

# Improving Trust and Understanding in Biostatistics: The Importance of Interpretable Machine Learning Models

Linie Safeti\*

Department of Biostatistics, University of Dhaka, Dhaka, Bangladesh

## Introduction

In the rapidly evolving field of biostatistics, the integration of machine learning (ML) has opened new doors for analyzing complex datasets and making predictions. However, with these advancements comes the challenge of making machine learning models more interpretable and transparent, especially when they are used in sensitive fields like healthcare and epidemiology. As biostatistical models increasingly guide critical decisions—such as disease diagnosis, treatment planning, and public health strategies—the need for models that are not only accurate but also interpretable has never been more urgent. Interpretability in machine learning refers to the ability to understand and explain how a model arrives at its predictions. While machine learning models, particularly deep learning models, are often praised for their high accuracy, they are often regarded as “black boxes,” where the decision-making process is obscure even to the experts who created them. This lack of transparency can be problematic in biostatistics, where the stakes are high, and decisions based on model outputs can directly affect patient outcomes or public health policies [1].

## Description

In biostatistics, models are frequently applied to critical areas, such as predicting disease outcomes, understanding the effects of treatments, or identifying health disparities. These applications require not only reliable results but also explanations that can be trusted by medical professionals, policymakers, and the public. When machine learning models are used in healthcare or public health, the people affected by the decisions—patients, clinicians, and communities—need to trust the model's predictions. If a model cannot explain why it made a certain decision, it is much harder for users to trust it. For example, if a machine learning model predicts that a patient is at high risk for a particular disease, clinicians need to understand the factors that influenced this prediction to make informed decisions about patient care. Without transparency, there is a risk that the model could be viewed as a “black box,” undermining its credibility. Machine learning models, particularly those trained on large datasets, can inadvertently learn biases present in the data. For instance, a model trained on biased demographic data may produce discriminatory predictions. Interpretability allows for the identification of such biases. By understanding how a model makes decisions, it becomes easier to spot whether certain factors are being unfairly weighted or whether there are discrepancies in the predictions for different groups of people [2].

In healthcare and other sectors, models often need to meet regulatory standards to be deemed trustworthy and safe for use. In many cases, these regulations require that models be interpretable. For example, the European Union's General Data Protection Regulation (GDPR) includes provisions

for the “right to explanation,” which ensures that individuals can ask how automated decisions were made about them. This requirement necessitates that biostatistical models be transparent, so individuals can challenge decisions or request further clarifications about how their data was used. Many machine learning models are used to support clinicians in making treatment decisions or diagnosing conditions. However, clinicians need to be able to understand the model's reasoning to make better decisions themselves. If the model cannot explain why it suggests a certain treatment or diagnosis, clinicians may be less likely to follow the model's recommendations, even if they are statistically accurate. Interpretability helps bridge this gap between data science and clinical practice, providing actionable insights that improve healthcare outcomes [3].

While the importance of interpretability is clear, achieving it in machine learning models, especially complex ones, is not always straightforward. As machine learning models, particularly deep learning algorithms, become more complex, they become increasingly difficult to interpret. These models often involve numerous layers of computation, each contributing to the final output. While these models may achieve impressive accuracy, their decision-making processes can be highly intricate and opaque. In biostatistics, where data is often highly multidimensional, simplifying these models without sacrificing predictive power is a delicate balancing act. In some cases, simpler models—such as decision trees or logistic regression—offer more transparency but may not achieve the same level of accuracy as more complex models like neural networks. This trade-off between interpretability and predictive performance is a central challenge in biostatistics. Researchers and practitioners must often make decisions about which models to use based on the specific context and goals of the analysis. While there has been significant progress in developing interpretability methods, there is still no universal toolkit that works well for all types of machine learning models. Different models and tasks may require different interpretability techniques, such as feature importance analysis, SHAP (Shapley Additive Explanations), or LIME (Local Interpretable Model-agnostic Explanations). The lack of standardization can create confusion and hinder widespread adoption of interpretable models in biostatistics [4].

**Model-Agnostic Interpretability Methods:** For more complex models, model-agnostic interpretability methods can help explain predictions. Techniques like LIME and SHAP provide explanations that are independent of the model type, making them suitable for a wide range of machine learning algorithms. These methods work by approximating complex models with simpler, interpretable models that can reveal which features are driving the predictions. One way to make a model more interpretable is through careful feature selection and engineering. By choosing a limited set of important features and engineering them in ways that are meaningful and understandable, biostatisticians can ensure that the model focuses on the most relevant information. This not only improves interpretability but can also lead to better model performance. To ensure that machine learning models are transparent, it is essential to have transparent data practices as well. Understanding the data used to train models, including how it was collected, cleaned, and preprocessed, is a crucial part of ensuring that the model's predictions are fair and reliable. Researchers should document their data sources and methodologies to help others understand how the models were developed and to identify potential sources of bias [5].

## Conclusion

The integration of machine learning in biostatistics has the potential to

\*Address for Correspondence: Linie Safeti, Department of Biostatistics, University of Dhaka, Dhaka, Bangladesh, E-mail: [linie1@edu.com](mailto:linie1@edu.com)

**Copyright:** © 2024 Safeti L. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Received:** 27 November, 2024, Manuscript No. jbmbs-25-158902; **Editor assigned:** 29 November, 2024, Pre QC No. P-158902; **Reviewed:** 13 December, 2024, QC No. Q-158902; **Revised:** 18 December, 2024, Manuscript No. R-158902; **Published:** 26 December, 2024, DOI: 10.37421/2155-6180.2024.15.241

revolutionize healthcare and public health by providing powerful predictive models that can inform decision-making. However, for these models to be trusted and effective, they must be interpretable. Transparency in how a model makes its predictions not only helps to build trust among clinicians, patients, and the public, but also enables better decision-making, identification of biases, and compliance with ethical and regulatory standards. The importance of interpretability in machine learning cannot be overstated, particularly in high-stakes fields like biostatistics. By adopting interpretable machine learning models and leveraging tools that enhance transparency, biostatisticians can ensure that their models not only achieve high accuracy but also contribute to better, more equitable health outcomes.

---

## Acknowledgement

None.

---

## Conflict of Interest

None.

---

## References

1. Rudin, Cynthia. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." *Nat Mach Intell* 1 (2019): 206-215.
2. Karatekin, T., Sancak, S., Celik, G. and Topcuoglu, S., et al. "Interpretable machine learning in healthcare through generalized additive model with pairwise interactions (GA2M): Predicting severe retinopathy of prematurity." In 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML) 61-66.
3. Kamath, Uday and John Liu. "Explainable artificial intelligence: An introduction to interpretable machine learning." *Springer* (2021).
4. He, Xinyu, Ruoyu Tang, Jie Lou and Ruiqi Wang. "Identifying key factors in cell fate decisions by machine learning interpretable strategies ." *J Biol Phys* (2023): 1-20.
5. Kos, Marko and Iztok Kramberger. "A wearable device and system for movement and biometric data acquisition for sports applications." *IEEE* 5 (2017): 6411-6420.

**How to cite this article:** Safeti ,Linie. "Improving Trust and Understanding in Biostatistics: The Importance of Interpretable Machine Learning Models." *J Biom Biosta* 14 (2024): 241.