

1) SinOCR + SinFUND (SriDoc)

- **SinOCR**: printed + handwritten Sinhala text images (the paper says it's large-scale).
- **SinFUND**: scanned, handwritten Sinhala forms (for form understanding / document structure).

<https://drive.google.com/drive/folders/1Czf8dQkEzWXBBcmVbBNa2OaMwdBBvBsx>

<https://github.com/SriDoc/datasets>

2) “Sinhala Letter and Modifications” (Kaggle) — handwritten characters (454 classes)

- Handwritten Sinhala characters, ~45k images, 454 classes.

<https://www.kaggle.com/datasets/sathiralamal/sinhala-letter-454>