|  |  |  |  |
| --- | --- | --- | --- |
| Title | BanglaWriting: A multi-purpose offline Bangla handwriting dataset | | |
| Author(s) Name | Muhammad F Mridha, Abu Quwsar Ohi, M Ameer Ali, Mazedul Islam Emon, Muhammad Mohsin Kabir | | |
| Contact Email(s) | firoz.mridha@aiub.edu | | |
| Published Journal Name | Data in Brief | | |
| Type of Publication | Journal | | |
| Volume | 34 | Issue |  |
| Publisher | Elsevier | | |
| Publication Date | 2021/2/1 | | |
| ISSN | 2352-3409 | | |
| DOI | 10.1016/j.dib.2020.106633 | | |
| URL | https://www.sciencedirect.com/science/article/pii/S2352340920315134 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| This article presents a Bangla handwriting dataset named BanglaWriting that contains single-page handwritings of 260 individuals of different personalities and ages. Each page includes bounding-boxes that bounds each word, along with the unicode representation of the writing. This dataset contains 21,234 words and 32,787 characters in total. Moreover, this dataset includes 5,470 unique words of Bangla vocabulary. Apart from the usual words, the dataset comprises 261 comprehensible overwriting and 450 handwritten strikes and mistakes. All of the bounding-boxes and word labels are manually-generated. The dataset can be used for complex optical character/word recognition, writer identification, handwritten word segmentation, and word generation. Furthermore, this dataset is suitable for extracting age-based and gender-based variation of handwriting. | |