RESEARCH ARTICLE

A causal learning framework for the analysis and interpretation of COVID-19 clinical data

Elisa Ferrari [ID]¹*, **Luna Gargani**², **Greta Barbieri**³,⁴, **Lorenzo Ghiadoni**⁵, **Francesco Faita**²ᵒ, **Davide Bacciu**⁶ᵒ

1 Scuola Normale Superiore, Pisa, Italy, **2** Institute of Clinical Physiology, C.N.R, Pisa, Italy, **3** Department of Surgical, Medical, Molecular and Critical Area Pathology, University of Pisa, Pisa, Italy, **4** Emergency Medicine Department, Pisa University Hospital, Pisa, Italy, **5** Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy, **6** Department of Computer Science, University of Pisa, Pisa, Italy

ᵒ These authors contributed equally to this work.
* elisa.ferrari@sns.it

Data Availability Statement: The dataset acquired and analyzed during the current study is not publicly available due to privacy policies of patient's information; in particular, the informed consent did not explicitly refer to the possibility of freely sharing data among research groups. Furthermore, the data cannot be published because our ethical committee stated that the informed consent provided to the patient does not allow us to make the data public, even if anonymized. In fact, the data have been collected within the small town of Pisa and the hospitalized cases in this area between March and

Abstract

We present a workflow for clinical data analysis that relies on Bayesian Structure Learning (BSL), an unsupervised learning approach, robust to noise and biases, that allows to incorporate prior medical knowledge into the learning process and that provides explainable results in the form of a graph showing the causal connections among the analyzed features. The workflow consists in a multi-step approach that goes from identifying the main causes of patient's outcome through BSL, to the realization of a tool suitable for clinical practice, based on a Binary Decision Tree (BDT), to recognize patients at high-risk with information available already at hospital admission time. We evaluate our approach on a feature-rich dataset of Coronavirus disease (COVID-19), showing that the proposed framework provides a schematic overview of the multi-factorial processes that jointly contribute to the outcome. We compare our findings with current literature on COVID-19, showing that this approach allows to re-discover established cause-effect relationships about the disease. Further, our approach yields to a highly interpretable tool correctly predicting the outcome of 85% of subjects based exclusively on 3 features: age, a previous history of chronic obstructive pulmonary disease and the PaO2/FiO2 ratio at the time of arrival to the hospital. The inclusion of additional information from 4 routine blood tests (Creatinine, Glucose, pO2 and Sodium) increases predictive accuracy to 94.5%.

Introduction

Coronavirus disease (COVID-19) first appeared in China in November 2019 and it is caused by the new coronavirus SARS-CoV-2. Its spreading was immediately very fast, prompting WHO to declare a pandemic status on March 12, 2020. Currently, COVID-19 counts over 116M accumulated cases worldwide with more than 2.5M deaths [1]. Early identification of critical patients is an imperative challenge for triage systems as the severity of cases is putting great pressure on hospital resources, leading to the need of alleviating the shortage of medical assets [2]. This is especially important during COVID-19 pandemic when the access to the

relatively less available supportive care and intensive care units (ICUs) is more frequent and timely interventions may reduce mortality.

Since the pandemic outbreak, several algorithms have been proposed to evaluate predictive scores helping in management of COVID-19 patients. These systems are based on both standard statistical approaches and Machine Learning (ML) methodologies [3–5]. The application of supervised ML algorithms to medical research has become widespread in the last years. It has been therefore natural to see a continuation of this trend in the analysis of COVID-19 clinical data, also thanks to the generally superior predictive performance of ML-based solutions over classical statistical approaches. However, the adoption of ML models is also known to be subjected to many pitfalls due to their black box and data-driven nature. In fact, the feckless application of supervised ML often results in overfitting problems, confounding bias effects and scarce interpretability; the latter being particularly troublesome for clinical practice. These phenomena have effectively been noted by a recent meta-analysis [6] where predictive models for COVID-19 are defined to be "poorly reported, at high risk of bias, and their reported performance is probably optimistic".

Another critical issue that needs careful consideration is the fact that commonly adopted ML approaches to diagnosis or outcome prediction are almost uniquely from the family of the supervised learning models. As such, they work on a purely associative approach, by identifying correlations between input data, such as symptoms, clinical history and treatments, and the target variable. Many examples show that the inability to distinguish correlation from causation can produce classifications that are misleading to the point of being potentially dangerous [7]. For instance, the paper by Caruana *et al.* [8] describes a model to predict the probability of death by pneumonia, trained on medical records of patients who have previously had pneumonia. Counter-intuitively, the model found that asthma lowers the risk of death, while it is known to be a severe condition in subjects with pneumonia. This misleading effect occurred because the patients with asthma in the training set received more care by the hospital system. In this example, the association learned from the dataset was correct, but clearly the aim of this application was to find only causal relations useful to prioritize care for patients with pneumonia.

A good option to avoid this kind of problems is given by Bayesian Structure Learning (BSL) approaches [9], which allow to learn from data a probabilistic directed acyclic graph (DAG), called Bayesian Network (BN), connecting the elements that according to the Bayes' rule have a causal relationship. With respect to supervised ML models, BSL has several features making it suitable for medical research [10, 11], such as:

- it is intrinsically interpretable given that it produces a graph representing the causal relationships among the input data;

- it allows to include domain knowledge [12], i.e. it can take into account a set of a-priori known mandatory and prohibited connections. This is a useful feature to combine consolidated medical knowledge with the ongoing research;

- it naturally avoids overfitting [13] and it is less sensitive to noise, because of its non-supervised nature and because it relies on a statistically based definition of conditional probability, instead of finding an arbitrary complex pattern fitting the data;

- it has shown good performances also with small sized dataset [14], a common situation in medical applications.

On a less rigorous note, it has been observed that building causal graphs with BSL forces the researchers to think clearly about the topic, and articulate that thinking in the form of the graph, which is often beneficial in and of itself [12, 15, 16].

All these features make BSL an optimal tool for medical research. In fact, it has found good application in the analysis of genetic data [17, 18], electronic health record time series [19], longitudinal and standard clinical data [20, 21]. In particular, the knowledge of causality relationship between potential predictors and mortality among patients diagnosed with COVID-19 would be crucial to target patients at highest risk and improve outcomes [22]. However, only few works applied Bayesian approaches to the study of COVID-19, mainly to analyse its spreading [23–25] and for contact tracing [26, 27]. One paper interestingly applied BSL to serological tests of COVID-19 to determine a more accurate estimation of the disease's infection and fatality rates taking into account the reliability of each kind of test [28].

Following these considerations, the main objective of this paper is to introduce the use of a BSL-based methodology to study the causal relationship between various predictors and mortality in COVID-19 patients. We aim at verifying whether BSL can reproduce and/or improve our current understanding of causal pathways of COVID-19 risk of mortality using a single-centre but feature-rich clinical dataset [29] which includes demographic information, clinical history of the patient, symptomatology, blood analysis data on admission and outcome. Our analysis is conducted both with and without incorporating prior medical information, to test whether there are significant differences. Following up the causal analysis, we leverage the features that are deemed to be causally related to the outcome to construct a Binary Decision Tree (BDT) that describes a practical flow chart to identify patients at high-risk using data available right at the admission in hospital.

In conclusion, the present work addresses the need of explainable and causally grounded approaches to learn clinical useful knowledge from retrospective data, which is a mainly unexplored research topic, especially in the emergency context of COVID-19. This is achieved by presenting a novel clinical data analysis framework that provides a multi-step approach that goes from the identification of causal relationships between medical data, using BSL, to the realization of a clinically suitable tool, based on a BDT. This tool is suitable for patients triage and is easy to use, straightforward to interpret and has solid clinical bases behind its predictions.

Materials and methods

Dataset description

The dataset includes COVID-19 diagnosed patients admitted between the 3rd of March 2020 and the 30th of April 2020 from three different units of the Pisa University Hospital (Emergency Room, Emergency Medicine Department and ICU) [29]. All data were acquired from both paper and electronic records and carefully checked for the presence of spurious and/or erroneous inputs.

Data collection was performed according to the principles stated in the Declaration of Helsinki and it conforms to standards currently applied in our country. No minors were involved in the study. The use of the data was approved by the Comitato Etico Area Vasta Nord Ovest (Internal Review Board) under the IRB number 230320 and the patient's informed consent was obtained in conformity with safety protocols adopted in the hospital during the emergency. In particular, oral consent was obtained and recorded in the medical chart by nurses from patients at the moment of admission. This procedure was approved by IRB in face of the extraordinary conditions present at the time of the first wave of COVID-19.

The main outcome of the data cohort is dismissal at home or death. The subjects included in the study are 265, of whom 71 died for COVID-19 or for its complications. The death rate of this sample is thus 26.8%. The original dataset included 125 variables. Among the various features collected during the triage and the hospitalization, 63 are missing in less than 50

subjects and thus are included in the study. For the purpose of this study, these features are grouped into six different logical categories as illustrated in Table 1. All the medical exams used in this analysis have been made on the first day of hospitalization.

