Defining Interpretability in Machine Learning for Epidemiology: Challenges and Paths Forward

Andrea Bellavia, PhD

TIMI Study Group, Brigham and Women's Hospital, Harvard Medical School
Department of Environmental Health, Harvard T.H. Chan School of Public Health

abellavia@bwh.harvard.edu

Karolinska Institutet, September 16, 2025

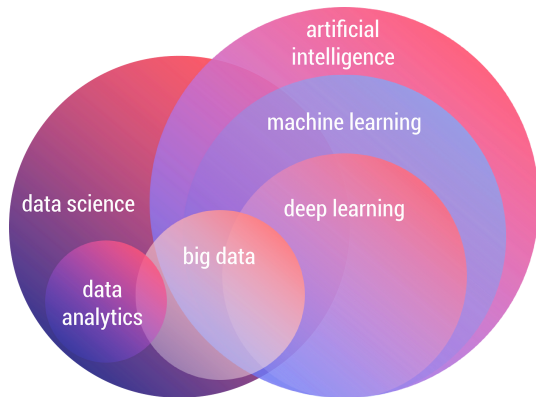# Complex methods for complex data: key considerations for interpretable and actionable results in exposome research

ESSAY | Published: 06 August 2025

(2025)   Cite this article

Marta Ponzano, Ran S Rotem & Andrea Bellavia ✉

1. What is Machine Learning?

1. What is Machine Learning?

1950's definition: *the field of study that gives computers the ability to learn without explicitly being programmed*

- Terminology such as *machine learning, deep learning, data science, statistics* can be confusing and often used interchangeably

- Clinical researchers and epidemiologists frequently view ML as an alternative to classical regression as a contrasting world

- Example from the Journal of Clinical Epidemiology, 2019:

REVIEW | VOLUME 110, P12-22, JUNE 2019   ⇩ Download Full Issue

A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models

Evangelia Christodoulou • Jie Ma • Gary S. Collins • Ewout W. Steyerberg • Jan Y. Verbakel • Ben Van Calster

- ▶ In contrast, here's how logistic regression is introduced in introductory computer science courses. From MIT's ML101:

- ▶ *Logistic Regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable*

- ▶ This binary framing can lead to misleading conclusions and unwarranted skepticism.

The distinction is not about the methods themselves, but rather about data assumptions. From Breiman 2001, "The two cultures":

▶ One [culture] assumes that the data are generated by a known stochastic model
▶ The other [culture] uses algorithmic models and treats the data mechanism as unknown

- Focus here is on scenarios where data complexity makes it difficult to formally define all assumptions, making data-driven approaches particularly appealing

- Even regression-based methods can help bridge this gap—for example, by using splines to relax linearity assumptions.

- A key challenge with data-driven approaches is overfitting: achieving high accuracy at the expense of generalizability

- Machine Learning robustly addresses this with techniques such as cross-validation and bootstrap resampling

2. Machine Learning in Epidemiology

2. Machine Learning in Epidemiology

Oversimplifying, two main factors drive the growing appeal of ML in epidemiology:

1. Reality is complex
2. The big data revolution enables the collection of data that can capture this complexity

- ▶ The Big data revolution is profoundly transforming epidemiology and clinical research.

- ▶ Vast amounts of data can now be collected rapidly, helping to uncover complex patterns and support more informed decision-making.

- ▶ Example of big data impacting clinical research:
  - ▶ Omics data
  - ▶ Hemodynamic parameters
  - ▶ Imaging
  - ▶ Data from wearable devices
  - ▶ High-resolution environmental exposures
  - ▶ . . .

▶ Complex data refers not only to the number of variables but also to the biological mechanisms involved - often featuring non-linearities and non-additive effects

Complexity rather arises from several features such as:

▶ The multitude of factors influencing health outcomes
▶ The intricate biological and social mechanisms involved
▶ The timing and duration of exposures

New research paradigms

Emerging frameworks shaped by this big data revolution include:

- Precision medicine: "Precision medicine, sometimes known as personalized medicine, is an innovative approach to tailoring disease prevention and treatment that takes into account differences in people's genes, environments, and lifestyles."

- Exposome: "the totality of environmental exposures, both internal and external, that a person experiences throughout their life, and how these exposures impact their health"[1]

---

[1] See Vermeulen et al, Science 2020

Analytical challenges

- To what extent can classical statistical approaches accommodate the complexity of these new scientific paradigms?

- Increasing interest in analytical methods that can handle complexity while mitigating overfitting:
  - Semi-parametric and non-parametric regression
  - Machine Learning
  - Deep learning

Complex data and clinical prediction

▶ Increasing interest in analytical methods that can handle complexity while mitigating overfitting:

  ▶ How do we balance the desire for individual precision with the need of parsimonious and pragmatic models?
  ▶ Are there contexts where one should be prioritized over the other?
  ▶ Does data complexity always imply interpretive complexity?
  ▶ How should complex results be communicated—to peers, or to patients?

