Preprints.org

Article

Not peer-reviewed version

# Bayesian networks in the management of hospital admissions: a comparison between explainable AI and black box AI during the pandemics.

Giovanna Nicora , Michele Catalano , Chandra Bortolotto * , Marina Francesca Achilli , Gaia Messana , Antonio Lo Tito , Alessio Consonni , Sara Cutti , Federico Comotto , Giulia Maria Stella , Angelo Guido Corsico , Stefano Perlini , Riccardo Bellazzi , Raffaele Bruno , Lorenzo Preda

Posted Date: 11 March 2024

doi: 10.20944/preprints202403.0647.v1

Keywords: Artificial Intelligence; Explainability; Machine Learning; Random Forest; Bayesian Networks; COVID-19

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

*Article*

# Bayesian Networks in the Management of Hospital Admissions: A Comparison between Explainable AI and Black Box AI during the Pandemics

**Nicora Giovanna [1], Michele Catalano [2], Chandra Bortolotto [2,*], Achilli Marina Francesca [2], Messana Gaia [2], Lo Tito Antonio [2], Consonni Alessio [2], Cutti Sara [3], Comotto Federico [4], Stella Giulia Maria [5], Corsico Angelo [5], Perlini Stefano [6], Bellazzi Riccardo [2], Bruno Raffaele [7] and Preda Lorenzo [1]**

[1] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy.
[2] Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy and Radiology Department, Fondazione IRCCS Policlinico San Matteo – Pavia, Italy;
[3] Medical Direction, Fondazione IRCCS Policlinico San Matteo – Pavia, Italy.
[4] Reply S.p.A. Corso Francia, 110, Turin, Italy.
[5] Department of Internal Medicine and Therapeutics, University of Pavia, Pavia, Italy and Dept. of Respiratory Diseases Unit, Fondazione IRCCS Policlinico San Matteo – Pavia, Italy.
[6] Department of Internal Medicine and Therapeutics, University of Pavia, Pavia, Italy and Dept. of Emergency Department, Fondazione IRCCS Policlinico San Matteo – Pavia, Italy.
[7] Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy and Infectious Diseases Unit, Fondazione IRCCS Policlinico San Matteo – Pavia, Italy.
* Correspondence: chandra.bortolotto@unipv.it

**Abstract:** Artificial Intelligence (AI) and Machine Learning (ML) approaches that could learn from large data sources have been identified as useful tools to support clinicians in their decisional process; AI and ML implementations have had a rapid acceleration during the recent COVID-19 pandemic. However, many ML classifiers are "black box" to the final user, since their underlying reasoning process is often obscure. Additionally, the performance of such models suffer from poor generalization ability in presence of dataset shift. Here, we present a comparison between an explainable-by-design ("white box") model (Bayesian Network (BN)) versus a black-box model (Random Forest), both studied with the aim to support clinicians of Policlinico San Matteo University Hospital in Pavia (Italy) during the triage of COVID-19 patients. Our aim is to evaluate whether the BN predictive performances are comparable with those of a widely used but less explainable ML model such as Random Forest and to test the generalization ability of the ML models across different waves of the pandemic.

**Keywords:** artificial intelligence; explainability; machine learning; random forest; Bayesian networks; COVID-19

## 1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) approaches are generally recognized as useful tools to support clinicians in their decision process. AI/ML tools have been developed to solve various medical problems, for instance to predict disease diagnosis [1], to support cancer detection in medical images [2] and even to determine the best treatment based on patient's clinical and genomics characteristics [3,4]. As research studies examining AI/ML applications in medicine have been growing during the years, interests have raised in the actual implementation of these systems into clinical practice [5]. The recent Sars-Cov-2 pandemics have promoted an acceleration of digital health technologies and AI/ML implementations to fight the pandemics, from supporting hospital admission to define new therapeutic strategies [6,7]. Nevertheless, building Machine Learning (ML)

© 2024 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



approaches that can be generalized over time and/or on patients coming from different hospitals is challenging. Since ML relies on data for training, the less the training data are representative of the true underlying population the less reliable the ML model will be when applied to new data that may deviate from the training population. As a matter of fact, ML inherently suffers from dataset shifts and poor generalization ability across different populations [8–10], which leads to a decrease of trust in AI/ML predictions. For instance, a recent paper showed the impact of data drift on the performance of AI models designed to predict sepsis, advocating for frequent re-training of such models [11].

