Deep Learning for Drug Discovery: A Case Study in Kinase Inhibitors for Anti-Cancer Therapies

Submission for NeurIPS 2025 One-Day Event on AI Education

Target Audience

This educational material represents a pedagogical bridge between the life sciences and artificial intelligence communities, demonstrating how interdisciplinary collaboration can address significant challenges in human health. Our primary audience consists of deep learning practitioners in industry and academia that are interested in understanding how their computational methods relate to drug discovery applications, specifically the screening of molecules as potential kinase inhibitors for anti-cancer therapies. Conversely, healthcare professionals and pharmaceutical industry practitioners interested in understanding AI applications in their field will find this material accessible and relevant to their work. Our secondary audience consists of undergraduate students in computer sciences who may not have formal training in the life sciences. The content is also appropriate for educators who seek to incorporate applied AI concepts into their curricula.

Expected Time Commitment

The material is designed with flexible engagement levels. A quick overview of the main concepts and visualizations can be completed in under 10 minutes and is sufficient for fundamental understanding of AI in drug discovery. For those seeking deeper engagement, an interactive walkthrough with code execution takes no more than 20 minutes. Optional exercises are presented at the end for learners who wish to explore the content further.

Educational Material Description

This interactive notebook introduces deep learning concepts through a real-world application in pharmaceutical drug discovery. Instead of beginning with abstract mathematical foundations, learners immediately engage with a compelling medical challenge: identifying potential cancer treatments by predicting which molecules might effectively inhibit disease-causing proteins.

The notebook's scaffold facilitates learning progression from simple concepts, such as representing molecules as numerical data, to more complex topics including neural network architecture and training. We use pharmaceutical data from the ChEMBL database, which is employed by professional drug discovery researchers. The notebook includes additional exercises that allow learners to modify AI model parameters and observe how these changes affect drug discovery performance, reinforcing theoretical concepts through practical experimentation.

Upon completion, learners will have developed a concrete understanding of how AI accelerates scientific discovery, gained familiarity with the fundamental components and training processes of neural networks, and acquired the ability to interpret AI model performance metrics in practical contexts. They will also have cultivated an appreciation for both the capabilities and limitations of AI in critical applications, while gaining hands-on experience with authentic scientific data and modern AI tools.

To maximize accessibility, the notebook requires no specialized hardware and can run on any standard computer without GPU acceleration and minimal dependencies outside a few Python packages.