic variables at times t - 1 and t could only influence dynamic variables at time t.

After K-nearest neighbour imputation, we trained the DBN on 1464 pairs of contiguous t - 1 and t slices via the max-min hill-climbing algorithm [11] according to the Bayesian Dirichlet equivalent uniform (BDeu) score [12], followed by a maximum a-posteriori estimation of the CPDs. To confirm the stability of the network’s topology, we retrained the DBN on 200 bootstrap samples, each 90% as large as the initial dataset. The main objective of this sensitivity analysis was to exclude any undue effects of resampling on the network’s general structure, and especially on its connected components.

2.5. Resampling, dynamic time warping, and prototyping

A trained DBN lends itself to meaningful considerations on the probabilistic relationships that exist over time among the variables on its nodes. In other words, it allows to substantiate statements such as “variable A is conditionally independent of variable set B given variable set C,” or, equivalently, “knowledge of variable set C probabilistically characterises variable A, regardless of variable set B,” or, again, “measuring variable set B provides enough information on variable A that variable set C can be ignored.” While this type of inference is extremely valuable to describe the stochastic process encoded by the DBN, on the one hand, it is prone to overinterpretation (e.g., typically, edge direction is misunderstood as a cause-effect relationship), on the other, especially when a lot of variables are involved, it fails to give an intuitive, time-aware overview of the entire network of relationships.

One way to address this issue is to 1) go back to the original signals, 2) calculate a summary of the available trajectories, stratifying by one or more characteristics of interest (e.g., death vs. survival), and 3) compare these summary evolutions by focusing on the relationships highlighted by the DBN. However, this approach is inadequate for the purposes of the present work (or, indeed, many others based on routinely acquired data), because it is very sensitive to the relatively low sample size, and ill-equipped to deal with the high missingness of the variables involved. Hence, to partly overcome these limitations, we developed an alternative framework to highlight the differences, encoded by the DBN, between patients who died and those who did not. Specifically, we proceeded in three steps by mirroring the three points outlined above. First, we sampled from the trained DBN, obtaining 10,000 sets of 31 signals (connected component including the death variable), corresponding to 30 (or fewer, in case of simulated death) days of hospitalisation of 10,000 synthetic patients. At the same time, we rescaled the 10,000 sets of signals by substituting the sampled quantised values with the sample means in the corresponding quantisation bin estimated in the original dataset (e.g., as per Table S2, we substituted a simulated, quantised value of “1” for age with the average value 67.8). Conceptually, this was an amplification step of sorts, whereby we took the information encoded by the DBN and decoded it into a sufficiently large sample to obtain discernible summary statistics. Second, after excluding the 123 simulated patients who immediately died, we stratified the 1055 dead and 8822 survived patients and calculated the barycenters of their trajectories via dynamic time warping (DTW) [13], and, specifically, using the softDTW distance [14]. Thus, we obtained two sets of signals corresponding to the prototype evolution of all encoded covariates in survivors and non-survivors. These prototypes were the summaries to be compared. Note that, since DTW privileges the shape of a signal over its exact coordinates, the shown prototypes may be stretched across the horizontal (slice) axis and are, thus, mainly illustrative. Third, to make sure that the comparison was valid, we also verified, via the area under the receiver-operating characteristic curve (AUROC), that the soft-DTW distance between each simulated subject and the prototype of dead patients’ evolution was a good discriminator between simulated dead vs. survived patients. The idea was that a good summary of dead vs. survived simulated patients implies that a subject who died should be closer to the dead subjects’ prototype in the softDTW sense.

To further exemplify the way in which dynamic relationships are encoded by the network, we also produced an animated video comparing the probabilistic trajectories across 30 days of two sets of 1000 patients with the following characteristics at the start of the simulation: male, aged 61 to 74, affected by diabetes, dyslipidaemia, hypertension, and history of kidney disease. The first set of 1000 men maintained creatinine values between 71 and 93 μmol/L for the entire simulated observation time; the second set also started from the 71 to 93 μmol/L range, but permanently switched to the >93 μmol/L range between days 4 and 6. We carried forward the last set of observed values upon simulated death, and encoded the mean quantised value of each variable by filling the corresponding node with a colour ranging from white (all samples belonging to the lowest category) to deep red (all samples belonging to the highest category).

