European Journal of Epidemiology
https://doi.org/10.1007/s10654-019-00564-9

CORRESPONDENCE

On the relationship of machine learning with causal inference

Sheng-Hsuan Lin[1] · **Mohammad Arfan Ikram**[2,3]

Received: 2 May 2019 / Accepted: 16 September 2019

There is rapid emergence of increasingly sophisticated informatics being applied in biomedical research. Machine learning is a collection of algorithms (including deep neural network learning, support vector machine, random forest, and super learning which integrates all above algorithms) often used in informatics and has been successfully applied in clinical decision making, diagnosis, and image identification [1, 2]. Given its success in these areas, there is great anticipation that a similarly impressive impact is to be expected in explanatory research, the area of research focused on identifying and understanding causes of disease. However, application of machine learning in explanatory research is not straightforward and might lead to misinterpretations [3]. Efforts are ongoing to develop an appropriate theoretical context for integrating machine learning with explanatory research [4]. In this Letter, we outline the fundamental differences between predictive research and explanatory research and summarize challenges and possibilities of applying machine learning in explanatory research.

Biomedical research is by and large focused on either of the following two aims: prediction and explanation. Epidemiology provides the theoretical framework as well as the corresponding statistics to properly carry out either aim. Predictive research is primarily focused on recognizing people with disease or those at increased risk of disease. Modern incarnations of this basic principle that additionally include therapeutic consequences based on such recognition are termed 'precision medicine'. On the other hand, explanatory research, sometimes called analytic research, is primarily focused on identifying causes of disease; to elaborate on the causality by demonstrating and quantifying possible mechanisms; and to investigate how to attenuate disease burden through interventions, i.e. trials. Due to the differences in aims, statistical models also differ accordingly. In predictive research, *fitting* a model to obtain higher accuracy for predicting the outcome is key, whilst in explanatory research the focus is on *estimating* the size of a certain parameter to be interpreted as the impact of the cause on the disease.

Machine learning can be regarded as an algorithm that comprehensively screens thousands of predictive models to select one with the best accuracy. Accuracy is here defined as how well the model recognize persons with disease (i.e. diagnostic research, image identification, screening) or those with highest risk of developing disease or responding to treatment (i.e. prognostic research). Typically, such algorithms require large sample sizes and intensive computational facilities, and only report the final predictive accuracy without necessarily revealing the model used. Nevertheless, machine learning has repeatedly far outperformed conventional predictive models. It is this success that had led to the expectation that machine learning will similarly impact the field of explanatory research. Yet, further scrutiny learns that the impact of machine learning in explanatory research might not be that straightforward.

Appropriate inference of causality is the cornerstone of explanatory research. In recent years, a causal inference framework has been formalized based on probability and mathematical models (Fig. 1). Initially, in the exploratory stage a causal relationship structure is built, usually by direct acyclic graphs or structural equation models. Next, the confirmatory stage consists of three further steps. First, a causal question of interest is formally defined as a causal parameter under the counterfactual outcome model. Second, since causal parameters under the counterfactual outcome model cannot be directly measured by real data, they should be thus identified as statistical parameters. Several causal assumptions, such as no confounding or selection bias, are required for this combining of theory with empirical data. Third, statistical inference estimates or tests the causal question of interest, in terms of statistical parameters.

✉ Mohammad Arfan Ikram
m.a.ikram@erasmusmc.nl

[1] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

[2] Department of Epidemiology, Erasmus Medical Center, Doctor Molewaterplein 40, 3015 GD Rotterdam, The Netherlands

[3] Department of Epidemiology, Harvard Chan School of Public Health, Boston, USA