The pandemic has been a perfect testing ground since several sources of data shifts had occurred. Several Sars-Cov-2 variants arises, from the Alpha variant recognized as Variant of Concern (VOC) in late December 2020, to the Omicron variant dominating the virus landscape from autumn 2021. Each VOCs exhibit different characteristics, such as increased transmissibility, reduction of treatment and vaccine efficacy, severity of symptoms. As our clinical and biological knowledge increased, new treatment protocols were defined over the pandemics. As a consequence, AI/ML models trained on data collected on particular time interval may not be calibrated to be used on data collected subsequently, and a re-training strategy should be defined. For instance, in [12] we developed a ML model to predict new Sars-Cov-2 variants of concern over the pandemics based on the Spike protein sequence. We simulated the implementation of this algorithm from the beginning of the pandemics to March 2023 and we periodically re-train the model when the World Health Organization (WHO) recognized a new variant as a variant of concern.

Another aspect that undermines trust in AI/ML is the perception the AI/ML classification process is obscure. Widely used algorithms, from Neural Networks to Gradient Boosting, allow fast learning with high performance from large data sources, but the reasoning process between input and output variables is usually a black-box to the user. These complex models are often perceived as more performing in comparison with “white box”, more transparent models, such as Logistic Regression or Bayesian network. Current research on AI Explainability (XAI) aims at making black box ML predictions more transparent. Towards this direction, different XAI approaches have been developed. Many of these providing explanations of single ML predictions by highlighting the important features that lead the classifier to its final decision[13]. Other XAI approaches approximate the complex model with a “white box” model on a local neighborhood of the prediction that needs to be explained [14]. However, as recently stated, in order to reach explainability in medicine we need to promote causability [15]. Additionally, popular XAI methods, such as LIME and SHAP showed to be unreliable in case of adversarial attacks [16]. Explanations derived from current XAI methods provide spurious correlations rather than cause/effects relationships, leading to erroneous or even biased explanations [17]. A recent study compared an explainable-by-design model (or “white box”, namely Bayesian Network (BN), with the explanations derived from XAI methods, by performing a survey with human participants. They found that BNs are easier to interpret compared to their associated XAI methods [18].

In the context of predicting hospital admission for COVID-19 through artificial intelligence and machine learning (AI/ML), we have a dual objective. The first objective is to explore the impact of data drift on the performance of a model developed during the first wave of the COVID-19 pandemic. Data drift refers to any change in data over time that could affect the performance of the model. This is particularly relevant in the context of COVID-19, where the characteristics of the disease and treatment strategies and health resources have changed over time. The second objective is to quantify whether a transparent model, such as a Bayesian Network (BN), has similar performance compared to a more complex, black box type model. A black box model is a type of AI/ML model that produces accurate predictions, but the process through which these predictions occur is not easily understandable or interpretable.

In particular, we developed AI/ML models that suggest to doctors whether patients with COVID-19 in triage could be treated at home or need to be hospitalized. This is a critical task, as effective management of patients with COVID-19 can have a significant impact on patient health and the use of health resources.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



We have trained a Bayesian Network (BN) for this purpose. A BN is a probabilistic graphical model that allows modeling the conditional dependencies of a set of variables. It presents itself in the form of an acyclic graph where each node represents a data variable, and the edges represent the probability dependencies between nodes. Systems based on BN can model complex relationships between variables when conditions of causality and conditional independence are involved. This is of great importance in the clinical decision-making process, where understanding the relationships between variables can guide more informed and accurate decisions. In addition, the joint probability distributions can be updated if new evidence is available using Bayes' theorem. This means that the model can adapt and improve over time as new data is collected.

We were therefore able to model existing medical evidence and suggest potential cause/effect relationships between clinical variables and hospital admission. This can help doctors better understand the factors that influence the need for hospital admission for patients with COVID-19, and therefore make more informed and effective decisions.

We then evaluated whether the BN predictive performances were comparable with those of a widely used but less explainable ML model, e.g. Random Forest. We also tested the generalization ability of the ML models across different waves of the pandemic.

## 2. Materials and Methods

During the "first wave" of the COVID-19 pandemics in Italy (from March 2020 to May 2020), we gathered data from 660 COVID-19 patients treated at the Fondazione IRCCS Policlinico San Matteo hospital in Pavia, an excellence center that is known to have successfully treated the first diagnosed COVID-19 patients in western countries. Half of these patients were hospitalized, while the remaining showed a better prognosis, and were treated at home. For each patient, we collected information about age, gender, clinical features, such as C-reactive protein (CRP), symptoms, such as cough and breathing difficulties, and presence of comorbidities, such as hypertension, cardiovascular diseases, cancer.