Causal learning for COVID-19 data

The methodology put forward in this paper is articulated in three steps, leveraging different machine learning and statistical methods. First we apply BSL to analyze how the variables of the sample are causally interconnected. Then, we evaluate with Bivariate Statistical Tests (BST) the strength of the connections between the outcome and its neighbours in the BSL graph. This second analysis may seem redundant with respect to the first step. However, while BSL performs a multivariate analysis taking into consideration also latent variables and the possibility of mediated connections, the second one evaluates singularly how each variable affects the outcome. Thus, on the one hand, the causal graph obtained with BSL helps to understand the logical cause-effect sequence bringing to the outcome and to select the features of interest; on the other hand, BST helps to rank the dependency between each feature and the outcome. The third step targets selecting the most relevant features to train a Binary Decision Tree (BDT) useful for clinical practice.

The following sections provide a synthetic background on Bayesian Networks, describe the BSL algorithm used and the workflow of the three analyses discussed above.

Bayesian Network (BNs) background. BNs are graphical models (G, P), where P is a joint probability of random variables $X_V = (X_1, \ldots, X_N)$ associated with nodes $V = \{1, \ldots, N\}$. Each random variable X_i can possibly be of different nature (binomial, multinomial, ordinal, continuous). The graph $G = (V, E)$ is a directed acyclic graph (DAG) whose edges E encode the joint probabilistic relationships among the N random variables. The graph is a visual representation of the joint distribution of the data, where a directed edge e_{ij} from node i to j indicates that i is the *parent* of j as part of a conditional dependence relationship between the two nodes [30].

Broadly speaking, there are two main families of methods to learn the structure of a BN from data: constraint-based methods and score-based ones [31, 32]. The first class of methods learns the conditional independence relations of the BN from which, in turn, it generates the network. Score-based approaches, instead, cast structure learning as an optimization problem, often addressed as an heuristic search task leveraging a score function to drive exploration of the space of the graph structures in search of the optimal DAG.

Causal graph analysis. *BSL algorithm.* In order to single out the best approach for our analysis, we have considered the fact that constraint-based algorithms have been shown to be more accurate than score-based algorithms for small sample sizes, and that score based algorithms tend to scale less well to high-dimensional data [33]. Additionally, constraint-based algorithms can naturally integrate domain knowledge in the form of already-known results of independence tests, while it is more difficult to include this kind of information in score-based methods. Since we would like to incorporate prior knowledge in our analysis, and since the single samples in our dataset are highly dimensional while the dataset has a relatively low number of subjects, in this paper we use a constraint-based algorithm called Fast Causal Inference (FCI) [34]. FCI is an extension of the popular PC [35] algorithm that considers also the presence of hidden variables, that are random variables for which no observable input data is available. Briefly, the FCI algorithm comprises two steps:

- Skeleton definition. The skeleton search phase consists in finding the undirected graph that has the same edges as the true DAG but no edge orientations. The search starts with a complete undirected graph. Then, each pair of variables is tested for mutual independence

Table 1. Summary of the dataset features. In this table the 63 features analyzed in this work, reported in the second column, are grouped into 6 classes, listed in the first column. The third and fourth columns show the feature occurrence in the dataset in dead and recovered subjects, respectively. The last column reports a description of the feature values: for continuous variables the 5th and the 95th percentiles indicated as $[P_5, P_{95}]$ are reported, while for categorical variables their values are reported within curly brackets.

Category	Feature	Available data for dead subjects (max 71)	Available data for recovered subjects (max 194)	Values
Demographic data	Age (years)	71	194	$[P_5, P_{95}] = [40, 89]$
	SEX	71	194	{M, F} = {0, 1}
	Smoke(y/n)	66	187	{No, Yes} = {0, 1}
	Smoke(ex/y/n)	64	187	{Yes, Ex, No} = {0, 1, 2}
Prior respiratory problems	COPD (chronic obstructive pulmonary disease)	71	194	{No, Yes} = {0, 1}
	Asma	70	194	{No, Yes} = {0, 1}
	Other resp. disease	71	194	{No, Yes} = {0, 1}
Prior diseases	Diabetes	71	194	{No, Yes} = {0, 1}
	Hypertension	71	194	{No, Yes} = {0, 1}
	Cardio.disease	71	194	{No, Yes} = {0, 1}
	Hypercolest	71	194	{No, Yes} = {0, 1}
	Cerebrovasc. disease	71	194	{No, Yes} = {0, 1}
	Neuro. disease	71	194	{No, Yes} = {0, 1}
	Dementia	71	194	{No, Yes} = {0, 1}
	Cancer	71	194	{No, Yes} = {0, 1}
	Blood cancer	71	194	{No, Yes} = {0, 1}
	Kidney disease	71	194	{No, Yes} = {0, 1}
	Liver disease	71	194	{No, Yes} = {0, 1}
	Cirrhosis	71	194	{No, Yes} = {0, 1}
	Autoimmune disease	71	194	{No, Yes} = {0, 1}
Ongoing treatments	Anticoag	71	194	{No, Yes} = {0, 1}
	RAAS BLOCK (renin-angiotensin-aldosterone system)	71	194	{No, Yes} = {0, 1}
	Immunos. therapy	71	194	{No, Yes} = {0, 1}
	Dialysis	71	194	{No, Yes} = {0, 1}
Symptoms on admission	Fever	71	194	{No, Yes} = {0, 1}
	Conjunct. congest.	71	194	{No, Yes} = {0, 1}
	Nasal congestion	71	194	{No, Yes} = {0, 1}
	Headache	71	194	{No, Yes} = {0, 1}
	Cough	71	194	{No, Yes} = {0, 1}
	Sore throat	71	194	{No, Yes} = {0, 1}
	Sputum	71	194	{No, Yes} = {0, 1}
	Fatigue	71	194	{No, Yes} = {0, 1}
	Hemoptysis	71	194	{No, Yes} = {0, 1}
	Short breath	71	194	{No, Yes} = {0, 1}
	Nausea	71	194	{No, Yes} = {0, 1}
	Diarrhea	71	194	{No, Yes} = {0, 1}
	Myalgia	71	194	{No, Yes} = {0, 1}
	Rash	71	194	{No, Yes} = {0, 1}
	FC (cardiac frequency, bpm)	63	176	$[P_5, P_{95}] = [61, 119]$
	PAS (systolic arterial pressure, mmHg)	65	174	$[P_5, P_{95}] = [100, 170]$
	PAD (diastolic arterial pressure mmHg)	65	174	$[P_5, P_{95}] = [55, 100]$
	Chest pain	71	194	{No, Yes} = {0, 1}
	Confusion	71	194	{No, Yes} = {0, 1}

(*Continued*)

Table 1. (Continued)

Category	Feature	Available data for dead subjects (max 71)	Available data for recovered subjects (max 194)	Values
Blood analysis on admission	Haemoglobin (g/dl)	68	188	$[P_5, P_{95}] = [9.3, 16]$
	WBC (white blood cells, cells/nl)	69	192	$[P_5, P_{95}] = [3.37, 15.78]$
	Lymphocyte (cells/μl)	68	191	$[P_5, P_{95}] = [370, 2380]$
	Neutrophils (cells/μl)	69	188	$[P_5, P_{95}] = [1740, 13540]$
	Haematocrit (%)	69	187	$[P_5, P_{95}] = [27.5, 46.4]$
	Platelets (cells/nl)	69	191	$[P_5, P_{95}] = [83, 423]$
	INR (international normalized ratio)	65	184	$[P_5, P_{95}] = [1.03, 1.75]$
	Bilirubin (mg/dl)	64	186	$[P_5, P_{95}] = [0.24, 1.6]$
	AST (aspartate aminotransferase, IU/l)	59	181	$[P_5, P_{95}] = [14, 115]$
	ALT (alanine aminotransferase, IU/l)	66	187	$[P_5, P_{95}] = [10, 125]$
	Glucose (mg/dl)	64	186	$[P_5, P_{95}] = [84, 233]$
	Creatinine (mg/dl)	67	191	$[P_5, P_{95}] = [0.62, 2.56]$
	BUN (blood urea nitrogen,mg/dl)	58	178	$[P_5, P_{95}] = [9, 74]$
	Sodium (mEq/l)	68	190	$[P_5, P_{95}] = [131, 145]$
	Potassium (mmol/l)	67	187	$[P_5, P_{95}] = [3.21, 5.04]$
	pH	57	162	$[P_5, P_{95}] = [7.35, 7.53]$
	pO2 (O2 partial pressure, mmHg)	62	173	$[P_5, P_{95}] = [40, 143]$
	pCO2 (CO2 partial pressure, mmHg)	61	167	$[P_5, P_{95}] = [24, 45]$
	PF (pO2/FIO2 ratio, %)	65	188	$[P_5, P_{95}] = [126.25, 450]$
	PCR (C-reactive protein, mg/dl)	67	177	$[P_5, P_{95}] = [0.48, 27.64]$
Outcome		71	194	{Death, Recovery} = {0, 1}

considering also all the subsets of the other variables. When two variables are found to be independent given at least one subset of other variables, the edge connecting them is removed. In this step we see how easy it is to include knowledge about prohibited or mandatory connections between variables: the information supersedes the result of the independence test. Different independence tests can be used for different kinds of variables.

- DAG learning. In this step, the direction of the edges is estimated. During this operation hidden confounders are also considered, i.e., unobserved variables which are a common cause of at least two observed variables [36].

In our analysis, variable independence is tested with a Gaussian conditional independence test for continuous variables, and the G^2 test for categorical variables [34]. Other parameters relevant for the reproduction of this study are reported in S1 Table.