▶ Moreover, Incorporating omics data into individualized prediction introduces challenges such as biological interactions

Complex data and etiological research

Similar challenges arise in etiological research (e.g.within frameworks like the exposome) with some other field-specific questions:

- ▶ How can we preserve the rigor of causal inference when incorporating complex predictors?
- ▶ How do we address confounding, effect modification, and mediation?
- ▶ How should we incorporate social determinants of health?

Motivating example 1 [clinical prediction]

- Research grant on proteomic discovery for heart failure risk in diabetic patients[2]
- $\sim$ 70 cases and 70 controls
- Blood sample used to quantify $\sim$ 400 proteins
- Validation of selected proteins in a larger cohort of 14,000 trial participants

---

[2]Preliminary results recently presented here

Analytical challenge: how to identify the top predictors (out of ~ 400) with only 70 events?

▶ Individual regressions are unreliable due to co-confounding and require multiple comparison assessment
▶ Multivariable models won't converge ($p > n$)
▶ High isk of overfitting

Motivating example 2 [exposomic research]

▶ Identification of potential risk factors for ALS[3]
▶ Focus on history of medication
▶ 501 ALS cases and matched controls
▶ Data on over 1000 medications

---

[3]Rotem et al 2024

Analytical challenge: can we use these data to identify potential risk factors?

In addition to previous concerns, etiological studies possibly require:

► Confounding control
► Causal interpretation
► Potential for translation into interventions

3. ML in Epidemiology: Specific Challenges

"Black-box" models

- ▶ Machine Learning methods are often labeled as *"black box"* approaches
- ▶ This term reflects the fact that their internal mechanisms do not need to be explicitly specified or evaluated
- ▶ While this is generally not a problem for prediction tasks[4] it can be a limitation in epidemiological research
- ▶ For example, imagine we feed a set of variables into a machine learning model and obtain a highly accurate prediction tool. We still face several important questions:

---

[4]Most ML methods were originally developed with predictive goals in mind

- ▶ Are these variables feasible to collect for all patients? If not, which ones should be prioritized?
  - ▶ Example: Collecting omics data is time-consuming and expensive. What if a decision is needed within hours?
- ▶ Is collecting all these variables ethically appropriate?
  - ▶ Example: How should we handle sensitive data like social determinants of health?
- ▶ Are some variables causally related? What mechanisms underlie the prediction?
  - ▶ Example: If both BMI and physical activity are included, understanding their interaction could inform personalized recommendations.

A primary goal of epidemiological research is to identify actions to be taken—whether public health recommendations, clinical interventions, or policy changes. A thorough understanding of mechanisms is key

From the introduction of the exceptional Molnar's textbook[5] on the topic:

"computers usually don't explain their predictions, which can cause many problems, ranging from trust issues to undetected bugs [...] model-agnostic methods for interpreting black box models [...] to explain individual predictions [...] to get insights about the more general relations between features and predictions"

Interpretable ML methods include:

▶ Tools for ranking predictors
▶ Visualizations of functional relationships
▶ Methods for ranking interactions

---

[5]Available for free here

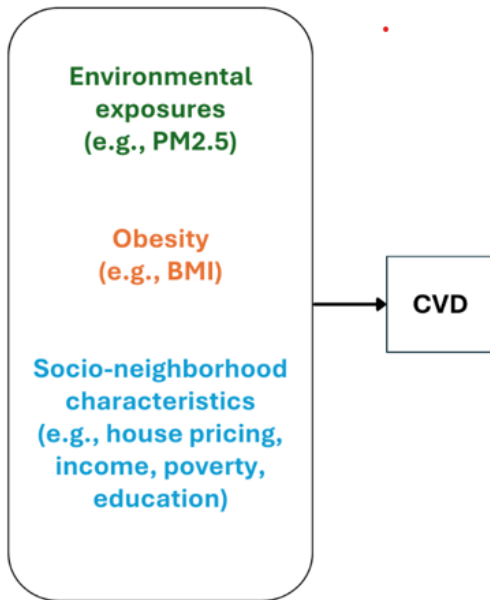# 4. ML Interpretability in Epidemiology

4. ML Interpretability in Epidemiology

- Common methods for interpretable ML address the important—but preliminary—aspect of statistical interpretability
- These tools help assess and explain a model's internal mechanisms (e.g. displaying functional forms without linearity assumptions, as splines do)

- ▶ While statistical interpretability is a useful first step, it leaves several key questions unanswered.
- ▶ Remember: the ultimate goal of epidemiological research is to identify *actions to be taken*
- ▶ The next question becomes: are the findings causally interpreted

Example:

- ▶ Research study aimed at investigating the determinants of cardiovascular diseases (CVD) risk in the general population using an exposome-wide approach
- ▶ Hundreds of potential risk factors are considered, including:
    - ▶ Pollutant and chemical exposures
    - ▶ Dietary factors
    - ▶ Anthropometric measures
    - ▶ Neighborhood characteristics
- ▶ Interpretable ML is used to screen for potential risk factors

Environmental exposures (e.g., PM2.5)

Obesity (e.g., BMI)

Socio-neighborhood characteristics (e.g., house pricing, income, poverty, education)

CVD

- ► Suppose the top-ranked predictors are:
    - ► 1. Obesity
    - ► 2. Neighborhood housing price
    - ► 3. PM2.5 levels
    - ► 4. Socioeconomic status (SES)
- ► Can we conclude that BMI is the main predictor of CVD?
- ► What about the relationship between PM2.5 and BMI?
- ► BMI might be a mediator in the PM2.5 - >CVD pathway. If so, including BMI in a regression model—or in ML—could obscure the true causal relationship.