**Table 1.** The table provides a schematic overview of first and third waves of COVID-19 cases, detailing demographic data, comorbidities, and symptoms presented by patients.

| | | I wave | III wave |
|---|---|---|---|
| **Demographic data** | Men | 441 | 266 |
| | Women | 219 | 196 |
| | Age (average) | 66,2 | 68,1 |
| | Age (median) | 68,0 | 70,5 |
| **Comorbidities** | Number of comorbidities (average) | 1,5 | 0,9 |
| | More than two comorbidities | 148 | 90 |
| **Signs and symptoms** | Fever (T>37,8 °C) | 153 | 447 |
| | Cough | 291 | 129 |
| | Dyspnea | 349 | 248 |
| | $SatO_2$ | 91 | 94 |

A Deep Learning algorithm was used to extract features from chest radiographs (CXR) images through the X-RAIS platform, developed by Reply™. X-RAIS is a Deep Network able to analyze different types of medical images and to extract relevant information for diagnosis. In our context, X-RAIS transformed the CXR image of each patient into 5 numerical clinically relevant features: Consolidation, Infiltration, Edema, Effusion and Lung Opacity. These 5 features, together with 19 other clinical features, represented the input of a ML model that would predict whether a patient should be hospitalized (class 1) or not (class 0). [Figure 1] We randomly selected 90% of the patients as a training set. The remaining 10% of patients were kept for testing and selecting the best performing model. During the "third wave" (from March to May 2021), 462 additional patients were triaged in our Emergency Department, and 68% of them were hospitalized. This dataset is based on ALFABETO (ALl FAster BEtter TOghe-ter) project, whose aim is to develop an AI-based pipeline

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



integrating data from diagnostic tools and clinical features to support clinicians during the triage of COVID-19 patients [19]. The third wave set was exploited as a validation set.

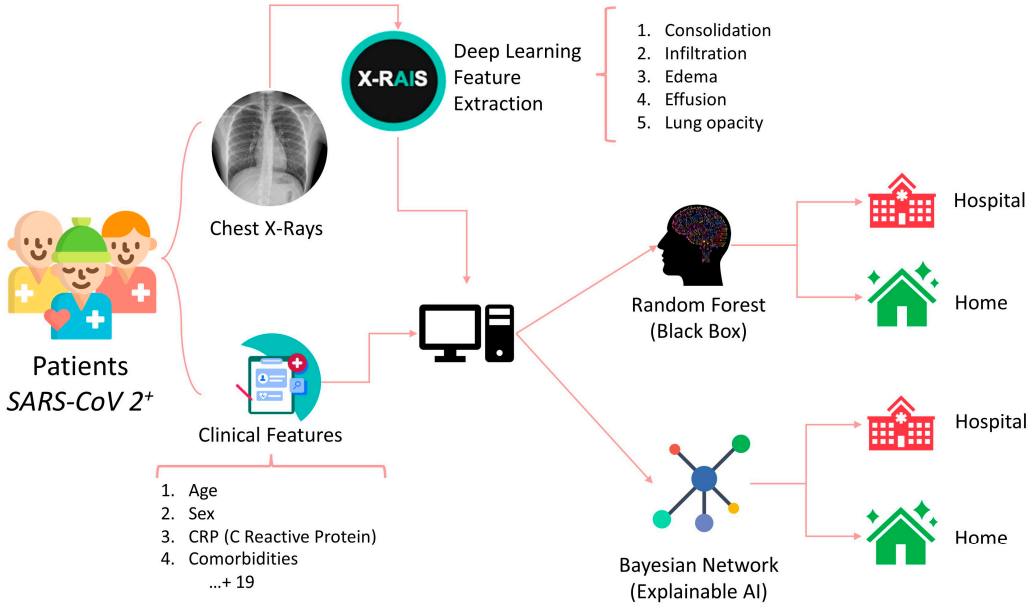


**Figure 1.** Graphical representation of patient's medical data gathered and deep learning informations used to estimate hospitalization or at-home management. Data is processed by Random Forest and Bayesian Network algorithms, determining the appropriate treatment option.

