# CHAPTER 2

# EXPLAINABLE ARTIFICIAL INTELLIGENCE IN EVIDENCE BASED MEDICAL STATISTICS EDUCATION

**Debasish Paul**
Ph.D. Researcher, IIT Kharagpur,West Bengal, India

***Abstract***

*Explainable Artificial Intelligence (XAI) in evidence-based medical statistics education can be described as a revolutionary innovation. It helps medical students to acquire better understanding of medical statistics. Although, there have some current challenges in the educational environment of medical institution. Statistical education has largely focused on output from algorithms and the interpretation of numbers. Explainable Artificial Intelligence allows students to understand how predictive outputs are influenced by individual clinical variables. This capability promotes a more in-depth understanding of the fundamental principles of statistics. On the other hand, it promotes the orientation of future clinical decisions toward evidence-based medicine. Through interactive visualization, model explanation and case-based learning scenarios, students explore complex relationships in statistics. They also identify biases and assess model reliability. Applying XAI in medical education, students acquire different skills like questioning and interpreting AI-driven recommendations. Generally, XAI in medical statistics education fits perfectly in the chasm between computational approaches and clinical reasoning. So, it turns future healthcare professionals into a differently trained analytical expert.*

*Keywords:* Explainable Artificial Intelligence (XAI)*,* Medical Statistics Education*, Clinical* Practice*, Interactive Visualization, Black Box***.**

**Introduction:**

Artificial Intelligence (AI) has significantly transformed healthcare by enhancing processes from data generation to advanced analysis and interpretation. Despite these advancements, medical statistics education continues to rely largely on formula-based instruction and software-driven outputs, which, while effective for building foundational knowledge, often fail to promote deeper cognitive reasoning among medical students. As a result, learners may struggle to interpret statistical outcomes critically and apply them meaningfully in clinical contexts. In evidence-based medicine, however, interpretive competence is more crucial than mere numerical literacy, as clinicians must evaluate data quality, understand uncertainty, and make informed decisions for patient care (Harden, 2017; Shortliffe, 2018). In this context, Explainable Artificial Intelligence (XAI) emerges as a promising pedagogical innovation that bridges the gap between computational results and human understanding. XAI enhances transparency and interpretability, enabling learners to comprehend how and why specific outputs are generated (Arrieta et al., 2020; Cutillo et al., 2020). This aligns with the need for developing higher-order cognitive skills, as emphasized in Bloom's theory of mastery learning (Bloom, 1984). Techniques such as model-agnostic explanations and interpretable machine learning approaches further support this educational transformation by fostering trust, usability, and critical thinking (Doshi-Velez & Kim, 2017; Ribeiro et al., 2016; Lundberg & Lee, 2017). Moreover, the integration of human-in-the-loop systems ensures active engagement and continuous learning, particularly in complex healthcare environments (Holzinger, 2016; Holzinger et al., 2020). Scholars have also emphasized that in high-stakes domains like healthcare, interpretable models should be prioritized to ensure ethical and responsible decision-making (Rudin, 2019). Thus, XAI not only strengthens the interpretive capabilities of medical students but also aligns with the broader vision of high-performance medicine, where human expertise and artificial intelligence converge to improve clinical outcomes (Topol, 2019).

**Literature review:**

**(i) Traditional Approaches to Medical Statistics Education:** Although traditional approaches were sufficient for providing basic technical competencies (such as performing statistical tests like z-test, t-test, and chi-square), they largely emphasized procedural knowledge over conceptual understanding. In other words, students often learned how to arrive at particular results without fully understanding the underlying reasons for those outcomes.

**(ii) Emergence of Artificial Intelligence in Medical Education:** The introduction of artificial intelligence brought adaptive learning systems (as proposed by Benjamin Bloom), predictive analysis, and simulation-based training (aligned with the "SPICES model" proposed by Harden R. M.) into the medical education curriculum. These innovations increased efficiency and personalization in learning. However, most of these AI applications functioned as "black boxes," thereby limiting their educational value, particularly when used to teach reasoning processes.

