# Bangla Document Classification using Character Level Deep Learning

[Md. Mahbubur Rahman](https://ieeexplore.ieee.org/author/37088557445); [Rifat Sadik](https://ieeexplore.ieee.org/author/37088558318); [Al Amin Biswas](https://ieeexplore.ieee.org/author/37087139260)

**Abstract:**

Last few decades, the availability and accessibility of the Bangla document and its content have rapidly increased due to the rapid technological advancement. Intense research needs to be performed on various Bangla documents due to the diversity of the language and associated sentiment. Document classification is one of the fundamental problems of Natural Language Processing. To handle miss-classification and convenient indexing and searching of Bangla documents on the web, researchers nowadays exploring different fields of computer science to classify Bangla documents. In this paper, Deep Learning based approaches are implemented to classify Bangla text documents. Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) is used here for the classification task. Here we have implemented an advanced technique that encoded the documents at their character level. Documents from three different data sources are used to validate and test of the working models. The highest classification accuracy is 95.42% that is achieved on the Prothom Alo data set using LSTM. Furthermore, we presented a comparison between two models and explained how well the classification task can be carried out using our character level approach with higher accuracy.

|  |
| --- |
| **Conference / Journal Link:**  <https://ieeexplore.ieee.org/document/9254416> |

**DOI:** [10.1109/ISMSIT50672.2020.9254416](https://doi.org/10.1109/ISMSIT50672.2020.9254416)