To implement the BN, we first designed a graph based on our pre-existing knowledge and experts advice. This graph contains relationships between a few variables that may represent the clinicians reasoning process during triage. The graph is represented in Figure 2a: the node labeled "Treatment (Home vs Hospital)" is the target node representing our outcome of interest, i.e. whether the patient should be hospitalized or not. To make this decision, we assumed that the clinician would evaluate at least the age, the gender (male patients are more likely to incur more severe consequences from the infection) and whether the patient had breathing difficulties. The target node depends on these 3 variables, and we also assumed a direct dependency between age and breathing difficulties. We then automatically enriched the structure of this graph with the remaining collected variables, by using the hill climbing search algorithm applied on the training data: starting from the constraints represented in Figure 2a, this method implements a greedy local search and performs single-edge manipulations that maximally increase a score of fitness. The search terminates once a local maximum is found [20]. The resulting graph is shown in Figure 2b: notably, the Boolean feature indicating whether the patient has more than 2 comorbidities ("ComorbiditiesGreaterThan2") is explicitly linked to comorbidities nodes, such as the presence of cancer or cardiovascular diseases. Interestingly, the outcome is not directly linked to the node "ComorbiditiesGreaterThan2", but it can be linked to the presence of comorbidities through the patient's age. Some DL features directly depend on the target node. BN is implemented in Python 3.7, using the bnlearn package.

On the test set (from the first wave, therefore coming from the same population of the training set) and on the validation set (from the third wave, when new variants, treatment protocols and hospital facilities appeared), we evaluated several performance metrics. As True Positive (TP), we defined the number of hospitalized patients correctly classified by the model, while as True Negative (TN) we considered the number of patients treated at home correctly classified. The False Negative (FN) represented the number of hospitalized patients incorrectly classified as "treat at home", while as False Positive we indicated the number of patients treated at home that were incorrectly classified

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



as "hospital". In particular, we computed the Area Under the Curve (AUC) and Precision Recall Curve (PRC), the Accuracy (as the proportion of correctly classified patients over the total number of patients), the Precision (as the ratio between the TP and the sum of TP and FO), the Recall (as the ration between the TP and the total number of "hospital" patients), the sensitivity (as the ration between the TN and the total number of "home" patients), and the F1 score (harmonic mean between specificity and recall).

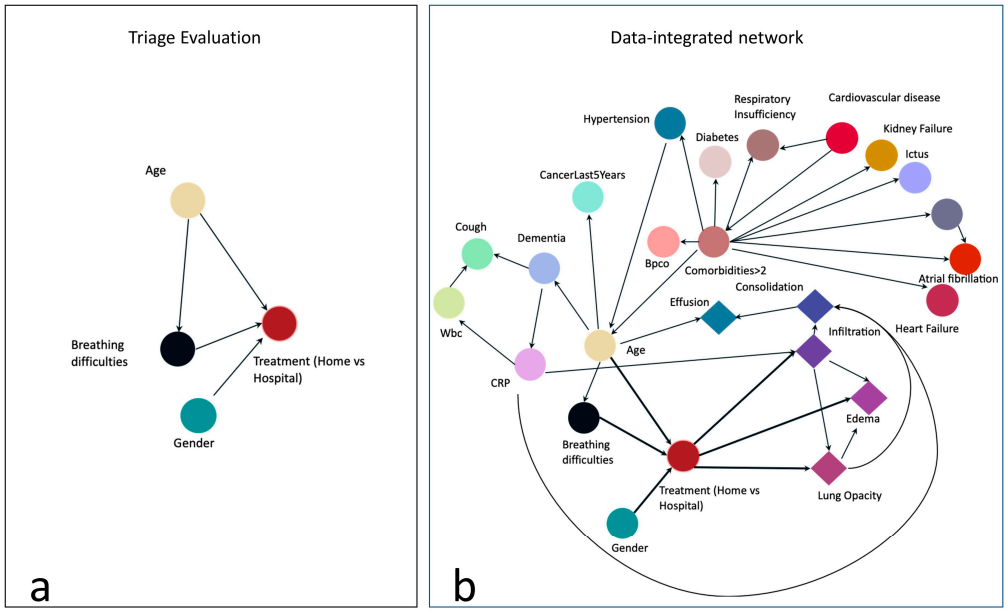


**Figure 2.** The left graph (**a**) includes relationships between some variables that represent the reasoning process of clinicians during emergency triage. The right graph (**b**) is the result of enriching it through the hill climbing search algorithm with the remaining collected variables.

