Addressing Model Drift through Advanced Uncertainty Quantification

Jules Udahemuka, Carnegie Mellon University,
Advisor: Dr. John Kalafut, Asher Orion Group

As FDA-approved AI models proliferate in healthcare, ensuring their reliability and safety over time becomes increasingly critical. Model drift, resulting from changes in input data distributions, poses significant challenges to performance monitoring and regulatory compliance. This proposal outlines a novel non-intrusive methodology for quantifying uncertainty using Monte Carlo Dropout, achieving real-time performance monitoring while preserving FDA validation. This framework has already demonstrated its merit, securing a $3000 prize at a Nucleate BioHack.

Introduction

FDA-approved AI systems are transforming modern healthcare by offering automated diagnostics, enhanced workflows, and decision support. Initially developed for a Nucleate BioHack, this methodology won a $3000 prize, recognizing its potential to address critical challenges in AI-driven healthcare systems. However, the post-deployment reliability of these systems remains a critical concern due to:

- **Model Drift**: Variations in input data distributions over time can lead to degraded model performance, posing risks to patient safety and care outcomes (1), (2).
- **Black-Box Nature**: The opacity of many AI models limits the ability to identify and address reliability concerns.
- **Regulatory Constraints**: Regulatory frameworks, such as FDA approval processes, restrict direct modifications to validated models, complicating efforts to maintain system integrity (3).

These challenges underscore the need for robust uncertainty quantification techniques to monitor and mitigate model drift effectively. Existing research highlights the consequences of model drift, including compromised diagnostic accuracy and regulatory violations (1), (2). This proposal introduces a non-intrusive uncertainty quantification framework leveraging Monte Carlo Dropout to address these issues.

Proposed Methodology

The proposed methodology builds upon Monte Carlo Dropout as a foundational approach to enable robust uncertainty estimation in FDA-approved AI models. Monte Carlo Dropout involves maintaining dropout layers active during inference to generate multiple stochastic predictions (4). This approach leverages the variance of these predictions to calculate uncertainty scores and derive statistical confidence intervals, providing actionable insights into model reliability. By quantifying prediction uncertainty, the methodology enables a more transparent and reliable deployment of AI systems in healthcare settings.

To complement this, a rigorous statistical analysis framework is incorporated to continuously monitor input-output data distributions. Using statistical metrics such as Jensen-Shannon Divergence (5), the framework identifies drift in data distributions that may signal performance degradation. Alerts are triggered when significant shifts are detected, allowing early intervention to mitigate potential risks to patient safety and system reliability.

The proposed methodology also emphasizes maintaining the integrity of FDA-approved models by employing non-intrusive techniques. By operating externally, this approach avoids direct modifications to validated models, ensuring compliance with regulatory frameworks such as the FDA's SaMD guidelines (3). This design ensures that the framework can be seamlessly integrated into existing clinical workflows, preserving the validated status of AI systems while addressing performance concerns.

A critical extension of this approach involves integrating real-time monitoring capabilities. These include visualizing confidence scores, tracking trend analysis, and automating anomaly detection to provide proactive insights. Such tools empower radiologists, safety teams, and medical device manufacturers to preemptively address potential model failures or drift, enhancing system reliability across diverse clinical settings.

Additionally, the incorporation of adaptive monitoring techniques is proposed to dynamically adjust thresholds for drift detection based on evolving operational contexts. This adaptability allows the system to remain effective in environments with changing clinical requirements or population shifts, ensuring that patient safety remains uncompromised over time.

Proposed Research Directions

1. Improvements on the Medical Imaging Models: One key direction is to explore the literature on out-of-distribution prediction methods and evaluate their applicability to medical AI models. These methods could help identify when models encounter data significantly different from their training distribution, thereby mitigating risks associated with unreliable predictions. Additionally, research can focus on developing techniques to make models aware of their mistakes, enabling them to flag potentially erroneous outputs proactively.

2. Data Augmentation Techniques: The development of innovative data augmentation techniques tailored to healthcare applications is another crucial area. These methods should mimic real-life datasets as closely as possible, incorporating variations observed in clinical practice. Effective data augmentation could enhance model robustness and performance across diverse patient populations and imaging conditions.

3. Monte Carlo Dropout Refinement: A systematic evaluation of Monte Carlo Dropout across various FDA-approved models is essential. This research would quantify the method's effectiveness in real-world settings and refine it into a robust framework that AI developers can adopt. Establishing best practices for implementing Monte Carlo Dropout would contribute to standardizing uncertainty quantification techniques in healthcare AI.

4. Hybrid Uncertainty Estimation Techniques: Combining Monte Carlo Dropout with other approaches, such as Bayesian Neural Networks, could improve reliability and scalability in uncertainty quantification. This hybrid methodology might provide deeper insights into model behavior under uncertain conditions.

Conclusion

This framework establishes a scalable, non-intrusive approach to addressing model drift and uncertainty quantification in FDA-approved AI systems.

References

[1] A. S. M. C. G. K. D. Kelly, Christopher J.; Karthikesalingam, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, no. 1, p. 195, 2019. [Online]. Available: https://libkey.io/10.1186/s12916-019-1426-2

[2] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, Eds., *Dataset Shift in Machine Learning*, ser. Neural Information Processing series. Cambridge, MA, USA: The MIT Press, 2009.

[3] Z. Lowell, "Life science companies should work with fda to shape ai regulation." [Online]. Available: https://www.hoganlovells.com/en/publications/life-science-companies-should-work-with-fda-to-shape-ai-regulation

[4] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," 2016. [Online]. Available: https://arxiv.org/abs/1506.02142

[5] A. Dhinakaran, "How to understand and use the jensen-shannon divergence," Mar 2023. [Online]. Available: https://towardsdatascience.com/how-to-understand-and-use-jensen-shannon-divergence-b10e11b03fd6