A significant issue with constraint-based algorithms is that the number of independence tests needed to remove an edge increases significantly with the number of vertices and with the number of values assumed by the random variables (for the discrete case). For this reason, heuristic search algorithms are usually employed to avoid testing a combinatorial number of cases [34]. As a consequence, the true DAG can seldom be identified and multiple DAGs with different edge directions are found to be compatible with the input data. For this reason, in our analysis we do not consider edge orientation by the algorithm as reliable information, leaving this task to the clinician interpretation.

Quantifying causal effects. For every DAG identified by FCI, we estimate the effect that each variable has on the variables it is connected to, representing an indicator of the strength of

their causal connection, using covariate adjustment. Basically, for each pair of variables, X and Y, we compute a linear regression of Y on X using the parent set of X, extracted from the DAG, as covariate adjustment. When multiple DAGs are available, we average the effects estimated from each DAG. However, in this study we have observed that the DAG is usually estimated without ambiguity and thus the cases in which multiple different effects are found are very rare. Note that this method describes the effect that a unitary increase in X has on Y and, therefore, it is highly sensitive to the range of both X and Y. In order to be able to qualitatively compare the effects, we standardized the variables before this calculation, applying a Z-score transformation.

BSL strategy on our cohort. The application of the BSL pipeline described in the previous subsection followed some considerations related to the nature of the dataset available in our study. First, it has to be noted that only a portion of the features in Table 1 has been collected or is available for all subjects in our dataset. As most data-driven methods, BSL performs better when the sample size is higher than the number of features. Therefore, we opted for applying the BSL pipeline separately for the 6 categories of features identified in Table 1, in order to maximize the amount of samples available. The outcome variable has been included in each category-specific analysis. After building the BNs (one per category), we identified those features of each skeleton that reported either a direct causal connection with the outcome, or an indirect connection mediated by a single intermediate feature. These served to define the candidate pool of relevant features which have been jointly analyzed with another round of the BSL algorithm (i.e. representing samples only in terms of the filtered features) leading to an integrated causal graph.

The BSL algorithm attempts to infer causal relationships from input data, which are unavoidably biased, because the sample analyzed is composed by subjects whose medical conditions required hospitalization. Thus, the results from this analysis should be interpreted as representative of the most severe situations.

The causal graph analysis pipeline discussed so far has been repeated under two different conditions. In the first we provide to the BSL process a conservative list of prohibited causal connections (drawn up by our clinicians and based on the fact that there is vast medical agreement on the nonexistence of such relationships), whose severance is enforced during the graph identification process. The second one, instead, comprises a fully data driven setting, where no prior knowledge is supplied to the BSL algorithm. This second condition is tested for comparison, to assess the capability of the method to autonomously identify causal relationships that are not blatantly in contrast with widely agreed clinical knowledge.

Bivariate statistical analysis. The focus of this step of our analysis is to study how the outcome depends singularly from each relevant feature identified at the previous stage. To this end, we use the Fisher Exact Test (FET) [37] for categorical variables and the Point-Biserial Correlation (PBC) [38] for the continuous ones. The FET method tests whether two binary variables are independent, such as "death-recovery" and "male-female". The PBC, instead, is a correlation test, such as the Pearson's correlation, specifically developed to correctly estimate the correlation between a categorical variable, in this case the outcome, and a continuous variable. The bivariate statistical analysis is applied to all the features considered in the integrated causal graph.

Decision tree analysis. BDT models are characterized by an inferential process which has some resemblance with human reasoning, which makes them amenable to human interpretation. However, they are known to severely suffer from overfitting [39], thus it is fundamental to reduce the complexity, and thus the depth, of the tree to avoid the identification of poorly generalizable classification rules.

The scope of our BDT analysis is two-fold. On the one hand we would like to identify a compact BDT to provide an interpretable and procedural view on what are the discriminative cut-point values for the features identified by the BSL causal analysis, and if those values are in agreement with clinical knowledge. On the other hand, we would like to measure the predictive value of such candidate features on our dataset.

The first objective (interpretation) has been addressed by training a BDT with a depth fixed to 4 levels, considering only those features present in the integrated graph and only those subjects having those features fully available. Other BDT parameters relevant to reproduce this work are recapitulated in S1 Table.

The interpretable BDT is trained on the full data (train-all setting), as it is never used for prediction. To assess the impact of the identified causal features on predictive performance, we train two additional models. The first is a BDT with the same structural constraints and features of the intepretable one, but it is trained and assessed in a 10-fold cross-validation scheme. Its predictive performance is assessed as the average accuracy on the 10 validation folds. The second is a permutation test BDT, again trained and evaluated in 10-fold cross-validation, but using features randomly drawn from the set of features in Table 1.

The comparison between the last two BDTs is intended to verify the effectiveness of the feature identified by the BSL causal analysis. However, a fair comparison requires to ensure that the two BTDs have a comparable number of subjects in training, otherwise it is impossible to understand whether a performance reduction is to be imputed to the lower training size or to the different features. Thus, in the permutation test, we train about 1000 different BDTs of depth-4 with number of training subjects and features matching those used for training the causal BDT.

As already said at the beginning of this paragraph, the choice of using a BDT is mainly motivated by its interpretability and ease of clinical usage, but at the same time we want to maximize the prediction performance, otherwise the benefit deriving from an interpretable tool would be less useful. Thus, we performed an additional analysis to compare the performance of the BDT with 3 commonly adopted machine learning algorithms: logistic regression (LR) and support vector machine (SVM) with linear and polynomial kernels. In this analysis the algorithms are trained only with the same features used by the BDT, following the same steps of the BDT analysis: train-all, 10-fold cross-validation and permutation test using random features.

Results

We report the main results of the empirical analysis run on our COVID-19 dataset following the three-fold structure of our methodology. A pictorial overview of the experimental analysis is provided in Fig 1. Here, we highlight the explorative causal analysis run separately on the different feature categories (Step I), followed by the analysis on selected causal features (Step II) comprising the integrated causal graph, the BST and interpretable BDT. The final stage comprises a predictive analysis comparing the BDT on causal features with the BDT on randomly drawn features (Step III).

Causal graph analysis

The results of the BSL step run separately on the single feature categories are provided in Fig 2, while Fig 3 depicts the corresponding integrated graph. All graphs have been obtained including prior knowledge on prohibited connections through a list provided by our physicians. Their fully data-driven counterparts are reported in S1 and S2 Figs in the Supporting information to ease readability. One can easily appreciate that no substantial difference exists between
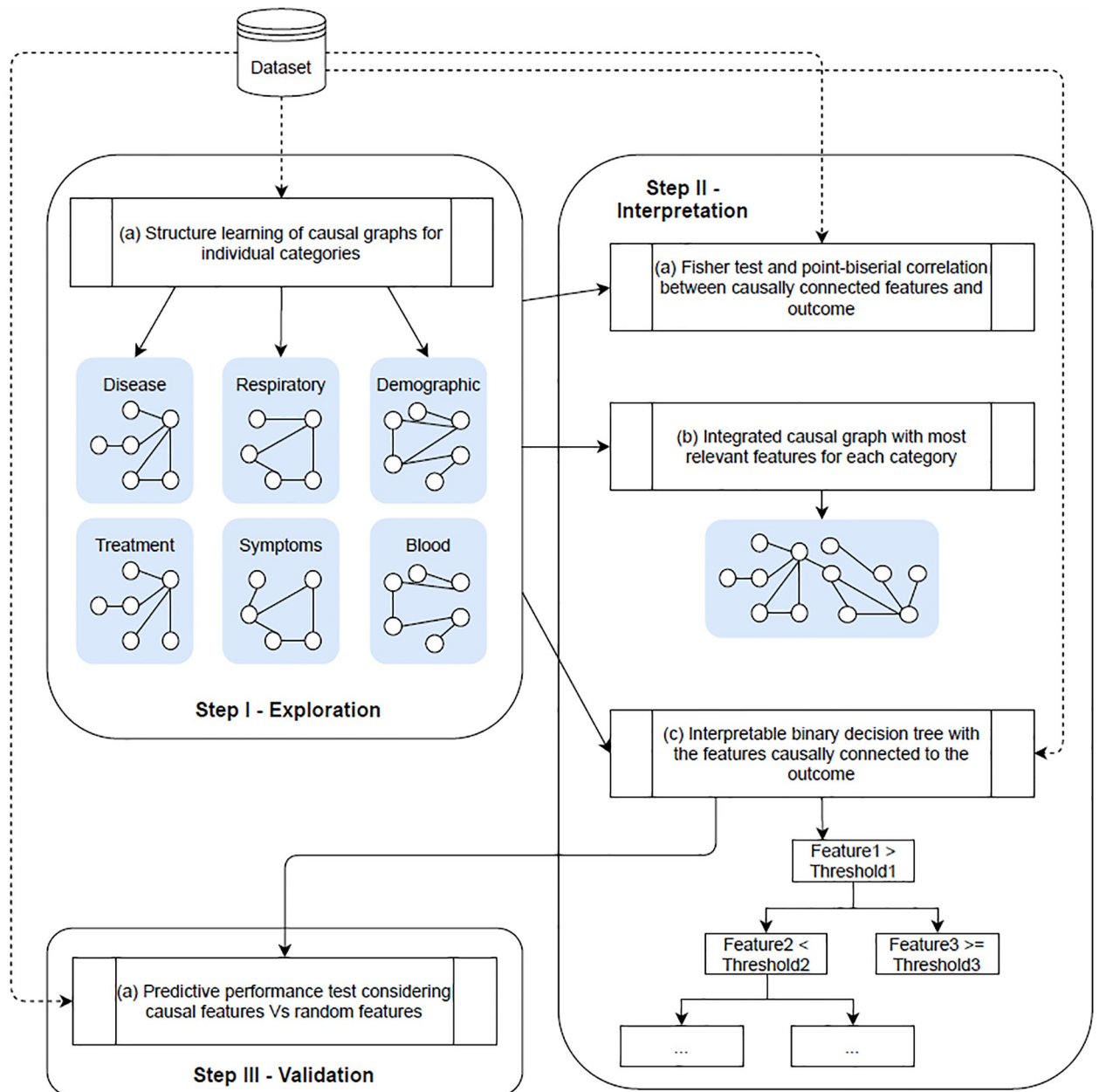
Fig 1. Flow chart of the proposed approach. Our empirical analysis is mainly based on three steps: (1) an explorative analysis applied to different classes of features separately; (2) an integrative and interpretative step; (3) a quantitative validation.