## 3. Results

Herein we report the predicted performance of the BN in Figure 2b, whose structure is based both on evidence and from data. The simplest network, only based on evidence (Figure 2a), shows good recall (i.e. the ability to correctly classify hospitalization) on the Test Set (85%) but low specificity (around 20%), and for this reason it was excluded from the analysis. We trained and tested three additional models: a regularized Logistic Regression, Gradient Boosting and Random Forest (RF). We show the performance of the RF only, since it outperforms the other two models on test data. RF is a widely applied ensemble classifier that works by training several decision trees through bagging. Table 2 reports BN and RF classification performance on 66 patients of the Test set, in terms of various metrics, such as Area Under the ROC Curve (AUC) and Area Under the Precision-Recall Curve (PRC), described above. On the relatively small test set, the BN showed higher performance in comparison with the RF for all the metrics of around 5%.

**Table 2.** Predictive performance of the Bayesian Network (BN) and the Random Forest (RF) on the Test set (66 Patients).

| | AUC | PRC | Accuracy | Precision | Recall | Specificity | F1 Score |
|---|---|---|---|---|---|---|---|
| **BN** | 0.80 | 0.85 | 0.76 | 0.82 | 0.78 | 0.73 | 0.79 |
| **RF** | 0.76 | 0.84 | 0.71 | 0.78 | 0.72 | 0.69 | 0.75 |

To test whether the error rates of the two approaches are significantly different, we apply the McNemar's Test [21], a statistical test for paired nominal data. In the context of AI/ML the McNemar's Test can be used to compare the predictive performance of two models on the same test set. P-value

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



is 0.6, and we can-not reject the null hypothesis, i.e. the two classifiers have the same error rates. In Table 3 we can observe the prediction ability on third wave patients, where RF has slightly higher performance. Also in this case the p-value of McNemar's test is high (0.7). In comparison with Test results, both BN and RF show lower recall, but higher specificity, meaning that they are more accurate in predicting the "home" class. In this case, the RF performance are higher in terms of F1 score, Accuracy and Recall.

**Table 3.** Predictive performance of the Bayesian Network (BN) and the Random Forest (RF) during the third wave (462 Patients).

| | AUC | PRC | Accuracy | Precision | Recall | Specificity | F1 Score |
|---|---|---|---|---|---|---|---|
| **BN** | 0.80 | 0.89 | 0.71 | 0.89 | 0.65 | 0.82 | 0.75 |
| **RF** | 0.83 | 0.91 | 0.75 | 0.89 | 0.71 | 0.82 | 0.79 |

We examined the RF features importance by computing the mean decrease impurity [22]. The most important feature for classification resulted in the C reactive protein (CRP), which was also directly linked to the outcome by the BN structure learning algorithm. In RF, CRP is followed by four DL-extracted features (LungOpacity, Edema, Consolidation and Infiltration). Interestingly, the CRP has been identified as a relevant characteristics significantly higher in sever COVID-19 patients on data from the 4CE consortium[23]. All the DL-extracted features, except for Consolidation, have a direct link to the outcome (1b). Age, gender and breathing difficulties are placed in the 8th, 9th and 10th positions.

We additionally analyzed the BN and RF performance and their misclassification errors in light of prognostic outcomes that were collected after the initial diagnosis for each patient. In summary, each patient included in the study (both treated at home or in hospital), underwent follow up, and clinicians recorded the outcomes after some time from the initial diagnosis. Patients with mild outcome were those not hospitalized or hospitalized without the need for ventilation support. Patients with moderate outcome were hospitalized with airway pressure device support, while severe patients were those hospitalized with invasive ventilatory support or deceased. We analyzed misclassification errors by the BN, and we correlate it with the final outcome. We found that for False Positive patients (i.e. patients predicted "hospital" where in fact they were treated at home at the beginning), higher percentage of severe and moderate patients were detected. In False Negative patients, most of them had a mild outcome.

## 4. Discussion

In this work we showed the development of a Bayesian Network (BN), whose goal is to predict the need for a COVID-19 patient to be hospitalized rather than followed up at home based on their clinical characteristics. In a recently published paper, Shen et al have proposed a BN-based model as support tool to assess the severity of COVID-19 symptoms in patients, taking into account the epidemiology, clinical symptoms, imaging findings and laboratory tests [24]. Performances are quite similar to RF, but BN shows slightly higher values for all the metrics. Yet, this study tested the RF and the BN to a significantly smaller number of samples.

The majority of our results have been verified by many studies which analyzed the risk factors for COVID-19-associated hospitalization [25–27]; in particular, the presence of at least two comorbidities was indirectly associated with hospitalization, but strictly related to age, while age > 70, male gender, high CRP blood levels and breathing difficulties were directly correlated with increased risk for hospitalization, as well as three distinct radiological features, i.e. Lung Opacity, Edema and Infiltration.

