Summary

- You can use convolutional networks on top of characters (called learning from scratch)

- It works best for large datasets where it beats classical approaches (like BOW)

- Sometimes it even beats LSTM that works on word level