a *training set*, each of which is labelled as either malignant or benign, where the labels are obtained from a biopsy test that is considered to provide the true class of the lesion. The training set is used to determine the values of some 25 million adjustable parameters, known as weights, in a deep neural network. This process of setting the parameter values from data is known as *learning* or *training*. The goal is for the trained network to predict the correct label for a new lesion just from the image alone without needing the time-consuming step of taking a biopsy. This is an example of a *supervised learning* problem because, for each training example, the network is told the correct label. It is also an example of a *classification* problem because each input must be assigned to a discrete set of classes (benign or malignant in this case). Applications in which the output consists of one or more continuous variables are called *regression* problems. An example of a regression problem would be the prediction of the yield in a chemical manufacturing process in which the inputs consist of the temperature, the pressure, and the concentrations of reactants.

An interesting aspect of this application is that the number of labelled training images available, roughly 129,000, is considered relatively small, and so the deep neural network was first trained on a much larger data set of 1.28 million images of everyday objects (such as dogs, buildings, and mushrooms) and then *fine-tuned* on the data set of lesion images. This is an example of *transfer learning* in which the network learns the general properties of natural images from the large data set of everyday objects and is then specialized to the specific problem of lesion classification. Through the use of deep learning, the classification of skin lesion images has reached a level of accuracy that exceeds that of professional dermatologists (Brinker *et al.*, 2019).

1.1.2 Protein structure

Proteins are sometimes called the building blocks of living organisms. They are biological molecules that consist of one or more long chains of units called amino acids, of which there are 22 different types, and the protein is specified by the sequence of amino acids. Once a protein has been synthesized inside a living cell, it folds into a complex three-dimensional structure whose behaviour and interactions are strongly determined by its shape. Calculating this 3D structure, given the amino acid sequence, has been a fundamental open problem in biology for half a century that had seen relatively little progress until the advent of deep learning.

The 3D structure can be measured experimentally using techniques such as X-ray crystallography, cryogenic electron microscopy, or nuclear magnetic resonance spectroscopy. However, this can be extremely time-consuming and for some proteins can prove to be challenging, for example due to the difficulty of obtaining a pure sample or because the structure is dependent on the context. In contrast, the amino acid sequence of a protein can be determined experimentally at lower cost and higher throughput. Consequently, there is considerable interest in being able to predict the 3D structures of proteins directly from their amino acid sequences in order to better understand biological processes or for practical applications such as drug discovery. A deep learning model can be trained to take an amino acid sequence as input and generate the 3D structure as output, in which the training data