In our research, both BN and RF very able to generalize during the third wave of COVID-19 pandemic, despite some population variables (such as age) changed in comparison to the first wave patients used for training. Additionally, the predictive capacity of the BN was substantially comparable to a Random Forest approach, particularly in the third wave. The values of Precision,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



Accuracy, Recall and Specificity of the BN were even higher with respect to those of the Random Forest during the first wave, showing that the learned structure did not suffer from dataset shift.

While the need for clinical decision support systems that implement ML is growing, understanding how AI algorithms correlate and classify entered data is increasingly important. In fact, these approaches are able to detect useful and hidden patterns in the collected data, which can then be exploited to support the discovery of knowledge and/or to subsequently implement highly accurate automatic classifiers. For machine learning to be safely integrated into daily clinical practice, it is necessary to fully understand the classification process implemented by the algorithm and this is currently not fully feasible, as many high-performance classifiers are perceived as "black boxes" for which it is difficult to understand how the data is processed and how the output is handled.

As the need for clinical decision support systems implementing ML continues to grow, understanding how AI algorithms correlate and classify input data is becoming increasingly important. These approaches are capable of detecting useful and hidden patterns in collected data, which can then be leveraged to support knowledge discovery and/or subsequently implement highly accurate automatic classifiers. This ability to identify subtle connections and predictions based on complex data is crucial in the context of medicine, where a swift and precise diagnosis can make a difference in a patient's life.

To securely integrate machine learning into everyday clinical practice, a full understanding of the classification process implemented by the algorithm is necessary. However, this is a task that currently presents significant challenges. Indeed, many of the high-performance classifiers used are perceived as "black boxes," as it is difficult for physicians and healthcare providers not only to understand exactly how data is processed but also how the output is managed.

This becomes particularly critical when it comes to making decisions that directly impact patient health and well-being. The lack of transparency in AI classification processes can lead to a lack of trust in the use of such systems. Therefore, current research is increasingly focusing on interpreting and explaining AI models, seeking to make these algorithms more transparent and understandable for those using them in clinical practice. Only through a deeper understanding of how AI analyzes and interprets data will it be possible to maximize its potential in the field of medicine, ensuring both the safety and effectiveness of clinical decisions.

The BN allows not only to develop an explainable model by design, but also to combine known evidence of dependence on variables with information encoded in the anamnestic and clinical-radiological data of patients, with the ultimate objective to assist physicians in the decision-making process. The resulting network structure can then be inspected by doctors to understand the algorithm classification process and to improve or correct correlations between data.

Considering the medico-legal aspects related to the use of artificial intelligence, as supported by Gleeson et al, "for medico-legal reasons, all contributions from non-human entities must be identifiable. In decision support systems that integrate information from imaging, medical reports, lab results, etc. for the probability of diagnoses, the recommendations change dynamically, as new information is added" [28].

Addressing the challenges of defining explainability also requires consideration of the legal context of healthcare. The European Union's General Data Protection Regulation (GDPR) has mandated the provision of meaningful information to data subjects regarding the logic of AI algorithms, their significance, and their consequences. Although the interpretation of the GDPR is still widely debated by legal experts, the regulation primarily aims to safeguard an individual's right to comprehend the decision-making process and assess the reasonableness of AI decisions. Thus, the explainability requirement does not equate to providing a causal account but rather involves elucidating the choices about the decision-making model, how data was collected, the features that were and were not considered, and the anticipated effects of the automated decision-making process. These statements suggest that explainability should take into account the implications of using AI in specific clinical contexts [29].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



## 5. Limitations

The main limitation of our project was the relatively small sample size, which could be increased through further retrospective studies.

The study could also be hindered by the fact that only some clinical features were included in the BN; we sought to analyze mostly variables who had already been validated by previous studies as important in the clinical evolution of the disease, but the usage of additional clinical criteria could potentially improve the width of the network or increase its correlation strength, thus giving more information to the clinician.

Data from the second wave of the pandemics were not collected, and therefore we were not able to perform our analysis on that period. Another aspect that should be investigated is the possibility to use the model on patients from other hospital, thus investigating the generalizability not only during time but also in space, across different hospitals, where different clinical decision protocols and patients' populations may occur.

## 6. Conclusions

The BN allows the development of a graphically representable model, which combines known information on the dependence of clinical variables, with information encoded in each patient's data identified by the ML model. It also has predictive capabilities similar to a data-only approach such as RF, but greater interpretability, as it is possible to inspect the structure of the BN, to understand the classification process and relationships between variables.