**(iii) Explainable Artificial Intelligence (XAI):** XAI techniques addressed the lack of transparency in AI models by providing interpretable outputs. Various explainability methods based on feature attribution, local explanations, and visual analytics were proposed to facilitate understanding of how specific variables influenced model predictions. In medical statistics, such transparency was crucial because it built trust in the system, ensured accountability, and supported the ethical use of decision-making processes.

**(iv) Explainable Artificial Intelligence in Educational Contexts:** Several scholars, including Finale Doshi-Velez, Been Kim, and Cynthia Rudin, pointed out that explainability techniques enhanced learning by improving student engagement and critical thinking. They observed that when students understood how inputs were transformed into outputs, they were better able to question underlying assumptions, reflect on their understanding, and explore alternative interpretations.

**Conceptual framework:**

The place of XAI within medical statistics education have been summarized in the following three dimensions:

- **Transparency:** makes statistical and Artificial Intelligence processes visible and understandable.
- **Interactivity:** allows students to manipulate variables and to observe outcomes.
- **Clinical relevance:** constructs link between statistical reasoning and real-world medical decisions.

**Role of XAI in enhancing learning:**

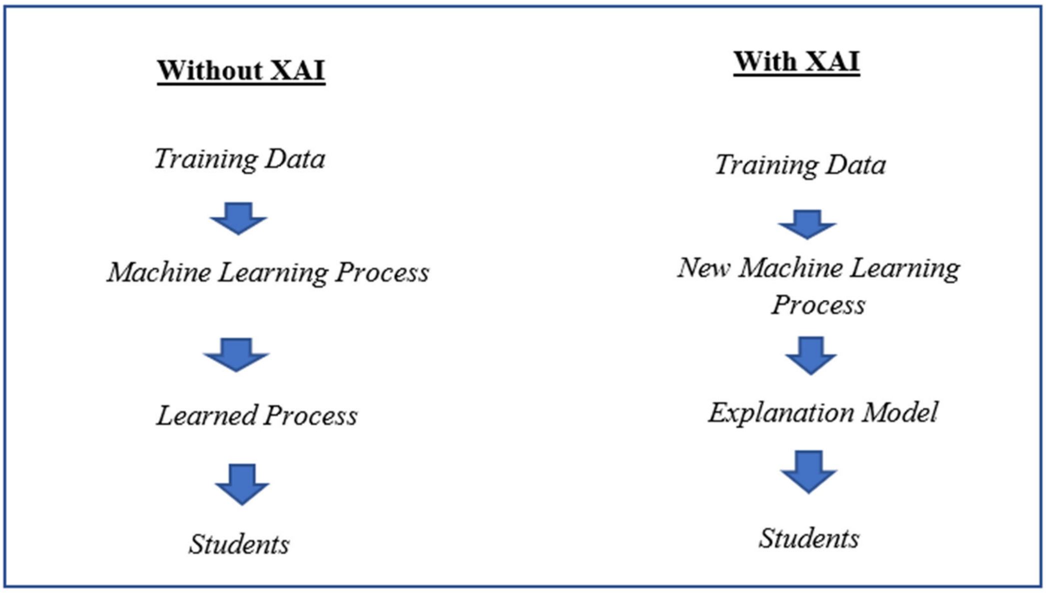
**(i) Improving conceptual understanding:**

XAI allows students of medical statistics to visualize relationship between input variables and predicted outcome. Hence core statistical concepts including probability, correlation and regression can be revisited within XAI.

**(ii) Encouraging critical thinking:**

XAI builds thinking process of students because instead of accepting model's predictions at face value, they are encouraged to ask questions and to think about possible biases in the data. It's very essential which have been used for training the model, to find out limitations of the model and to check the reliability of the predictions.

**Flow Chart -1 Represent the XAI Model**




**Source**: Developed by Researcher

**(iii) Bridging theory and clinical practice:**

There is often a disconnection between the abstract theory of statistics and the clinical reality in which medical students find themselves. In this way XAI facilitates linking the theoretical framework of statistics with clinical cases. It helps medical students to understand that statistical evidences have played the most crucial role in developing diagnostic, prognostic and therapeutic medical decisions.

**Pedagogical strategies for integration:**

**(i) Interactive visualization tools:**

The use of visual dashboards and explainable interfaces like Tableau, Power BI, R Shiny etc. promotes the dynamic manipulation of data sets by students. Such interactivity brings statistical associations into more concrete and graspable forms.