2.6. Software implementation

We carried out the analyses with the following software: bnstruct v1.0.8 R package to train the DBN [15], pgmmpy v0.1.11 python library to sample from it [16], tlearn v0.4.1 python library to implement DTW and compute the prototypes [17]. R version 4.0.2; python version 3.7.7.

3. Results

As shown in Table 1, the 394 patients encoded by the DBN were, on average, 64.7 years old, predominantly male (58.8%) and affected by hypertension (50.5%), likely to have diabetes (20.3%), dyslipidaemia (22.3%), cardiovascular disease (17.0%), or cancer (15.7%), and with a 71.5% chance of having chronic kidney disease or chronic obstructive pulmonary disease. 17.0% of the patients entered the ICU at some point during hospitalisation, and 12.7% died. The rate of missing basic patient information was very low (usually 0%, max 3.8%).

The resulting structure of the trained DBN comprised a main connected component, including most of the variables and death, and several independent, smaller components (1 to 4 variables each). The main connected component is shown in Fig. 1. The remaining, unshown, variables were LDL cholesterol, HDL cholesterol,
triglycerides, glucose, antithrombin-III, activated partial thromboplastin time, prothrombin time, sodium, potassium, urinary haemoglobin, proteins and ketone bodies, liver function markers (aspartate transaminase, alanine transaminase, alkaline phosphatase, gamma-glutamyl transferase, total bilirubin), lactate dehydrogenase, and swab test PCR results for SARS-CoV-2. All dynamic variables exhibited an inter-slice dependency from their previous values.

From the analysis of Markov blankets, we found that death was conditionally independent of all other variables given procalcitonin and D-dimer. Similarly, entering the ICU was conditionally independent of all other variables given urea, albumin, oxygen saturation, and arterial pCO2.

**Fig. 2** illustrates the prototypal temporal evolution of the 17 dynamic variables belonging to the DBN’s main connected component in simulated cases of patients recovering or dying from COVID-19. As expected, procalcitonin and D-dimer exhibited the most marked divergence between those who died and those who recovered. Elevated levels of procalcitonin and D-dimer were, indeed, associated with subsequent death. Levels of C-reactive protein, urea, creatinine, and fibrinogen were also higher among patients who died, but the difference between the two groups were much smaller than for procalcitonin and D-dimer. Prototyped temporal evolution of the variables of interest is also reported graphically as a heatmap in Figure S1.

The confirmatory analysis on the discrimination ability of the prototypes yielded a satisfactory 0.80 AUROC, suggesting that these were sufficiently representative of the simulated dead vs. survived patient populations.

**Fig. S2** presents a practical example of the way in which DBNs encode probabilistic relationships over time (also illustrated in the heatmap of Figure S3). Consistently with the barycentres shown in **Fig. 2,** of the 1000 simulated subjects with increasing creatinine levels, 24.4% died before the end of the 30-day simulations vs. 9.5% of those with consistently lower levels, despite both resampled subpopulations starting from approximately the same baseline conditions. Different probabilistic trajectories are also apparent in the neighbourhood of the creatinine node (e.g., urea, white blood cell count, procalcitonin, and C-reactive protein), confirming that the DBN has indeed encoded not only time-varying, but also multivariate dependencies. The links between urea and creatinine concentrations, as well as among inflammatory markers, highlight the biological plausibility of the resulting network.

### 4. Discussion

In this study, we used dynamic Bayesian networks (DBN) to analyse clinical variables time series from patients with COVID-19 initially hospitalised in a non-critical care setting. With this analysis, we found that admittance to the ICU was connected to blood gases reflecting respiratory function (oxygen saturation and pCO2), one marker of general anabolic status (albumin), and one marker of renal function (blood urea, which was itself linked to serum creatinine). Death was tightly connected to levels of procalcitonin and D-dimer. Although causation cannot be inferred from direction of the edges in the trained DBN, these findings are in line with prior knowledge on the clinical variables associated with progression from mild/moderate to severe COVID-19 requiring intensive care and eventually leading to patient’s death. Indeed, worsening respiratory function is the most common cause of admittance to the ICU, which is the appropriate setting for invasive ventilation. On the other side, variables connected to death were related to secondary infection (procalcitonin) and thrombotic microangiopathy (D-dimer), which, in fact, have been identified as frequent causes of death in COVID-19 patients.