the two sets of graphs, confirming the robustness of the causal inference process. Note that the distance between nodes in the graphs has no meaning and it is automatically adjusted to make the diagrams visually appealing. Instead, the thickness of the edges reflects the intensity of the connection and the red (blue) edges connect positively (negatively) correlated variables. The definition of positively/negatively correlated variables may be confusing when one variable is categorical and the other is continuous. For instance in Fig 3, age (continuous) appears to be negatively correlated with the outcome, which has value 1 for death and 0 for recovery. This means that older subjects have a higher risk of death than younger ones. To fully understand
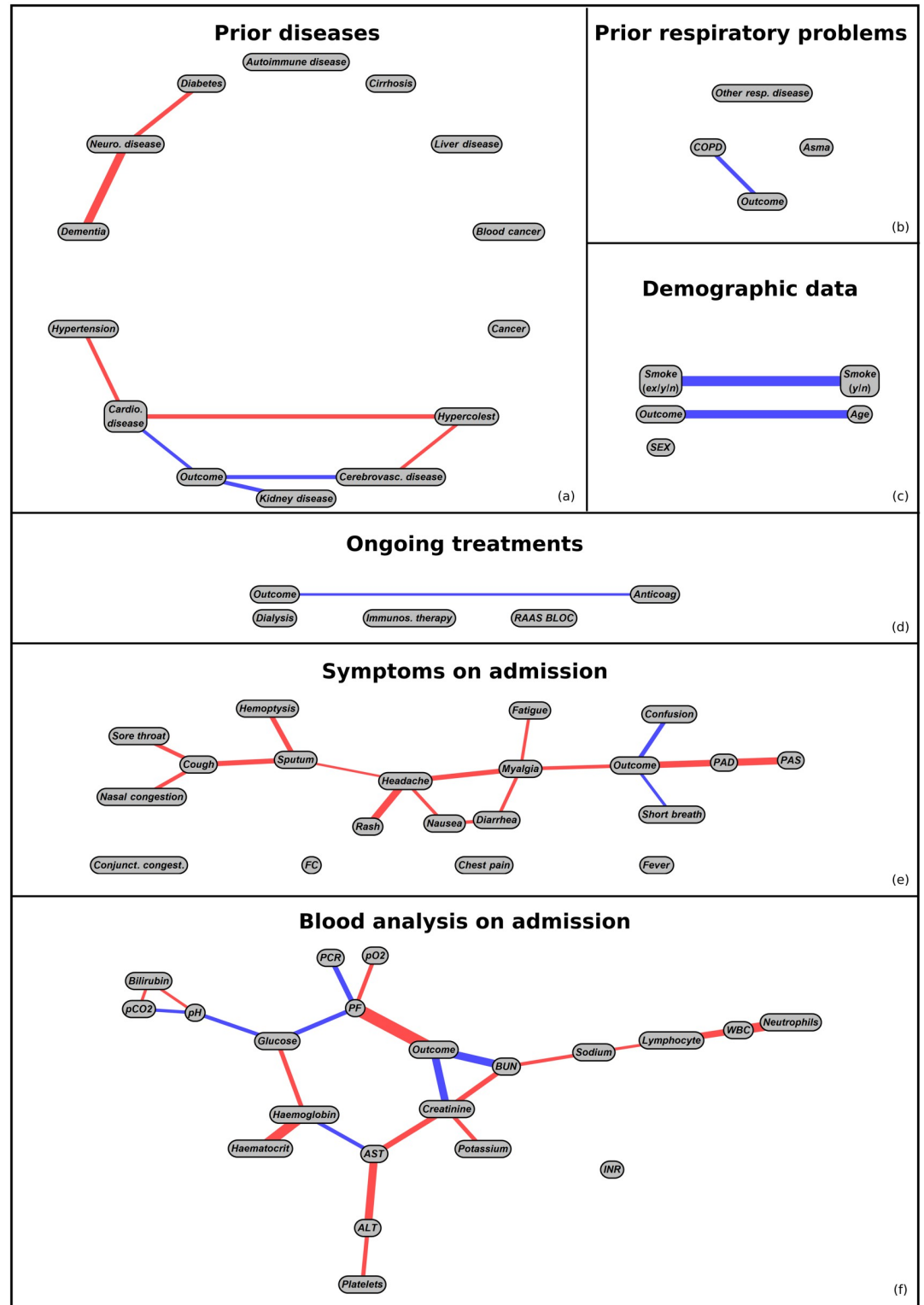
Fig 2. BSL analysis on separate classes of features. Image showing the BSL analysis applied to different categories of features. All the illustrated graphs are generated taking the information provided by clinicians into account.
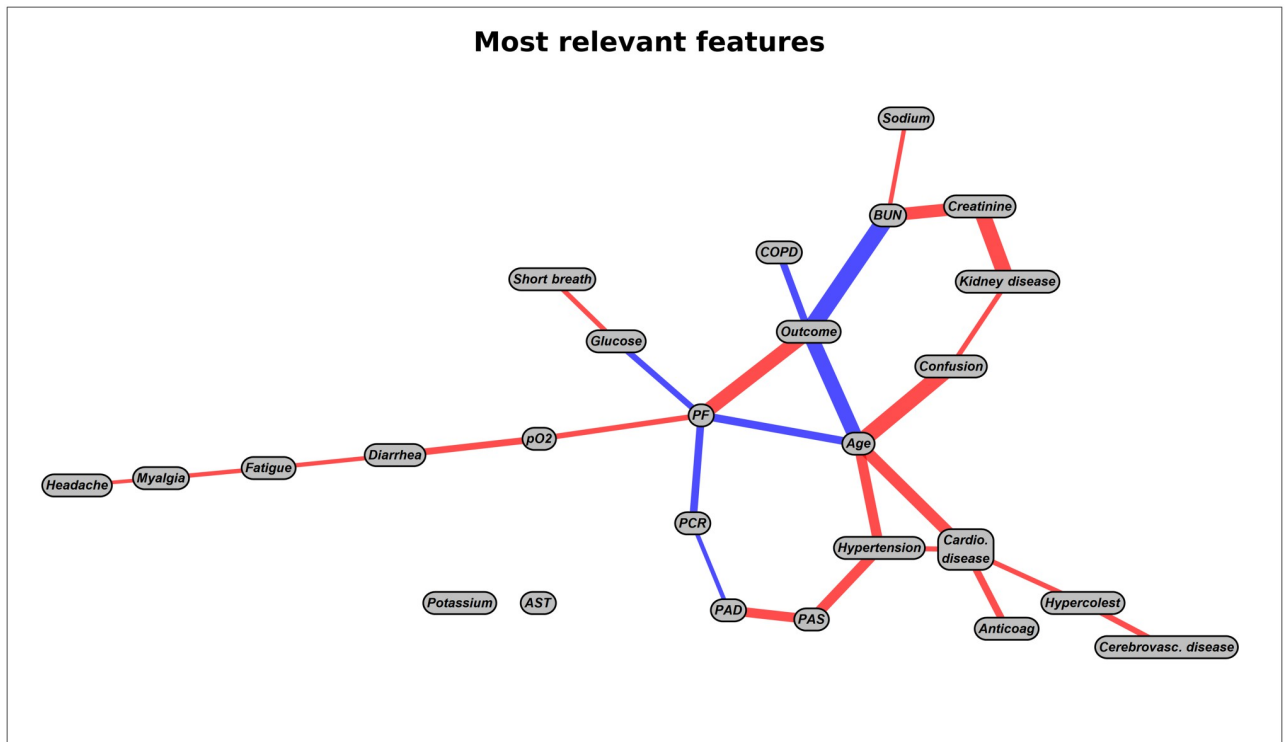
Fig 3. BSL analysis on the most relevant features. Graph generated with the most relevant features found from the graphs shown in Fig 2, taking the information provided by clinicians into account.

https://doi.org/10.1371/journal.pone.0268327.g003

Table 2. Results of the bivariate statistical analysis for categorical variables. Contingency table and results of the Fisher test between the categorical variables present in Fig 3 and the outcome. The most impactful deaths (recoveries) fold increases with respect to the dataset average are reported within the brackets in the second and third columns, respectively.

