Challenge

- Can the two languages be distinguished using a bag-of-words approach? Explain why.

  No. Bag-of-words doesn't relate to the order of the characters, because a sum is commutative. For example, the bag-of-words will output the same result on the two words 11a11b11c11d11, 11a11c11b11d11, because the sum of their embeddings (no matter what the embeddings are) is equal, while those words have different labels.

- Can the two languages be distinguished using a bigram or trigram based approach? Explain why.

  No. trigram, and bigram preserve some local order, but not global. It means that two for example when a trigram will see '1b1' it doesn't consider the position of this sequence in the entire word, which means it can be on the start or on the end. This can cause the model to misidentify.

- Can the two languages be distinguished using a convolutional neural network? Explain why.

  Yes. ConvNets preserves local and global order, which means that a ConvNet can identify if 'b' appears before 'c' in the entire word.