Resampling the initial case series, prototyping patients who died vs. those who survived, and DTW allowed a graphical representation of the time trends in the variables included in the network. Notably, an early separation was observed for the trend in procalcitonin (at admission) and D-dimer (at 3–6 days) values, which remained markedly higher in prototyped patients who died vs. those who survived during a simulated 30-day disease course. On the other side, CRP was only mildly higher in patients who subsequently died, whereas it slowly decreased in patients who recovered. Interestingly, procalcitonin, more than CRP, can differentiate inflammation due to secondary bacterial infection from the hyperinflammatory state occurring in response to SARS-CoV-2 infection [18]. Therefore, an elevated procalcitonin level at admission, more than any elevation of CRP, should spotlight patients at higher risk of death and drive prompt therapeutic choices, including, e.g., use of antibiotics [19].

D-dimer levels have been shown to be elevated in patients with moderate COVID-19, possibly reflecting a diffuse thrombotic microangiopathy, rather than a localised pulmonary thromboembolism. At variance with procalcitonin, a steep increase in D-dimer levels within the first 3–6 days of admission, rather than merely elevated levels at admission, were associated with subsequent death. Therefore, closely looking at D-dimer dynamics during the early state in the non-intensive care setting can inform on mortality risk and possibly drive more aggressive therapy, including, e.g., use of anti-coagulants [20].

Based on these observations, results of our study also help clinicians in deciding whether and when biomarkers should be rechecked during the hospital stay in order to assess a patient’s risk. It appears that death is associated with elevated procalcitonin levels upon admission, such that further check may not change the expected patient’s trajectory. On the contrary, re-checking D-dimer at 3–6 days may identify patients who are likely to develop more severe COVID-19 and die.

### Table 1

Patient characteristics. Age, the only continuous variable, is shown as mean ± SD; all other variables are shown as percentages. The third column reports the fraction of missing values.

| Variable (units) | Values | Missing (%) [N = 394] |
|------------------|--------|-----------------------|
| Age (years)      | 64.7 ± 15.4 | 0.0 |
| Male sex (Y/N)   | 58.9% | 0.0 |
| Diabetes (Y/N)   | 20.3% | 0.0 |
| Hypertension (Y/N) | 50.5% | 0.0 |
| Dyslipidaemia (Y/N) | 22.3% | 0.0 |
| Cardiovascular disease (Y/N) | 17.0% | 3.6 |
| Chronic kidney disease (Y/N) | 7.1% | 0.0 |
| Chronic obstructive pulmonary disease (Y/N) | 7.1% | 3.6 |
| Cancer (Y/N)     | 15.7% | 3.8 |
| Death (Y/N)      | 12.7% | 0.0 |
| Entered ICU (Y/N) | 17.0% | 1.8 |
Outstanding is the observation that the probabilistic trajectories from static variables (such as comorbidities) to outcomes (ICU or death) inevitably proceeded through markers of kidney function (urea and creatinine). In the network, the link between ICU and death is only made possible through kidney function. Closely looking at the trained DBN, the only alternative pathway linking ICU admittance to static variables involves markers of cardiac damage (BNP and TnI), which is another established feature of multi-organ failure occurring during severe COVID-19 [21]. This pathway is, however, much less linear, spanning more nodes, and is not linked to death independently from kidney function. Figure S4 shows a simplified view of the DBN where these paths are highlighted.