Feature X	% of deaths in patients with X	% of recoveries in patients with X	% of deaths in patients without X	% of recoveries in patients without X	P-Value (Fisher test)
COPD	69.0% (×2.6)	31.0%	21.6%	78.4%	5.10^{-7}
Kidney disease	58.3% (×2.2)	41.7%	23.7%	76.3%	6.10^{-4}
Cerebrovasc. disease	59.3% (×2.2)	40.7%	23.1%	76.9%	$1.7*10^{-4}$
Cardio. disease	42.5% (×1.6)	57.5%	19.1%	80.9%	$6.4*10^{-5}$
Anticoag.	51.5% (×1.9)	48.5%	23.3%	76.7%	$1.1*10^{-3}$
Myalgia	2.7%	97.3% (×1.3)	30.7%	69.3%	$6.1*10^{-5}$
Confusion	71.9% (×2.7)	28.1%	20.6%	79.4%	$1.4*10^{-8}$
Short breath	34.1% (×1.3)	65.9%	20.1%	79.9%	$7.5*10^{-3}$
Dialysis	50.0% (×1.9)	50.0%	26.4%	73.6%	0.29
Hypercolesterolemia	37.0% (×1.4)	63.0%	24.7%	75.3%	$6.6*10^{-2}$
Hypertension	35.5% (×1.3)	64.5%	19.1%	80.9%	$2.1*10^{-3}$
Diarrhea	11.8%	88.2% (×1.2)	30.4%	69.6%	$4.0*10^{-3}$
Fatigue	23.3%	76.7%	27.5%	72.5%	0.36
Headache	4.3%	95.7% (×1.3)	28.9%	71.1%	$5.5*10^{-3}$

https://doi.org/10.1371/journal.pone.0268327.t002

Table 3. Results of the bivariate statistical analysis for continuous variables. Point-biserial correlation values and significance tests between the continuous variables present in Fig 3 and the outcome.

Feature	Correlation	P-Value
Age	-0.46	1.3^*10^{-15}
PAS	0.21	4.6^*10^{-4}
PAD	0.24	8.1^*10^{-5}
AST	-0.11	0.052
Glucose	-0.19	1.5^*10^{-3}
Creatinine	-0.20	5.6^*10^{-4}
BUN	-0.45	1.1^*10^{-13}
Sodium	-0.16	5.0^*10^{-3}
Potassium	-0.18	1.6^*10^{-3}
PCR	-0.21	4.4^*10^{-4}
pO2	0.12	0.032
PF	0.46	1.1^*10^{-14}

https://doi.org/10.1371/journal.pone.0268327.t003

the meaning of each connection, we refer to Table 1 which reports the values of each categorical variable under examination.

Bivariate statistical analysis

Table 2 reports the contingency table and the p-values of the Fisher exact test computed for all the categorical variables found relevant in the single-category BSL analyses and thus included in the integrated causal graph in Fig 3. With *relevant* we mean that these features are connected to the outcome within 2 hops. To ease intepretation of the results we report also, in the second (third) within the brackets, the death (recovery) excess factors with respect to the occurrence in the whole dataset.

Table 3 instead reports PBCs and their p-values. As it can be noted, these measures emphasize some differences with respect to the BSL-based analysis. For instance, in the integrated graph fatigue and headache seem both to be linked to the outcome through myalgia. The Fisher analysis instead reveals that headache is unrelated to the outcome and has to be simply related to myalgia, while fatigue seems to be actually causally connected to the outcome.

Decision tree analysis

Fig 4 shows the BDT trained with all the 198 subjects having the 26 features included in the integrated causal graph Fig 3. The interpretable BDT chose to base its prediction exclusively on 7 features: AGE, PF, COPD, Creatinine, Glucose, pO2 and Sodium. Remarkably, all of them are connected to the outcome within 2 hops in the integrated causal graph (Fig 3). When computing the classification accuracy of the interpretable BDT on its training set, we appreciate that only 5.5% of the data are misclassified. Despite a few leaves being clearly tailored to fit the dataset (and thus being good candidates for pruning), all the learned patterns appear to be in line with current clinical understanding of how these features affects the outcome.

The first line in Table 4 shows various performance metrics of the BDT classifier evaluated in train-all setting (i.e. trained and validated on the same data comprising all the available subjects) and in a 10-fold cross-validation assessment. As it can be noted, the results are quite coherent, suggesting that the overall pattern is really informative and almost insensitive to overfitting. The permutation test, whose results are reported in the second line of Table 4, shows instead a completely different behaviour. Even with randomly selected features it is
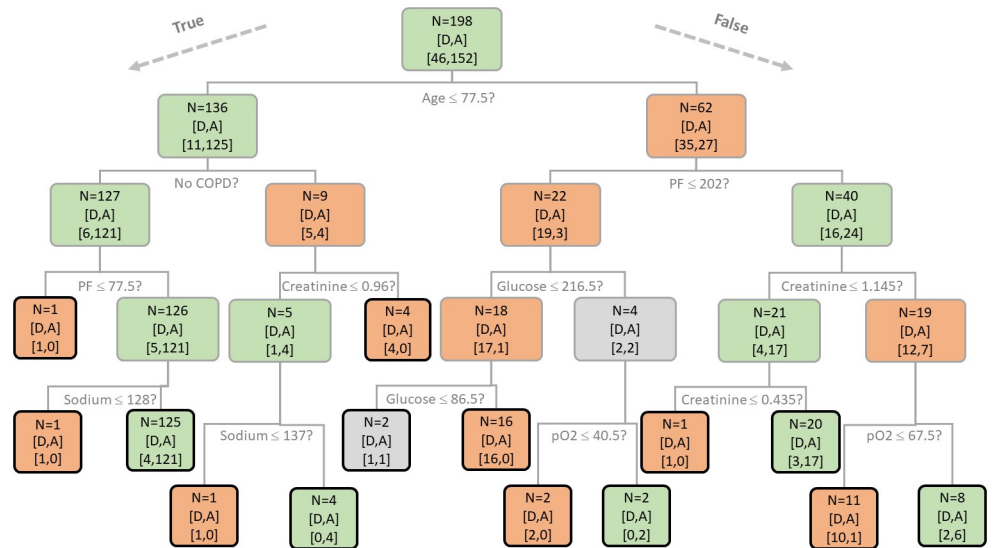
Fig 4. BDT obtained with a train-all setting. BDT trained with the features included in the graph analysis reported in Fig 3. The color of the squares represents the class prevalence: red for deaths, green for recoveries and grey in case of parity. The number of subjects is indicated in the first line of each square, while the last line reports deaths/recoveries. Black edges denote leaves.

https://doi.org/10.1371/journal.pone.0268327.g004

possible to achieve high performances in train-all setting, but these performances are not maintained in cross-validation.

Fig 5 depicts the average proportion of misclassified patients in cross-validation of our BDT (red line) and of the permutation test (orange columns), showing that the errors in the former tree are significantly lower from what would be obtained from one considering 7 features randomly taken from the dataset.

Finally, the additional comparison with less interpretable but well regarded algorithms such as LR and SVM (reported in S2 Table of the supplementary materials) shows that these algorithms reach overall inferior performance. In fact, they are strongly misled by the bias in the number of recoveries with respect to deaths and tend to classify most of the subjects as recovered. Furthermore, they are more prone to overfitting, as shown by the performance loss when moving from a train-all configuration to a 10-fold cross-validation. The permutation test results, reported in S2 Table and in the S3–S5 Figs show that also for these algorithms, the 7 features selected by our pipeline allow a better classification of the subjects with respect to a random selection of other 7 features.

We believe that these results support the methodological choice to select the features causally connected to the outcome through BSL and then training a BDT to obtain an intepretable clinical decisional tool.

Table 4. BDT performance evaluation. Comparison between the performance of the classifier developed in this study and classifiers trained on a random set of 7 features. The results of the permutation tests are the average of those obtained from 1,000 permutations.