- ▶ More broadly, when feeding hundreds of variables into a machine learning model, you are implicitly making causal assumptions:
  - ▶ What happens if you include highly correlated variables?
  - ▶ What if both confounders and mediators are present?
  - ▶ How do you handle repeated measurements over time?

▶ Causality in ML is an active area of research[6] with specific methodological extensions under development
▶ Still, there are practical steps we can take to improve causal interpretability:
▶ First, we could use high-dimensional DAGs to hypothesize relationships between key variables

---

[6]Methods overview, Moccia et al.
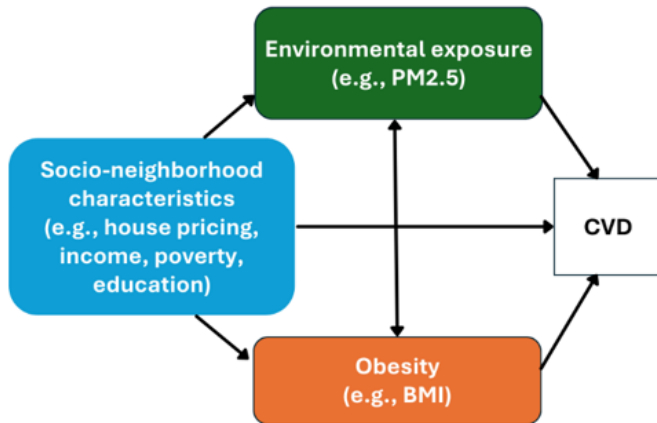
These hypotheses can guide:

- ▶ Running separate ML models within each subcategory
- ▶ Applying methods for high-dimensional mediation analysis
- ▶ Performing variable selection among highly correlated predictors

When using ML for etiological research, don't forget about causality. The machine doesn't distinguish between confounders, mediators, or main predictors—it treats all variables equally.

Actionable interpretability

▶ A third level of interpretability is required to identify potential actions.
▶ Suppose we've identified a set of pathways and are confident they represent causal relationships.
▶ The next key questions become:
  ▶ Are these causal pathways actionable?
  ▶ Will these actions be effective across all individuals in the population?

Social determinants of health (SDH)

- In our example, socioeconomic status (SES) was identified as a top predictor.
- However, SES offers limited room for direct intervention.
- SDH encompasses multiple dimensions—social, economic, environmental—that are often difficult to quantify individually.
- Despite this, SDHs are central to exposomic research and must be integrated. The key question is: how?
- This is a relevant topic in epidemiology broadly, not just in ML applications.

- First, we must unpack the meaning of social factors and identify proximal, actionable determinants
- For example, collecting more detailed data on components of SES—such as income, housing type, or education—can help identify clearer intervention pathways.
- Ideally, this should be considered during study design, to ensure the necessary data are collected.

- ▶ A key component of ML applications in epidemiology is a careful, human-driven selection of variables to be fed into the model.
- ▶ This selection should be informed by causal assumptions, ethical and practical considerations, and guided by the goal of identifying actionable interventions and public health recommendations.

Assessing effect modification with complex data

- ▶ Final question: are the identified actions effective for all individuals?
- ▶ In regression, this is typically addressed through interaction terms or stratified analyses.
- ▶ Similar strategies can be applied in ML settings:
- ▶ A simple first step is to run separate models within population subgroups.

- ► More formal ML approaches for stratifying across multiple characteristics have been developed, particularly in clinical epidemiology
- ► These methods aim to assess multivariable heterogeneity of treatment effects (HTE)[7]

---

[7]See Bellavia/Murphy 2025, Circulation (link here) for a general overview

5. Summary and Conclusions

5. Summary and Conclusions

- The big data revolution is driving a growing and well-justified interest in modern methodologies that can handle data complexity and relax the assumptions and limitations of classical regression.
- Addressing overfitting — through techniques like training/validation splits and cross-validation — is essential when adopting these approaches.
- In epidemiology, several additional challenges arise:
  - Interest in understanding underlying mechanisms
  - Practical and ethical considerations

▶ When applying machine learning in epidemiology, interpretability should span all dimensions:

   ▶ Statistical: understanding model behavior
   ▶ Causal: identifying meaningful relationships
   ▶ Actionable: guiding interventions and recommendations

▶ While this is an active area of research, many practical steps can already be taken using existing tools and frameworks.

Thank you for your attention!

You can download these slide from https://andreabellavia.github.io/presentations/