Future works will delve into several avenues of research. This includes the exploration of new network configurations, incorporating additional layers of medical knowledge to enhance the system's understanding and predictive capabilities. Additionally, there will be a focused effort on investigating potential causal relationships between variables, aiming to uncover deeper insights into the complex interactions within medical data.

**Author Contributions:** Conceptualization: C.B., R.B., L.P.; Methodology C.B., G.N., R.B., L.P.; Administrative support: S.C.; Data curation: M.C., M.F.A., A.C., G.M., A.L.T.; Writing—original draft preparation: M.C., C.B., M.F.A., G.M., A.L.T.; Writing—review and editing: G.N., M.C., C.B., G.M.S., A.C., S.P., R.B, L.P.; Software: F.C., G.N.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, 'Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda', *J Ambient Intell Human Comput*, vol. 14, no. 7, pp. 8459–8486, Jul. 2023, doi: 10.1007/s12652-021-03612-z.
2. C. Kaur and U. Garg, 'Artificial intelligence techniques for cancer detection in medical image processing: A review', *Materials Today: Proceedings*, vol. 81, pp. 806–809, Jan. 2023, doi: 10.1016/j.matpr.2021.04.241.
3. A. Mukhopadhyay *et al.*, 'Personalised Dosing Using the CURATE.AI Algorithm: Protocol for a Feasibility Study in Patients with Hypertension and Type II Diabetes Mellitus', *Int J Environ Res Public Health*, vol. 19, no. 15, p. 8979, Jul. 2022, doi: 10.3390/ijerph19158979.
4. C. Gallo, 'Artificial Intelligence for Personalized Genetics and New Drug Development: Benefits and Cautions', *Bioengineering*, vol. 10, no. 5, Art. no. 5, May 2023, doi: 10.3390/bioengineering10050613.
5. V. Kaul, S. Enslin, and S. A. Gross, 'History of artificial intelligence in medicine', *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, Oct. 2020, doi: 10.1016/j.gie.2020.06.040.
6. F. Piccialli, V. S. di Cola, F. Giampaolo, and S. Cuomo, 'The Role of Artificial Intelligence in Fighting the COVID-19 Pandemic', *Inf Syst Front*, vol. 23, no. 6, pp. 1467–1497, 2021, doi: 10.1007/s10796-021-10131-x.
7. K. H. Almotairi *et al.*, 'Impact of Artificial Intelligence on COVID-19 Pandemic: A Survey of Image Processing, Tracking of Disease, Prediction of Outcomes, and Computational Medicine', *Big Data and Cognitive Computing*, vol. 7, no. 1, Art. no. 1, Mar. 2023, doi: 10.3390/bdcc7010011.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1




