

A Benchmark Dataset for Manipuri Meetei-Mayek Handwritten Character Recognition

Contributor(s): Pangambam Sendash Singh¹

**Department of Computer Science,
Institute of Science,
Banaras Hindu University, Varanasi, India**

A benchmark dataset is always required for any classification or recognition system. To the best of our knowledge, no benchmark dataset exists for handwritten character recognition of Manipuri Meetei-Mayek script in public domain so far.

Manipuri, also referred to as Meiteilon or sometimes Meiteilon, is a Sino-Tibetan language and also one of the Eight Scheduled languages of Indian Constitution. It is the official language and lingua franca of the southeastern Himalayan state of Manipur, in northeastern India. This language is also used by a significant number of people as their communicating language over the north-east India, and some parts of Bangladesh and Myanmar. It is the most widely spoken language in Northeast India after Bengali and Assamese languages.

In this work, we introduce a handwritten Manipuri Meetei-Mayek character dataset which consists of more than 5000 data samples which were collected from a diverse population group that belongs to different age groups (from 4 years to 60 years), genders, educational backgrounds, occupations, communities from three different districts of Manipur, India (Imphal East District, Thoubal District and Kangpokpi District) during March and April 2019. Each individual was asked to write down all the Manipuri characters on one A4-size paper. The recorded responses are scanned with the help of a scanner and then each character is manually segmented from the scanned images.

This dataset consists of segmented scanned images of handwritten Manipuri Meetei-Mayek characters (Mapi Mayek, Lonsum Mayek, Cheitap Mayek, Cheising Mayek, Khutam Mayek) of size 128X128 pixels in .JPG format as well as in .MAT format.

Cite this dataset as: Pangambam Singh, "A Benchmark Dataset for Manipuri Meetei-Mayek Handwritten Character Recognition", IEEE Dataport, 2019. [Online].

Available: <http://dx.doi.org/10.21227/fwax-yr43>.

This dataset is licensed under a Creative Commons Attribution 4.0 International licence. What does this mean? You can share, copy and modify this dataset so long as you give appropriate credit, provide a link to the CC BY license, and indicate if changes were made, but you may not do so in a way that suggests the rights holder has endorsed you or your use of the dataset. Note that further permission may be required for any content within the dataset that is identified as belonging to a third party.

¹ pangambams.singh4@bhu.ac.in