DEEP LEARNING FOR THE LIFE SCIENCES

By Bharath Ramsundar, Peter Eastman, Patrick Walters, and Vijay Pande

We are no longer in an era where data can be processed simply by loading it onto a spreadsheet and generating a few graphs. To distil scientific knowledge from datasets, we must be able to identify and extract nonobvious relationships.

One technique that has emerged in recent years as a powerful tool for identifying patterns and relationships in data is deep learning, a class of algorithms that has revolutionized approaches to tasks like image analysis, language translation, and speech recognition.

"Machine learning algorithms are now a key component of everything from online shopping to social media. Teams of computer scientists are developing algorithms that enable digital assistants such as the Amazon Echo or Google Home to understand speech," said the authors. "Advances in machine learning have enabled routine on-the-fly translation of webpages between spoken languages. In addition to machine learning's impact on everyday life, it has impacted many areas of the physical and life sciences," they add.

Deep learning algorithms excel at deciphering patterns in large datasets, which is why it has increasingly being applied to various life science disciplines including genetics, drug discovery, and medical diagnosis. The book provides examples of this and accompanies them with code illustrations, thus providing a practical introduction to the methods and giving readers a starting point for future research and exploration. The book will be of benefit to developers and scientists, helping them to apply deep learning methods to genomics, chemistry, biophysics, microscopy, medical analysis, and other fields.

Growing therapeutic uses

The authors explain the basics of performing machine learning on molecular data and they clarify how to use deep learning to analyse microscopic images and medical scans. "In recent years, life sciences and data science have converged. Advances in robotics and automation have enabled chemists and biologists to generate enormous amounts of data. Scientists today are
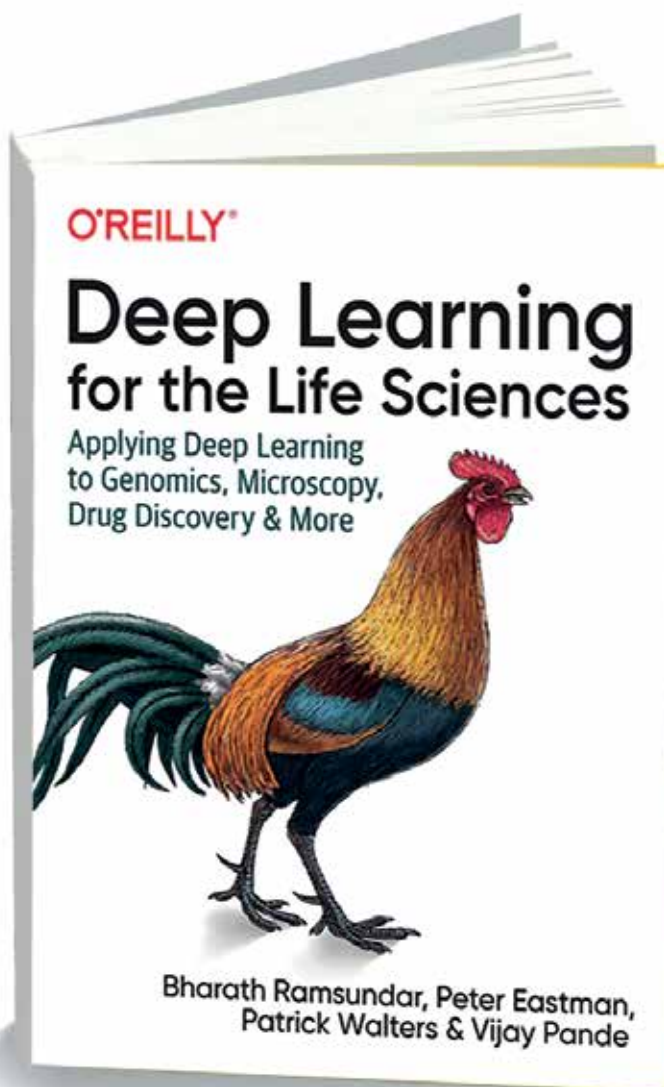
capable of generating more data in a day than their predecessors 20 years ago could have generated in an entire career," said the authors. "This ability to rapidly generate data has also created a number of new scientific challenges."

Technologies that are driven by deep learning include the speech recognition used in cell phones, computers, televisions, and other internet-connected devices, and the image recognition used in self-driving cars and in internet searches. Some deep learning methods that were first used in consumer applications are now being applied in biomedical research—for example, the classification of tumor cells into different types.

Companies like Amazon and Netflix apply deep learning techniques to entice customers to make more purchases with their 'customers who bought this also bought' approach. According to the authors, "many of the ideas behind these recommender systems are being used to identify new molecules that may provide starting points for drug discovery efforts."

"A powerful and rapidly advancing tool," the authors argue that "if you work in the life sciences, you need to be aware of it, because it's going to transform your field. Equally, if you work in deep learning, the life sciences are an incredibly important domain that deserves your attention."

The book comprises 12 chapters: Why Life Science?, Introduction to Deep Learning, Machine Learning with DeepChem, Machine Learning for Molecules, Biophysical Machine Learning, Deep Learning for Genomics, Machine Learning for Microscopy, Deep Learning for Medicine, Generative Models, Interpretation of Deep Models, A Virtual Screening Workflow Example, and Prospects and Perspectives.

Above:
Bharath Ramsundar
Peter Eastman
Patrick Walters.
Vijay Pande

Deep learning is a class of algorithms that has revolutionized approaches to tasks such as image analysis, language translation, and speech recognition