Input data	Sensitivity		Specificity		F1 score	
	Train all	10f cv	Train all	10f cv	Train all	10f cv
7 feat of the tree in Fig 4	0.99	0.90	0.95	0.95	0.97	0.92
7 random feat	0.99	0.88	0.82	0.49	0.97	0.86

https://doi.org/10.1371/journal.pone.0268327.t004
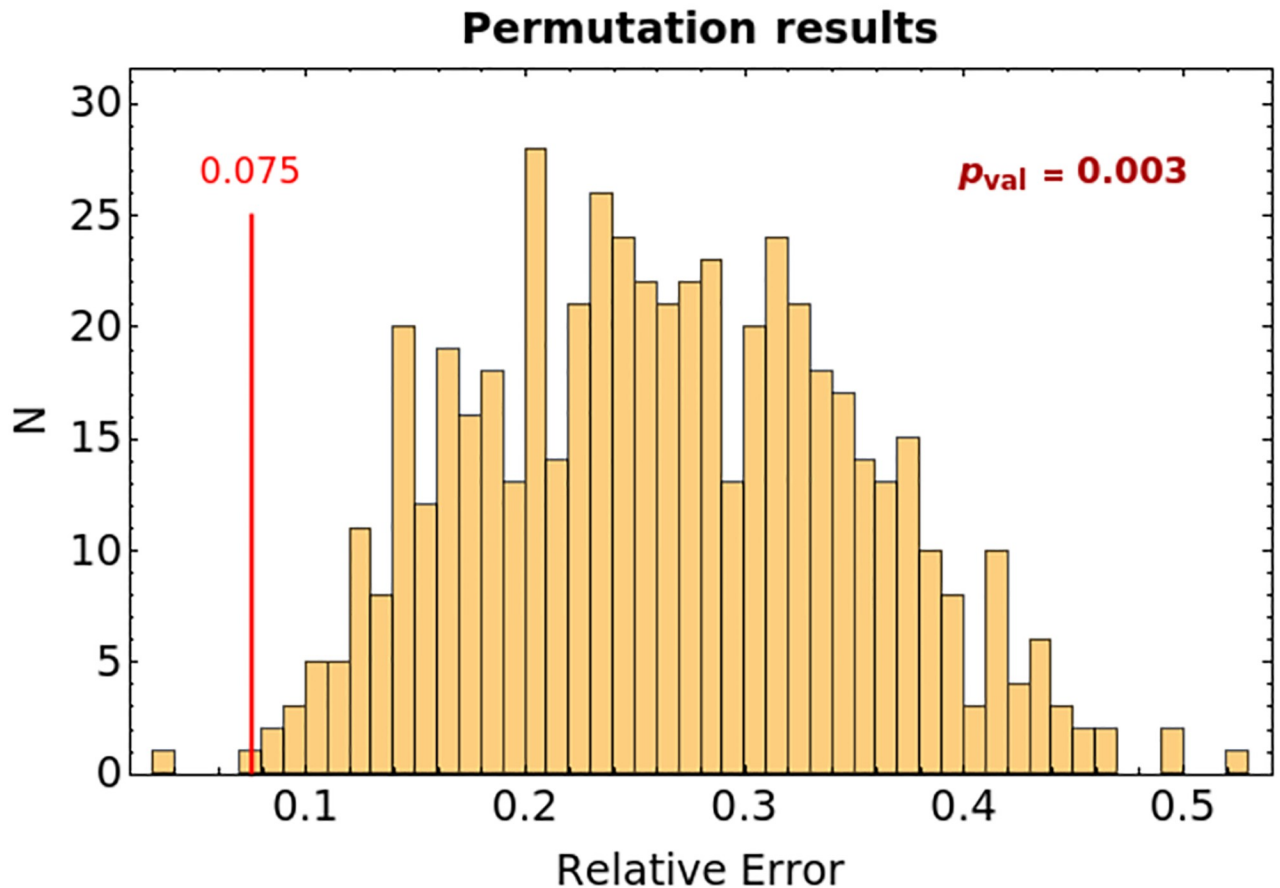
Permutation results

Fig 5. BDT permutation evaluation. Histogram reporting the percentage of misclassified subjects in the permutation test. The red bar shows the performance of the BDT reported in Fig 4. All the misclassification rates are calculated as the average over a 10-fold cross-validation.

Discussion

In this section, we briefly discuss the experimental results obtained by our causal analysis. We begin by considering the results from the single-category causal graphs in Fig 2.

Among respiratory diseases (graph b), COPD is the only pathology in direct causal relationship with mortality. This result was recently confirmed by a meta-analysis on the associations between respiratory diseases and COVID-19 outcomes [40].

As regards prior (non-respiratory) diseases (graph a), not surprisingly, cardiovascular disease has a direct relationship with the outcome [41]. Furthermore, it is the mediator of the causal effects of hypertension and hypercholesterolemia. On the other hand, hypercholesterolemia also has a causal effect on the presence of cerebrovascular disease which has a direct impact on the outcome [42]. From graph a also kideny diseases appear to be directly connected to the outcome, as widely documented [43]. Remarkably, the three pathologies directly connected to the outcome (i.e. cardiovascular, cerebrovascular and kidney diseases) exert a strong effect on it, increasing the probability of death by 1.6, 2.2 and 2.2 with respect to the average in the dataset, respectively (see the analysis in Table 2). In contrast to published results, in our analysis liver diseases (including cirrhosis) [44], diabetes [45] and the presence of dementia or neurological diseases [46] seem to be not causally connected to the outcome. This discrepancy is mainly imputable to their scarce representation in the dataset, which does not allow to satisfy

the conditional probability criterion with which the graph connections are identified. In fact, despite their limited size, in all the subsets of subjects presenting the aforementioned diseases, the risk of mortality rises at least by a factor of 2.3. However, these increases cannot be attributed to these diseases alone because at least 75% of the subjects who died and were affected by one of them, presented also one of the three pathologies directly connected to the outcome.

Age is one of the most cited risk factors due to its close relationship with mortality in COVID-19 patients [47]. Accordingly, the BSL approach (see graph c) identifies a strong causal relationship with the outcome also in the patients enrolled in Pisa. The importance of this effect can be appreciated even more by looking at the thickness of the "age-outcome" edge in the integrated causal graph, Fig 3. On the other hand, in our cohort neither sex nor smoking seem to have a direct causal role with mortality, contrary to what was claimed by recent meta-analysis [48, 49]. The apparent inconsistency on the effect of sex is explained by the intrinsically biased cohort which has been selected based on the symptomatology severity of the patients in admission. In fact, only 32% of the examined sample is composed by females, which means that they are somewhat protected from acute forms of COVID-19. Thus, according to our analysis being a female has no protective effect against death when the patient already shows severe symptoms that necessitate hospitalization, while this protective effect is present *ab origin* and it is detectable by the gender bias in hospitalized patients.

The BSL analysis applied to treatment features (graph d) shows also a direct connection between the assumption of anticoagulant before/during Coronavirus infection and the outcome. As it can be noted from Table 2, the prescription of anticoagulants increases the risk of mortality by a factor of 1.9, while the cardiovascular diseases by only 1.6. This effect suggests that the subjects that are prescribed to take anticoagulants have a more severe form of cardiovascular disease, and that the true causal effect is given by cardiovascular disease alone. Following this idea, given that kidney diseases increase the risk of mortality, we would expect a connection between outcome and dialysis, however only 4 subjects of the cohort were on dialysis.

Results from the analysis of the symptoms present at hospital admission (graph e) are particularly intriguing. In fact, only a relatively small number of symptoms show a direct causal relationship with the outcome. Among these, myalgia seems to be protective against the risk of mortality. This fact, which has been recently confirmed by Zheng et al. [50], could be explained by the fact that patients with less severe form of disease are reporting a generic muscle pain with augmented sensitivity as they do not suffer from more severe and debilitating symptoms. Accordingly, myalgia presents upstream causal relationships with many other symptoms of medium and mild severity, such as nasal congestion, sputum and nausea. On the other hand, a strong causal link between confusion and mortality was detected (leading to an increased risk of death by 2.7 with respect to the average in Table 2), probably associated with the older age of patients presenting this symptom. In the same way, the effects of PAD (and indirectly PAS) could be explained by the already demonstrated causal relationship between cardiovascular disease and outcome. In contrast to what is present in the literature, the presence of fever at the time of admission seems to lack a direct causality link with mortality [51], however differences in fever assessment may alter the results of this analysis. For instance, fever may be measured at the hospital at the time of admission or assessed by the patient and in both cases its value may be altered by antipyretic assumption, whose frequency depends on self-medication habits.

Among laboratory tests (graph f), only three parameters have a direct causal relationship with mortality. PF ratio is an index of respiratory function and, given that the respiratory system is one of the systems most affected by COVID-19, its causal relationship with mortality is highly understandable. This connection can be observed both in the single-category as well as

in the integrated causal graph. Other clinical variables selected by our automatic method are BUN and creatinine. Elevated levels of both BUN and creatinine can be associated to chronic kidney disease which was already demonstrated in direct causal relationship with mortality. Furthermore, a link between creatinine, BUN and mortality has already been observed in COVID-19 patients by other authors [52].

The study of causality between classes of clinical variables and mortality in COVID-19 patients revealed some already known relationships as well as offering some new insights. However, the picture becomes more clear when all clinical variables are used together to assess causal effects on the outcome as in the integrated causal graph of Fig 3. In this case, only 4 variables show a direct causal effect on the outcome. However, these direct relationships are very strong, with PF, age and BUN being particularly correlated to mortality (see Table 3). Furthermore, these parameters also exhibit a strong upstream causal relationship with practically all the clinical variables that are directly associated with the outcome in the analysis divided by single categories. In fact, age has causal relationships with confusion, cardiovascular and cerebrovascular diseases, as well as with the use of anticoagulants and high blood pressure. Similarly, BUN levels mask a causal relationship with creatinine levels and the presence of nephrological disease. Finally, the PF mediates the relationship with shortness of breath. On the contrary, myalgia seems to have a weaker and more distant causal role with the outcome when all the clinical variables are considered. This seems to sustain the hypothesis that the apparently causal of myalgia on symptoms is due to the fact that less severe patients report its occurrence together with the deficiency of more critical symptoms.