8. C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, 'Key challenges for delivering clinical impact with artificial intelligence', *BMC Med*, vol. 17, no. 1, Art. no. 1, Dec. 2019, doi: 10.1186/s12916-019-1426-2.
9. J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera, 'A unifying view on dataset shift in classification', *Pattern Recognition*, vol. 45, no. 1, pp. 521–530, Jan. 2012, doi: 10.1016/j.patcog.2011.06.019.
10. R. D. Riley, A. Pate, P. Dhiman, L. Archer, G. P. Martin, and G. S. Collins, 'Clinical prediction models and the multiverse of madness', *BMC Med*, vol. 21, p. 502, Dec. 2023, doi: 10.1186/s12916-023-03212-y.
11. K. Rahmani *et al.*, 'Assessing the effects of data drift on the performance of machine learning models used in clinical sepsis prediction', *International Journal of Medical Informatics*, vol. 173, p. 104930, May 2023, doi: 10.1016/j.ijmedinf.2022.104930.
12. G. Nicora, M. Salemi, S. Marini, and R. Bellazzi, 'Predicting emerging SARS-CoV-2 variants of concern through a One Class dynamic anomaly detection algorithm', *BMJ Health Care Inform*, vol. 29, no. 1, p. e100643, Dec. 2022, doi: 10.1136/bmjhci-2022-100643.
13. S. Lundberg and S.-I. Lee, 'A Unified Approach to Interpreting Model Predictions', *arXiv:1705.07874 [cs, stat]*, Nov. 2017, Accessed: Nov. 03, 2021. [Online]. Available: http://arxiv.org/abs/1705.07874
14. E. Parimbelli, T. M. Buonocore, G. Nicora, W. Michalowski, S. Wilk, and R. Bellazzi, 'Why did AI get this one wrong? — Tree-based explanations of machine learning model predictions', *Artificial Intelligence in Medicine*, vol. 135, p. 102471, Jan. 2023, doi: 10.1016/j.artmed.2022.102471.
15. A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, 'Causability and explainability of artificial intelligence in medicine', *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, p. e1312, 2019, doi: 10.1002/widm.1312.
16. D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju, 'Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods', in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York NY USA: ACM, Feb. 2020, pp. 180–186. doi: 10.1145/3375627.3375830.
17. Y.-L. Chou, C. Moreira, P. Bruza, C. Ouyang, and J. Jorge, 'Counterfactuals and Causability in Explainable Artificial Intelligence: Theory, Algorithms, and Applications', *arXiv:2103.04244 [cs]*, Jun. 2021, Accessed: Oct. 07, 2021. [Online]. Available: http://arxiv.org/abs/2103.04244
18. R. Butz, R. Schulz, A. Hommersom, and Marko van Eekelen, 'Investigating the understandability of XAI methods for enhanced user experience: When Bayesian network users became detectives', *Artificial Intelligence in Medicine*, vol. 134, p. 102438, Dec. 2022, doi: 10.1016/j.artmed.2022.102438.
19. M. Catalano *et al.*, 'Performance of an AI algorithm during the different phases of the COVID pandemics: what can we learn from the AI and vice versa.', *European Journal of Radiology Open*, vol. 11, p. 100497, Dec. 2023, doi: 10.1016/j.ejro.2023.100497.
20. M. Scutari, C. E. Graafland, and J. M. Gutiérrez, 'Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms', *International Journal of Approximate Reasoning*, vol. 115, pp. 235–253, Dec. 2019, doi: 10.1016/j.ijar.2019.10.003.
21. T. G. Dietterich, 'Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms', *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998, doi: 10.1162/089976698300017197.
22. G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, 'Understanding variable importances in forests of randomized trees', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2013. Accessed: Oct. 13, 2021. [Online]. Available: https://papers.nips.cc/paper/2013/hash/e3796ae838835da0b6f6ea37bcf8bcb7-Abstract.html
23. C. Hong *et al.*, 'Changes in laboratory value improvement and mortality rates over the course of the pandemic: an international retrospective cohort study of hospitalised patients infected with SARS-CoV-2', *BMJ Open*, vol. 12, no. 6, p. e057725, Jun. 2022, doi: 10.1136/bmjopen-2021-057725.
24. J. Shen, F. Liu, M. Xu, L. Fu, Z. Dong, and J. Wu, 'Decision support analysis for risk identification and control of patients affected by COVID-19 based on Bayesian Networks', *Expert Systems with Applications*, vol. 196, p. 116547, Jun. 2022, doi: 10.1016/j.eswa.2022.116547.
25. J. Y. Ko *et al.*, 'Risk Factors for Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization: COVID-19-Associated Hospitalization Surveillance Network and Behavioral Risk Factor Surveillance System', *Clin Infect Dis*, vol. 72, no. 11, pp. e695–e703, Jun. 2021, doi: 10.1093/cid/ciaa1419.
26. B. G. Pijls *et al.*, 'Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies', *BMJ Open*, vol. 11, no. 1, p. e044640, Jan. 2021, doi: 10.1136/bmjopen-2020-044640.
27. P. Malik *et al.*, 'Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis', *BMJ Evid Based Med*, vol. 26, no. 3, pp. 107–108, Jun. 2021, doi: 10.1136/bmjebm-2020-111536.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 March 2024 doi:10.20944/preprints202403.0647.v1



28. Gleeson F, Revel MP, Biederer J, Larici AR, Martini K, Frauenfelder T, Screaton N, Prosch H, Snoeckx A, Sverzellati N, Ghaye B, Parkar AP. 'Implementation of artificial intelligence in thoracic imaging-a what, how, and why guide from the European Society of Thoracic Imaging (ESTI).' *Eur Radiol. 2023 Jul*;33(7):5077-5086. doi: 10.1007/s00330-023-09409-2. Epub 2023 Feb 2. PMID: 36729173; PMCID: PMC9892666.
29. Arbelaez Ossa L, Starke G, Lorenzini G, Vogt JE, Shaw DM, Elger BS. 'Re-focusing explainability in medicine.' *Digit Health*. 2022 Feb 11;8:20552076221074488. doi: 10.1177/20552076221074488. PMID: 35173981; PMCID: PMC8841907.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.