Published online: 27 September 2019

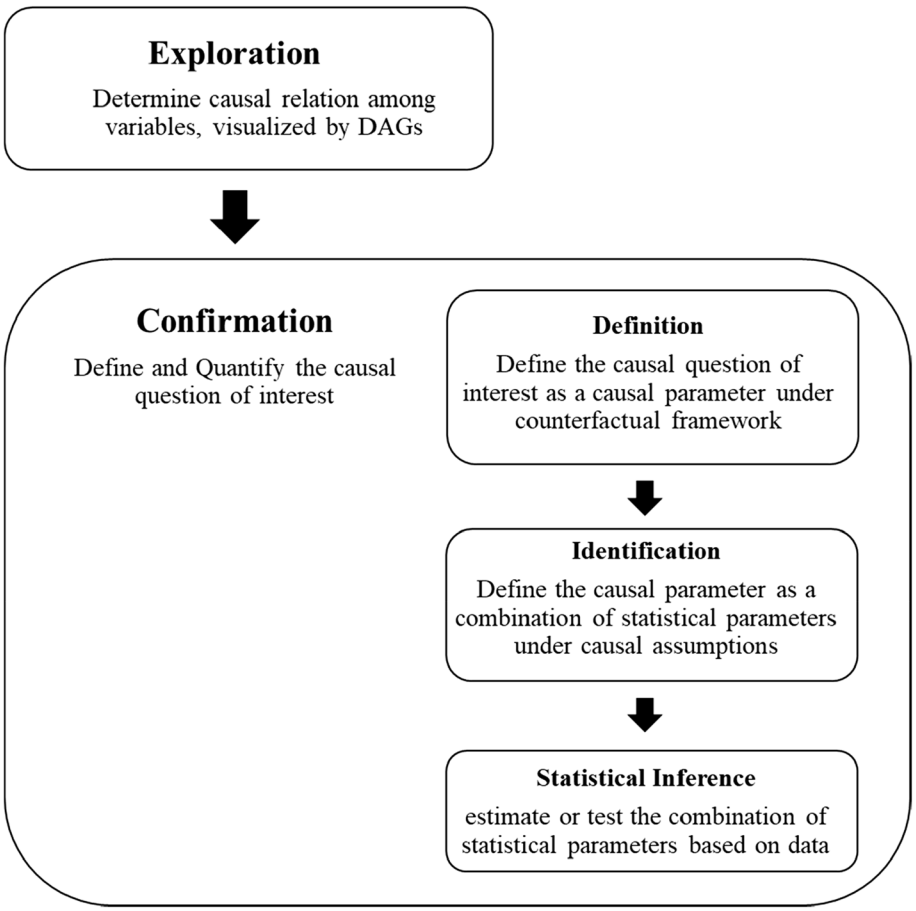

Fig. 1 Framework of causal inference

The main question is at what stage of the causal inference framework should machine learning be positioned. While there are claims that the causal relationship structure can be determined based on data, there is general consensus that the construction of this structure per definition relies on prior substantive information. Similarly, the definition and identification of the causal question is based on substantive knowledge of the investigator(s) on pathophysiology and biology. Against this background, machine learning can be best considered part of the final step pertaining to statistical inference. It thus follows that machine learning per se is per definition insufficient to infer causality.

The positioning of machine learning as part of statistical inference may still prove useful for causal inference. We illustrate this idea by discussing propensity scores and inverse-probability weighting. Propensity scores are a powerful way to control for confounding when the potential covariates are high dimensional. Usually, a propensity score model for treatment status is constructed regressed on all potential covariates. Next, these propensity scores are included in the final causal model instead of all separate covariates. Relating propensity scores to causal inference framework, it follows that constructing the two models belongs to the third step, i.e. statistical inference. Since the covariates are high-dimensional, it is likely insufficient to use regular statistical models to construct a propensity score without model misspecification. It may thus be fruitful to construct the propensity score model using machine learning. Inverse probability weighting (IPW) is another approach in causal inference used for control of confounding and selection bias. First, a model is constructed for treatment based on all potential covariates, and subsequently an outcome model, weighted by the inverse probability of treatment. Similar to propensity scores, these models can be considered part of the third step (statistical inference) and could therefore be improved by application of machine learning. Although machine learning may be powerful to avoid model misspecification for a given set of covariates, note that such approaches should not be used for the initial selection of covariates, as detailed elsewhere [5–7].

In conclusion, explanatory research focuses on identifying and understanding causes of disease and relies heavily on the causal inference framework. This framework involves prior knowledge on biology, knowledge on flow of information, and knowledge on methodological issues in study design, i.e. biases. These aspects cannot be derived from the data, which precludes machine learning per se to reach appropriate causal conclusions. Nevertheless, certain aspects

in causal inference, i.e. those related to model specification and statistical inference might benefit from machine learning.

Funding This study is supported by the grant from Ministry of Science and Technology in Taiwan (No. 108-2636-B-009-001).

References

1. Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. Eur J Epidemiol. 2018;33(5):459–64.
2. Breiman L. Stacked regressions. Mach Learn. 1996;24(1):49–64.
3. Keil AP, Edwards JK. You are smarter than you think: (super) machine learning in context. Eur J Epidemiol. 2018;33(5):437–40.
4. Van der Laan MJ, Rose S. Targeted learning: causal inference for observational and experimental data. Berlin: Springer; 2011.
5. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019;34(3):211–9.
6. Ikram MA. The disjunctive cause criterion by VanderWeele: an easy solution to a complex problem? Eur J Epidemiol. 2019;34(3):223–4.
7. Schneeweiss S. Theory meets practice: a commentary on VanderWeele's 'principles of confounder selection'. Eur J Epidemiol. 2019;34(3):221–2.