

# 深度学习与生物信息学

6 班 陈静雯

传统生物学数据分析方式已经不足以应对日益复杂庞大的生物序列数据，为此国内外学者将深度学习应用到生物学分析中，利用其处理高维数据的优势，取得一系列进展。

深度学习是学习样本数据的内在规律和表示层次，能够更好地解释文字、图像、声音等数据，最终目标是使机器达到像人一样具有分析和学习能力，能够识别外界信息，是一种机器学习算法。

蛋白质是生物研究的一个重要领域，随着生物技术的发展，其数据规模不断扩大，并且维度高、相关性较复杂<sup>[1]</sup>，而深度学习可以将复杂的数据转化为计算机可以学习的表示方式，并从中抽取潜在的高维数据特征。

2020 年，AlphaFold 在对蛋白质折叠的问题上取得了良好的效果<sup>[2]</sup>。蛋白质的复杂空间结构决定了它的功能特性，AlphaFold 通过利用 CNN，对氨基酸之间的距离以及其化学键的角度进行学习，给出了一个推荐结构，再用梯度下降算法调整优化整个结构预测结果，大大缩短了确定结构的时间<sup>[1]</sup>。

深度学习相对于传统的机器学习的优势在于<sup>[1]</sup>，它可以利用复杂的多层次神经网络，依靠前向与反向的传播，自行获得学习数据的隐藏特征，同时经过梯度下降等算法的优化，能够适应更多类型的、复杂的、体量庞大的数据，并且预测效果良好。利用深度学习方法进行生物分析和预测是未来生物信息技术重要发展方向之一。

## 参考文献:

[1]张冀东,王志晗,刘博.深度学习在生物序列分析领域的应用进展

[J].北京工业大学学报,2022,48(08):878-887.

[2]Senior Andrew W et al. Improved protein structure prediction using potentials from deep learning.[J]. Nature, 2020,

577(7792) : 706-710.