



S. M. RAKIB HASAN <sm.rakib.hasan@g.bracu.ac.bd>

## 2023 IEEE Fifth International Conference on Advances in Electronics, Computers and Communications (ICAIECC) : Submission (31) has been created.

2 messages

Microsoft CMT <email@msr-cmt.org>

Mon, May 1, 2023 at 1:35 PM

Reply-To: Microsoft CMT - Do Not Reply <noreply@msr-cmt.org>

To: sm.rakib.hasan@g.bracu.ac.bd

Hello,

The following submission has been created.

Track Name: ICAIECC2023

Paper ID: 31

Paper Title: Optical Text Recognition in Nepali and Bengali: A Transformer-based Approach

### Abstract:

Efforts on the research and development of OCR systems for Low-Resource Languages are relatively new. Low-Resource Languages have little training data available for training Machine Translation systems or other systems. Even though a vast amount of text has been digitized and made available on the internet the text is still in PDF and Image format, which are not instantly accessible. This paper discusses text recognition for two scripts: Bengali and Nepali; there are about 300 and 40 million Bengali and Nepali speakers respectively. In our study, using encoder-decoder transformers, a model was developed, and its efficacy was assessed using a collection of optical text images, both handwritten and printed. The results signify that our suggested technique corresponds with current approaches and achieves high precision in recognizing text in Bengali and Nepali. This study can pave the way for the advanced and accessible study of linguistics in South East Asia.

Created on: Mon, 01 May 2023 07:35:13 GMT

Last Modified: Mon, 01 May 2023 07:35:13 GMT

### Authors:

- [sm.rakib.hasan@g.bracu.ac.bd](mailto:sm.rakib.hasan@g.bracu.ac.bd) (Primary)
- [aakar.dhakal@g.bracu.ac.bd](mailto:aakar.dhakal@g.bracu.ac.bd)

Secondary Subject Areas: Not Entered

Submission Files: OCR in Bengali and Nepali Using Transformers.pdf (351 Kb, Mon, 01 May 2023 07:34:06 GMT)

### Submission Questions Response:

1. Contact number

+8801907791892

2. Contact number

+8801322821560

3. Keywords of the paper

Low-Resource Languages, OCR, Transformers, Machine Translation, Digitized text

4. Domain

Information Technology

Thanks,  
CMT team.

Download the CMT app to access submissions and reviews on the move and receive notifications:

<https://apps.apple.com/us/app/conference-management-toolkit/id1532488001>

<https://play.google.com/store/apps/details?id=com.microsoft.research.cmt>

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our [Privacy Statement](#).

Microsoft Corporation  
One [Microsoft Way](#)  
[Redmond, WA 98052](#)

---

**S. M. RAKIB HASAN** <sm.rakib.hasan@g.bracu.ac.bd>  
To: Annajiat Alim Rasel <annajiat@gmail.com>

Mon, May 1, 2023 at 1:52 PM

Assalamu Alaikum Sir,

We have submitted our paper to the 2023 IEEE Fifth International Conference on Advances in Electronics, Computers and Communications (ICAEECC). Here is the receipt email.

Regards,  
Team No-04  
22241038, S M Rakib Hasan  
20201203, Aakar Dhakal  
[Quoted text hidden]



Virus-free.[www.avast.com](http://www.avast.com)