The BBBP dataset contains binary labels for 2,055 drug-like molecules, indicating whether they can penetrate the blood-brain barrier (BBB). The input features are the SMILES strings of the molecules, which are sequences of characters that represent the chemical structure.

LSTM and BiLSTM are both types of recurrent neural networks (RNNs) that can process sequential data, such as SMILES strings. LSTM stands for long short-term memory, and it is designed to overcome the problem of vanishing or exploding gradients that affect the learning of long-term dependencies in standard RNNs. LSTM uses a special structure called a memory cell, which consists of three gates: an input gate, an output gate, and a forget gate. These gates control how much information is stored, retrieved, or discarded from the memory cell.

BiLSTM stands for bidirectional LSTM, and it is an extension of LSTM that can capture both past and future contexts. BiLSTM consists of two LSTMs: one that processes the input sequence from left to right (forward), and one that processes the input sequence from right to left (backward). The outputs of the two LSTMs are then concatenated and fed to the next layer.

The difference in accuracy between LSTM and BiLSTM can be explained by the fact that BiLSTM can utilize more information from the input sequence than LSTM. For example, consider the following SMILES string:

CC(C)(C)OC(=O)N1CCC(CC1)C(=O)N2CCN(CC2)C(=O)C3=CC=CC=C3

This molecule has a benzene ring at the end, which is a common substructure in organic chemistry. LSTM can only learn this information from the left to right direction, while BiLSTM can also learn it from the right to left direction. This can help BiLSTM to recognize patterns and features that are relevant for predicting the BBB permeability of the molecule.

Therefore, BiLSTM can achieve higher accuracy than LSTM by exploiting the bidirectional nature of the input sequence. However, BiLSTM also has some drawbacks, such as increased computational complexity and potential overfitting. Therefore, the choice of the network architecture should depend on the characteristics of the data and the task.