Moving to the BDT illustrated in Fig 4, remarkably the first two levels of the tree alone allow to correctly classify 85% of the subjects basing on their age, a previous history of COPD in young and middle-aged patients and the PF ratio value in the older ones. The PF threshold discovered by the algorithm to distinguish patients at high and low risk is almost equal to the criterion adopted by the American-European Consensus Conference Committee (AECC) to define hypoxemia, i.e. a PF ratio less than or equal to 200 mmHg [53]. Throughout the tree, creatinine levels have been questioned 3 times. Despite in healthy subjects creatinine normally can range between 0.6 to 1.3 mg/dl, from this analysis it seems that subjects affected by COVID-19 have an increased risk of mortality already when its value passes 0.96 for the younger ones and 1.145 for the older ones. Instead, binary decisions based on sodium and pO2 levels are coherent with normal range values. Sodium below 135 mEq/L indicates hyponatremia that may hint to heart, kidney or liver problems, which notoriously have a negative impact on subjects with COVID-19 and pO2 values below 80 mmHg indicates that a person is not getting enough oxygen [54]. The only binary decision of dubious interpretation from a medical perspective is the one based on glucose value at the third level of the tree. Probably, this is due to the overfitting problem that characterizes BDTs. However, this question and the following ones along its branches do not improve particularly the ability of the BDT to predict the risk of mortality. In fact, already before the debatable question on glucose the set was composed by 19 subjects who died and only 3 who survived.

As it emerges from this discussion, various limitations exists in this study:

- the limited size of the dataset;

- the dataset is unavoidably biased towards locally resident subjects with severe symptoms requiring hospitalization;

- patient-reported data may be very noisy: such as clinical history details, ongoing treatments, symptoms before hospitalization, etc.

Except for the first limitation, all the remaining ones represent common issues in the medical field. However, the coherence between the results obtained with this methodology on such a limited dataset, with the results available in literature, suggests that the implemented approach is resilient to this kind of issues. Overall, a percentage of 5.5% of misclassified subjects using solely data available at the triage suggests that the 3-step methodology presented in this paper could be successfully adopted to design an operative flowchart for clinical use when paired with a larger dataset and after an extensive validation.

Conclusion

Our work took place on the claim that causal learning, and in particular BSL methodologies, provide useful tools for delivering a robust and interpretable exploratory analysis of clinical data. We have put forward an integrated methodology, rooting on causal graph learning, that provides practitioners with a pipeline of analytical steps amenable to

- identify and measure the strength of causal association between features and outcomes in a fully-multivariate fashion;

- select causally relevant features and assess how they impact, separately, the outcome;

- provide a compact and interpretable procedural description of the outcome-determination process to support clinical decision making.

We have demonstrated the effectiveness of our approach on a COVID-19 case study, where we evaluated the causality relationship between clinical parameters and mortality in a cohort of COVID-19 patients, confirming existing results. The analysis was carried out both splitting the clinical variables into classes as well as using them altogether into an integrated causal graph. Our clinical discussion provides compelling evidence of the robustness of the causal relationships identified by our causal learning approach. The results of the causal analysis were further used to build a highly explainable predictive model of mortality as a decision tree that could be easily implemented in a real clinical context. Our empirical validation against a permutation test setting, confirms the quality and relevance of the features identified by our method, also in a predictive setting.

Thus, we propose the adoption of this workflow to discover complex cause and effect relationships in clinical high dimensional dataset, also for other applications. For instance, one possible future work may consist in the study of targeted drugs by applying the presented methodology to the data about the effects of such drugs on subjects with different symptoms, comorbidities, personal and familial clinical history. Within the context of this Coronavirus emergency, we hypothesize this multi-step approach may be useful for investigating how the symptomatology, the care received and the clinical history of the patient impact on COVID long-term effects, which are still very unclear [55, 56].

Finally, future improvements of the methodology itself may rely on the adoption of supervised deep learning models using causal inference [57], that have the advantage to integrally accomplish both the prediction task and the detection of causal relationships.

Supporting information

S1 Table. Important parameters. Parameters used for the BSL and BDT analyses. (PDF)

S2 Table. Performance evaluation of the LR and SVM (with linear and polynomial kernels) algorithms. Results obtained with the SVM (linear and polynomial kernels) and Logistic Regression algorithms. In all the 3 cases, the permutation test using random features shows

significantly inferior results with respect to the features chosen with the causal analysis. Note that the results of the permutation tests are the average of those obtained from 1,000 permutations.
(PDF)

S1 Fig. BSL analysis on separate classes of features, without incorporation of prior knowledge. Image showing the BSL analysis applied to different categories of features. All the illustrated graphs are generated without taking the information provided by clinicians into account.
(TIF)

S2 Fig. BSL analysis on the most relevant features, without incorporation of prior knowledge. Graph generated with the most relevant features found from the graphs shown in S1 Fig. This graph is generated without taking the information provided by clinicians into account.
(TIF)

S3 Fig. Logistic regression permutation evaluation. Results of the permutation test conducted with the logistic regression algorithm.
(TIF)

S4 Fig. Linear SVM permutation evaluation. Results of the permutation test conducted with the linear kernel SVM.
(TIF)

S5 Fig. Polynomial SVM permutation evaluation. Results of the permutation test conducted with the polynomial kernel SVM.
(TIF)

Author Contributions

Conceptualization: Elisa Ferrari, Davide Bacciu.

Data curation: Luna Gargani, Greta Barbieri, Lorenzo Ghiadoni, Francesco Faita.

Formal analysis: Elisa Ferrari, Francesco Faita, Davide Bacciu.

Investigation: Elisa Ferrari, Francesco Faita, Davide Bacciu.

Methodology: Elisa Ferrari, Davide Bacciu.

Software: Elisa Ferrari.

Supervision: Luna Gargani, Lorenzo Ghiadoni, Francesco Faita, Davide Bacciu.

Validation: Elisa Ferrari.

Visualization: Elisa Ferrari, Davide Bacciu.

Writing – original draft: Elisa Ferrari, Francesco Faita.

Writing – review & editing: Elisa Ferrari, Luna Gargani, Greta Barbieri, Lorenzo Ghiadoni, Francesco Faita, Davide Bacciu.

References

1. World Health Organisation. WHO Coronavirus Disease (COVID-19) Dashboard. Accessed 8th March 2020; 2021. Available from: https://covid19.who.int/.

2. Kirkpatrick JN, Hull SC, Fedson S, Mullen B, Goodlin SJ. Scarce-resource allocation and patient triage during the COVID-19 pandemic: JACC Review Topic of the Week. Journal of the American College of Cardiology. 2020; 76(1):85–92. https://doi.org/10.1016/j.jacc.2020.05.006 PMID: 32407772

3. Abd-Alrazaq A, Alajlani M, Alhuwail D, Schneider J, Al-Kuwari S, Shah Z, et al. Artificial Intelligence in the Fight Against COVID-19: Scoping Review. Journal of medical Internet research. 2020; 22(12): e20756. https://doi.org/10.2196/20756 PMID: 33284779

4. Ayoub A, Mahboob K, Javed AR, Rizwan M, Gadekallu TR, Abidi MH, et al. Classification and categorization of covid-19 outbreak in pakistan. Computers, Materials and Continua. 2021; p. 1253–1269. https://doi.org/10.32604/cmc.2021.015655

5. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. Multimedia Systems. 2021; p. 1–15. https://doi.org/10.1007/s00530-021-00774-w PMID: 33814730

6. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj. 2020; 369. https://doi.org/10.1136/bmj.m1328 PMID: 32265220

7. Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. Nature communications. 2020; 11(1):1–9. https://doi.org/10.1038/s41467-020-17419-7

8. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining; 2015. p. 1721–1730.

9. Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search. 2nd ed. MIT press; 2000.

10. Aliferis CF, Statnikov AR, Tsamardinos I, Brown LE. Causal Explorer: A Causal Probabilistic Network Learning Toolkit for Biomedical Discovery; 2003. p. 371–376.

11. Bacciu D, Etchells TA, Lisboa PJ, Whittaker J. Efficient Identification of Independence Networks Using Mutual Information. Computational Statistics. 2013; 28(2):621–646. https://doi.org/10.1007/s00180-012-0320-6

12. Uusitalo L. Advantages and challenges of Bayesian networks in environmental modelling. Ecological modelling. 2007; 203(3-4):312–318. https://doi.org/10.1016/j.ecolmodel.2006.11.033

13. Heckerman D. A tutorial on learning with Bayesian networks. Innovations in Bayesian networks. 2008; p. 33–82. https://doi.org/10.1007/978-3-540-85066-3_3

14. Kontkanen P, Myllymäki P, Silander T, Tirri H, Grunwald P. Comparing predictive inference methods for discrete domains. In: In Proceedings of the sixth international workshop on artificial intelligence and statistics. Citeseer; 1997.