Such findings are consistent with prior observations that kidney disease is associated with in-hospital death of patients with COVID-19 [22]. Several studies have reported that a history of chronic kidney disease at admission or the development of acute kidney injury during hospital stay predict poor COVID-19 outcomes [23, 24]. However, no prior study had examined the dynamic rela-
Fig. 2. Prototypes obtained via DWT. The 17 pairs of signals are the dynamic variables belonging to the connected components that includes death, extracted from the prototypes of dead and survived simulated patients.
tionships between changes in kidney function and mortality with the simultaneous changes of other relevant variables in a single model. This is particularly evident in the dynamic representation shown in Supplementary Fig. 1 where, at means of covariates, mild in-hospital elevations in serum creatinine drove changes in inflammatory markers and projected a 2.6 higher mortality rate. Of note, such trajectory is unrelated to changes in markers of respiratory function (oxygen saturation and CO2 levels)

BNs, in general, have been scarcely utilised to model COVID-19. The most prolific subfield of application appears to be that of population-level dynamics, investigating topics such as outbreak evolution [25], infection-fatality rates [26], or even the impact of economic disruptions [27]. Other applications involve the usage of DBNs as an explainability tool that is part of a larger experimental pipeline [28]. On the contrary, here, we have presented an approach that is fully based on DBNs and focused on hospitalised patients and their clinical parameters specifically, rather than the population as a whole.

We wish to acknowledge some limitations of the approach we have taken to describe the trajectories to COVID-19 outcomes. First, both the analysis of the DBN and that of the prototype signals are descriptive, rather than predictive. In fact, although we have trained the former using state-of-the-art methods and the latter passed a verification step regarding the separability of dead and survived simulated patients, the development of a prognostic model was out of scope for the present work. Moreover, although continuous variable quantisation and substitution with the statistical description of each quantile (here, the mean) is a relatively standard approach (e.g., for quantile normalisation), it inevitably leads to a loss in resolution with respect to biomarker values. Finally, since DWT is a warping technique rather than an instant-by-instant transformation, there was an implicit overweighting of the shape of each signal vs. the exact timing of shape changes, which, however, may be informative in their own right to describe the inhospital evolution of COVID-19.

Notwithstanding such limitations, our work has notable outputs. We combined innovative analytical approaches to model the dynamic clinical status of patients proceeding through different states of COVID-19 severity, from hospitalisation in non-critical conditions to ICU admittance and eventual death. This highlights the importance of choosing tools able to depict clinical trajectories in a way that is intelligible and useful to clinicians. DBN enabled us to describe the different dynamics of variables associated with COVID-19 mortality, especially procalcitonin, D-dimer and CRP. Such findings entail potential implications for clinical decision-making, including the timing of testing and the choice among therapeutic options. Finally, the DBN clearly identified kidney function as the fulcrum of pathways leading to poor COVID-19 outcomes, providing a strong contribution to our understanding of multi-organ dysfunction during COVID-19.

Author contribution

Study design: EL, MLM, GS, BDC, AA, PF, GF. Data collection and analysis: EL, MLM, AC, SLM, MT, AV, GG, FL. Manuscript writing: EL, BDC, PF, AA, RV, GF. Supervision: GS, AA, RV, GF. All authors revised the manuscript and approved the final version.

Declaration of Competing Interest

The authors declare no conflict of interest in relation to the content of this manuscript.

Data sharing policy

Original data analysed in this study are available from the corresponding author at a reasonable request.

Funding information

The study was supported by the Foundation CARIPARO grant to RV (project COVIDIMED).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2022.106873.