**(ii) Case based learning:**

Presenting students with real-world clinical cases supplemented by explanations derived from XAI would contextualize the application of statistical reasoning in realistic clinical decision-making. It would simultaneously foster the development of analytical and decision-making skills.

**(iii) Guided exploration:**

Educators could prepare assignments prompting students to explore the consequences of varying input variables on model predictions. This would promote active learning and a deeper level of engagement.

**Flow Chart-2 Represent Black -Box AI**

**Tools used to interpret the decision-making process of black-box AI**

***LIME*** ***SHAP*** ***DALE*** ***ELI5*** ***SKATER***

**Source**: Developed by Researcher

**Challenges and limitations:**

**(i) Risk of cognitive overload:**

Being very detailed, XAI explanations may be over complex, potentially leading to cognitive overload, particularly for novice students. Therefore, it is necessary to balance the richness and clarity of explanations.

**(ii) Misinterpretation of outputs:**

Students may misinterpret AI explanations by confusing association with causality or overestimating the explanatory power of the AI model. Thus, appropriate guidance is necessary to prevent such misinterpretations.

**(iii) Resource and training constraints:**

The adoption of XAI in teaching requires specific technological infrastructure for example computer, software tools, High speed Internet connection etc. and staff training which may not be available everywhere specially in Indian context.

**Future directions:**

Explainable Artificial Intelligence (XAI) in medical statistics education is still in its developmental stage and represents a significant avenue for future pedagogical innovation. As healthcare increasingly integrates artificial intelligence into clinical decision-making, the need for medical students to not only use but also understand AI-driven outputs has become essential. Traditional approaches to teaching medical statistics, which emphasize formulae and software-generated results, often fail to cultivate deeper interpretive and analytical skills. In this context, XAI offers a transformative opportunity to bridge the gap between computational processes and human reasoning by making complex models more transparent and understandable (Arrieta et al., 2020; Cutillo et al., 2020).

One of the important future directions in this domain is the development of student-friendly XAI platforms. These platforms should be designed with pedagogical sensitivity, enabling learners to interact with models, visualize decision pathways, and explore how different variables influence outcomes. By simplifying complex algorithms into intuitive representations, such platforms can enhance conceptual clarity and promote active learning. The importance of human-centered and interactive machine learning systems has been emphasized in health informatics, where user engagement plays a critical role in knowledge acquisition (Holzinger, 2016). Moreover, explainability tools such as SHAP and LIME demonstrate how model predictions can be broken down into interpretable components, thereby fostering trust and understanding among learners (Lundberg & Lee, 2017; Ribeiro et al., 2016).

Another important area for future research is the need for empirical studies that evaluate the impact of XAI on learning outcomes in medical education. While theoretical discussions highlight the potential benefits of XAI, there is a lack of systematic evidence demonstrating its effectiveness in improving students' interpretive skills, critical thinking, and clinical reasoning. Drawing from educational research, particularly the emphasis on mastery learning, it is clear that innovative teaching methods must be assessed rigorously to determine their efficacy (Bloom, 1984). Empirical investigations can provide insights into how XAI tools influence cognitive engagement and whether they lead to better application of statistical knowledge in real-world medical contexts.

The inclusion of ethics and bias in the medical curriculum is another crucial direction. AI systems, including those used in healthcare, are susceptible to biases that can lead to inequitable outcomes. Therefore, medical students must be trained to critically evaluate not only the outputs of AI systems but also the ethical implications underlying their use. Understanding issues such as algorithmic bias, fairness, and transparency is essential for responsible clinical practice (Rudin, 2019). XAI can support this by making hidden biases more visible and enabling learners to question and interpret results with a critical perspective, thereby aligning with the broader goals of trustworthy and ethical AI (Doshi-Velez & Kim, 2017).

Furthermore, enhanced collaboration between data scientists and medical educators is vital for the successful integration of XAI into medical statistics education. Such interdisciplinary partnerships can ensure that educational tools are both technically robust and pedagogically effective. Clinical decision support systems have already demonstrated the value of combining computational expertise with medical knowledge to improve patient care (Shortliffe, 2018). Extending this collaborative approach to education can lead to the development of curricula that are aligned with the evolving demands of AI-driven healthcare.