15. Marcot BG, Holthausen RS, Raphael MG, Rowland MM, Wisdom MJ. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. Forest ecology and management. 2001; 153(1-3):29–42. https://doi.org/10.1016/S0378-1127(01)00452-2

16. Walters CJ, Martell SJ. Fisheries ecology and management. Princeton University Press; 2004.

17. Imoto S, Kim S, Goto T, Aburatani S, Tashiro K, Kuhara S, et al. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. Journal of bioinformatics and computational biology. 2003; 1(02):231–252. https://doi.org/10.1142/S0219720003000071 PMID: 15290771

18. Jiang X, Neapolitan RE, Barmada MM, Visweswaran S. Learning genetic epistasis using Bayesian network scoring criteria. BMC bioinformatics. 2011; 12(1):1–12. https://doi.org/10.1186/1471-2105-12-89 PMID: 21453508

19. Van der Heijden M, Velikova M, Lucas PJ. Learning Bayesian networks for clinical time series analysis. Journal of biomedical informatics. 2014; 48:94–105. https://doi.org/10.1016/j.jbi.2013.12.007 PMID: 24361389

20. Onisko A, Druzdzel MJ, Austin RM. How to interpret the results of medical time series data analysis: classical statistical approaches versus dynamic Bayesian network modeling. Journal of pathology informatics. 2016; 7. https://doi.org/10.4103/2153-3539.197191 PMID: 28163973

21. Flores MJ, Nicholson AE, Brunskill A, Korb KB, Mascaro S. Incorporating expert knowledge when learning Bayesian network structure: a medical case study. Artificial intelligence in medicine. 2011; 53 (3):181–204. https://doi.org/10.1016/j.artmed.2011.08.004

22. Potere N, Valeriani E, Candeloro M, Tana M, Porreca E, Abbate A, et al. Acute complications and mortality in hospitalized patients with coronavirus disease 2019: a systematic review and meta-analysis. Critical care. 2020; 24(1):1–12. https://doi.org/10.1186/s13054-020-03022-1 PMID: 32616077

23. Dana S, Simas AB, Filardi BA, Rodriguez RN, da Costa Valiengo LL, Gallucci-Neto J. Brazilian Modeling of COVID-19 (BRAM-COD): a Bayesian Monte Carlo approach for COVID-19 spread in a limited data set context. MedRxiv. 2020;.

24. Wibbens PD, Koo WWY, McGahan AM. Which COVID policies are most effective? A Bayesian analysis of COVID-19 by jurisdiction. PloS one. 2020; 15(12):e0244177. https://doi.org/10.1371/journal.pone.0244177 PMID: 33373384

25. Mbuvha R, Marwala T. Bayesian inference of COVID-19 spreading rates in South Africa. PloS one. 2020; 15(8):e0237126. https://doi.org/10.1371/journal.pone.0237126 PMID: 32756608

26. Fenton N, McLachlan S, Lucas P, Dube K, Hitman G, Osman M, et al. A privacy-preserving Bayesian network model for personalised COVID19 risk assessment and contact tracing. medRxiv. 2020;.

27. McLachlan S, Lucas P, Dube K, McLachlan GS, Hitman GA, Osman M, et al. The fundamental limitations of COVID-19 contact tracing methods and how to resolve them with a Bayesian network approach. London, UK 2020;2(27042.66243).

28. Neil M, Fenton N, Osman M, McLachlan S. Bayesian Network Analysis of Covid-19 data reveals higher Infection Prevalence Rates and lower Fatality Rates than widely reported. Journal of Risk Research. 2020; 23(7-8):866–879. https://doi.org/10.1080/13669877.2020.1778771

29. Barbieri G, Spinelli S, Filippi M, Foltran F, Giraldi M, Martino MC, et al. COVID-19 pandemic management at the Emergency Department: the changing scenario at the University Hospital of Pisa. Emergency Care Journal. 2020; 16(2). https://doi.org/10.4081/ecj.2020.9146

30. Beretta S, Castelli M, Gonçalves I, Henriques R, Ramazzotti D. Learning the structure of Bayesian Networks: A quantitative assessment of the effect of different algorithmic schemes. Complexity. 2018; 2018. https://doi.org/10.1155/2018/1591878

31. Buntine W. A guide to the literature on learning probabilistic networks from data. IEEE Transactions on knowledge and data engineering. 1996; 8(2):195–210. https://doi.org/10.1109/69.494161

32. Daly R, Shen Q, Aitken S. Learning Bayesian networks: approaches and issues. The knowledge engineering review. 2011; 26(2):99. https://doi.org/10.1017/S0269888910000251

33. Scutari M, Graafland CE, Gutiérrez JM. Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms. International Journal of Approximate Reasoning. 2019; 115:235–253. https://doi.org/10.1016/j.ijar.2019.10.003

34. Spirtes P, Glymour CN, Scheines R, Heckerman D. Causation, prediction, and search. MIT press; 2000.

35. Spirtes P, Glymour C. An algorithm for fast recovery of sparse causal graphs. Social science computer review. 1991; 9(1):62–72. https://doi.org/10.1177/089443939100900106

36. Kalisch M, Bühlmann P. Causal structure learning and inference: a selective review. Quality Technology & Quantitative Management. 2014; 11(1):3–21. https://doi.org/10.1080/16843703.2014.11673322

37. Finney DJ. The Fisher-Yates test of significance in 2 × 2 contingency tables. Biometrika. 1948; 35(1/2):145–156. https://doi.org/10.2307/2332635 PMID: 18867419

38. Kornbrot D. Point biserial correlation. Encyclopedia of statistics in behavioral science. 2005;.

39. Bramer M. Avoiding overfitting of decision trees. Principles of data mining. 2007; p. 119–134.

40. Sanchez-Ramirez DC, Mackey D. Underlying respiratory diseases, specifically COPD, and smoking are associated with severe COVID-19 outcomes: A systematic review and meta-analysis. Respiratory medicine. 2020; p. 106096. https://doi.org/10.1016/j.rmed.2020.106096 PMID: 32763754

41. Gacche R, Gacche R, Chen J, Li H, Li G. Predictors of morbidity and mortality in COVID-19. European review for medical and pharmacological sciences. 2021; 25(3):1684–1707. PMID: 33629339

42. Xu J, Xiao W, Liang X, Zhang P, Shi L, Wang Y, et al. The association of cerebrovascular disease with adverse outcomes in COVID-19 patients: a meta-analysis based on adjusted effect estimates. Journal of Stroke and Cerebrovascular Diseases. 2020; 29(11):105283. https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105283 PMID: 33066922

43. Fathi M, Vakili K, Sayehmiri F, Mohamadkhani A, Hajiesmaeili M, Rezaei-Tavirani M, et al. The prognostic value of comorbidity for the severity of COVID-19: A systematic review and meta-analysis study. PloS one. 2021; 16(2):e0246190. https://doi.org/10.1371/journal.pone.0246190 PMID: 33592019

44. Sharma A, Jaiswal P, Kerakhan Y, Saravanan L, Murtaza Z, Zergham A, et al. Liver disease and outcomes among COVID-19 hospitalized patients-a systematic review and meta-analysis. Annals of hepatology. 2020;. https://doi.org/10.1016/j.aohep.2020.10.001 PMID: 33075578

45. Coppelli A, Giannarelli R, Aragona M, Penno G, Falcone M, Tiseo G, et al. Hyperglycemia at hospital admission is associated with severity of the prognosis in patients hospitalized for COVID-19: the Pisa COVID-19 Study. Diabetes Care. 2020; 43(10):2345–2348. https://doi.org/10.2337/dc20-1380 PMID: 32788285

46. July J, Pranata R. Prevalence of dementia and its impact on mortality in patients with coronavirus disease 2019: A systematic review and meta-analysis. Geriatrics & Gerontology International. 2021; 21(2):172–177. https://doi.org/10.1111/ggi.14107 PMID: 33340212

47. Yadaw AS, Li Yc, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. The Lancet Digital Health. 2020; 2(10): e516–e525. https://doi.org/10.1016/S2589-7500(20)30217-X PMID: 32984797

48. Peckham H, de Gruijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nature communications. 2020; 11(1):1–10. https://doi.org/10.1038/s41467-020-19741-6 PMID: 33298944

49. Zhang H, Ma S, Han T, Qu G, Cheng C, Uy JP, et al. Association of smoking history with severe and critical outcome in COVID-19 patients: A systemic review and meta-analysis. European journal of integrative medicine. 2021; p. 101313. https://doi.org/10.1016/j.eujim.2021.101313 PMID: 33619437

50. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. Journal of Infection. 2020;. https://doi.org/10.1016/j.jinf.2020.04.021

51. Qiu P, Zhou Y, Wang F, Wang H, Zhang M, Pan X, et al. Clinical characteristics, laboratory outcome characteristics, comorbidities, and complications of related COVID-19 deceased: a systematic review and meta-analysis. Aging clinical and experimental research. 2020; p. 1–10. https://doi.org/10.1007/s40520-020-01664-3 PMID: 32734576

52. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney impairment is associated with in-hospital death of COVID-19 patients. MedRxiv. 2020;.

53. Ware LB, Matthay MA. The acute respiratory distress syndrome. New England Journal of Medicine. 2000; 342(18):1334–1349. https://doi.org/10.1056/NEJM200005043421806 PMID: 10793167

54. Collins JA, Rudenski A, Gibson J, Howard L, O'Driscoll R. Relating oxygen partial pressure, saturation and content: the haemoglobin–oxygen dissociation curve. Breathe. 2015; 11(3):194–201. https://doi.org/10.1183/20734735.001415 PMID: 26632351

55. Morlacco A, Motterle G, Zattoni F. The multifaceted long-term effects of the COVID-19 pandemic on urology. Nature Reviews Urology. 2020; 17(7):365–367. https://doi.org/10.1038/s41585-020-0331-y PMID: 32377015

56. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 Long-term effects of COVID-19: a systematic review and meta-analysis. Available at SSRN 3769978. 2021;.

57. Luo Y, Peng J, Ma J. When causal inference meets deep learning. Nature Machine Intelligence. 2020; 2 (8):426–427. https://doi.org/10.1038/s42256-020-0218-x