References

[1] G.P. Fadini, et al., Newly-diagnosed diabetes and admission hyperglycaemia predict COVID-19 severity by aggravating respiratory deterioration, Diabetes Res. Clin. Pract. 168 (2020) 108374.
[2] C. de Roquetailleade, et al., Timing and causes of death in severe COVID-19 patients, Crit. Care 25 (2021) 224.
[3] S. Elezakurta, et al., Causes of death and comorbidities in hospitalized patients with COVID-19, Sci. Rep. 11 (2021) 4263.
[4] N.M. Burkhard-Koren, et al., Higher prevalence of pulmonary macrothrombi in SARS-CoV-2 than in influenza A: autopsy results from “Spanish flu” 1918/1919 in Switzerland to Coronavirus disease 2019, J. Pathol. Clin. Res. 7 (2021) 135–143.
[5] E.S. Romanova, et al., Cause of death based on systematic post-mortem studies in patients with positive SARS-CoV-2 tissue PCR during the COVID-19 pandemic, J. Intern. Med. (2021), doi: 10.1111/joim.13300.
[6] Treatments for COVID-19, Harvard Health (2020) [July 21, 2021].
[7] J. Amann, et al., Explainability for artificial intelligence in healthcare: a multi-disciplinary perspective, BMC Med. Inform. Decis. Mak. 20 (2020) 310.
[8] A. Darwiche, Chapter 11 Bayesian networks “in foundations of artificial intelligence, in: F. van Harmelen, V. Lifschitz, B. Porter (Eds.), Handbook of Knowledge Representation, Elsevier, 2008, pp. 467–509.
[9] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Elsevier, 2014.
[10] S. Parsons, Probabilistic graphical models: principles and techniques by Daphne Koller and Nir Friedman, MIT Press. 1231 pp., 995.00, ISBN 0-262-01319-3, Knowl. Eng. Rev. 26 (2011) 237–238.
[11] I. Tsimardinos, L.E. Brown, C.F. Aliferis, The max-min hill-climbing Bayesian network structure learning algorithm, Mach. Learn. 65 (2006) 31–78.
[12] D. Heckerman, D. Geiger, D.M. Chickering, Learning Bayesian networks: the combination of knowledge and statistical data, Mach. Learn. 20 (1995) 197–243.
[13] S. Aghabozorgi, A. Seyed Shirkhorshidi, T. Ying Wah, Time-series clustering – a decade review, Inf. Syst. 53 (2015) 16–38.
[14] M. Cuturi, M. Blondel, Soft-DTW: a differentiable loss function for time-series, in: Proceedings of the 34th International Conference on Machine Learning - 70, Sydney, NSW, Australia, 2017, pp. 894–903.
[15] A. Franzini, F. Sambo, B. Di Camillo, bstruct: an R package for Bayesian network system structure learning in the presence of missing data, Bioinformatics (2016) bvw807.
[16] A. Ankan, A. Panda, pgmpy: probabilistic graphical models using python in (2015), pp. 6–11.
[17] R. Tavenard, et al., Tilearn, a machine learning toolkit for time series data, J. Mach. Learn. Res. 21 (2020) 1–6.
[18] E.J. Williams, et al., Evaluation of procalcitonin as a contribution to antimicrobial stewardship in SARS-CoV-2 infection: a retrospective cohort study, J. Hosp. Infect. 110 (2021) 103–107.
[19] N. Vazzana, F. Digaola, S. Ogibnene, Procalcitonin and secondary bacterial infections in COVID-19: association with disease severity and outcomes, Acta Clin. Belg. (2020) 1–5.
[20] M. Artifoni, et al., Systematic assessment of venous thromboembolism in COVID-19 patients receiving thromboprophylaxis: incidence and role of D-dimer as predictive factors, J. Thromb. Thrombolysis 50 (2020) 211–216.
[21] S. Shi, et al., Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China, JAMA Cardiol. 5 (2020) 802–810.
[22] Y. Cheng, et al., Kidney disease is associated with in-hospital death of patients with COVID-19, Kidney Int. 97 (2020) 825–838.
[23] J.S. Hirsch, et al., Acute kidney injury in patients hospitalized with COVID-19, Kidney Int. 98 (2020) 209–218.
[24] B.M. Henry, G. Lipp, Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection, Int. Urol. Nephrol. 52 (2020) 1193–1194.
[25] V. Vitale, P. D’Urso, L. De Giovanni, Spatio-temporal object-oriented Bayesian network modelling of the COVID-19 Italian outbreak data, Spatial Statistics, 2021.

[26] M. Neil, N. Fenton, M. Osman, S. McLachlan, Bayesian network analysis of Covid-19 data reveals higher infection prevalence rates and lower fatality rates than widely reported, J. Risk Res. 23 (2020) 866–879.

[27] S. Hosseini, D. Ivanov, A multi-layer Bayesian network method for supply chain disruption modelling in the wake of the COVID-19 pandemic, Int. J. Prod. Res. (2021) https://doi.org/10.1080/00207543.2021.1953180.

[28] V.C. Pezoulas, et al., A multimodal approach for the risk prediction of intensive care and mortality in patients with COVID-19, Diagnostics 12 (2022) 56.