XAI should not be viewed as an "add-on" technology but as a fundamental shift in teaching and learning practices within medical statistics education. By promoting transparency, interpretability, and critical engagement, XAI has the potential to redefine how medical students understand and apply statistical knowledge. This aligns with the vision of high-performance medicine, where human intelligence and artificial intelligence work synergistically to enhance clinical outcomes and decision-making (Topol, 2019).

**Conclusion:**

Explainable Artificial Intelligence (XAI) has a profound impact on the teaching and learning process of medical statistics by transforming abstract computational processes into meaningful and interpretable knowledge. Traditionally, students entering the medical field encounter statistical tools and algorithms as "black boxes," where outputs are generated without a clear understanding of the internal mechanisms. This often limits their ability to critically engage with data and undermines their confidence in applying statistical reasoning in clinical contexts. However, XAI changes this paradigm by making the internal logic of computational models visible, interpretable, and interactive, thereby enhancing both conceptual understanding and analytical thinking (Arrieta et al., 2020; Cutillo et al., 2020).

Through XAI, the computational engine of statistical models becomes a transparent system where learners can explore how inputs are transformed into outputs. This transparency allows students to visualize relationships among variables, assess the contribution of different factors, and understand the reasoning behind predictions. Such an approach aligns with the principles of interactive and human-centered machine learning, which emphasize the importance of user engagement in knowledge construction (Holzinger, 2016). By actively involving students in the learning process, XAI fosters deeper cognitive engagement and promotes critical thinking skills that are essential in medical education.

Furthermore, XAI helps bridge the gap between numerical reasoning and clinical reasoning, which is a critical challenge in evidence-based medicine. While traditional statistical education equips students with computational skills, it often falls short in enabling them to interpret results within real-life clinical scenarios. XAI addresses this limitation by contextualizing statistical outputs and linking them to clinical decision-making processes. This integration ensures that students not only understand the "how" but also the "why" behind statistical results, thereby enhancing their ability to apply knowledge in patient care (Shortliffe, 2018). As a result, learners develop a more holistic understanding of medical data, which is essential for making informed and evidence-based decisions.

The pedagogical value of XAI is also supported by educational theories such as Bloom's mastery learning, which emphasizes the importance of deep understanding and individualized learning experiences (Bloom, 1984). XAI tools can simulate personalized learning environments by allowing students to explore models at their own pace, experiment with different scenarios, and receive immediate feedback. This not only improves comprehension but also builds confidence in handling complex statistical concepts. Additionally, techniques such as model-agnostic explanations and interpretable machine learning methods, including those proposed by Ribeiro et al. (2016) and Lundberg and Lee (2017), further enhance students' ability to critically evaluate and trust computational outputs.

Another significant contribution of XAI lies in promoting ethical awareness and responsible use of AI in healthcare. By making model decisions transparent, XAI enables students to identify potential biases and limitations in data and algorithms. This is particularly important in high-stakes medical contexts, where incorrect or biased interpretations can have serious consequences. Scholars have argued that interpretable models should be prioritized in such domains to ensure accountability and trustworthiness (Rudin, 2019; Doshi-Velez & Kim, 2017). Moreover, tools like the System Causability Scale provide frameworks for evaluating the quality of explanations, thereby supporting more rigorous and reflective learning (Holzinger et al., 2020).

## References


Arrieta, A. B.-R., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Bloom, B. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Cutillo, C. M., Sharma, K. R., Foschini, L., Kundu, S., Mackintosh, M., & Mandl, K. D. (2020). Machine intelligence in healthcare—Perspectives on trustworthiness, explainability, usability, and transparency. *npj Digital Medicine, 3*(1), 47.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. https://arxiv.org/abs/1702.08608

Harden, R. M. (2017). *Essential skills for a medical teacher: An introduction to teaching and learning in medicine*. Elsevier.

Holzinger, A. (2016). Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics, 2*, 119–131.

Holzinger, A., Carrington, A., & Müller, H. (2020). Measuring the quality of explanations: The system causability scale (SCS). *KI - Künstliche Intelligenz, 34*, 193–198.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS).*

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215.

Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA, 320*(21), 2199–2200.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56.