

Bengali.AI Computer Vision Challenge 2019

Sponsorship Booklet

Bengali.AI Computer Vision Challenge 2019

Featured on: Kaggle



National and International Researchers
2.5 Month Long Competition
3rd Biggest Language Dataset
Revolutionizing Bangla language research using AI
Endless possibilities



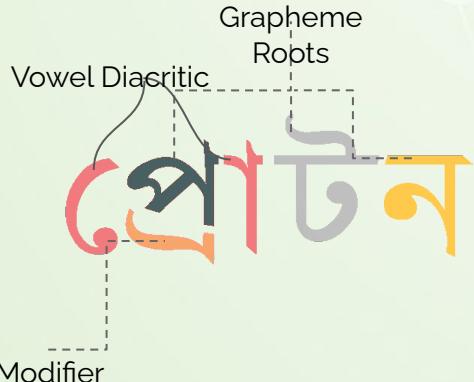
Grapheme recognition is a stepping stone towards
Bengali Optical Character Recognition, which
can digitize any form of written document, i.e. class
lectures, prescriptions, ledgers, etc

Bengali.AI presents Biggest Bangla Handwritten Grapheme dataset: (Anushilon??)

494,000
Samples Collected

1296
Unique Graphemes

2894
Contributors



The image displays a survey form at the top and a grid of handwritten Bengali characters below it. The survey form includes fields for Name, Roll ID, Age (0-12, 13-17, 18-24, 25-35, 36-50, 50+), Gender (Male, Female, Other), Dominant Hand (Right, Left), Medium of Primary Edu (Bangla, English, Madrasa), and Division of Primary Edu (DHA, CTG, SYL, RAJ, RAN, COM, BAR, KHU). The grid of characters shows various handwritten samples of Bengali letters and words, such as 'বা', 'ক্ত', 'মে', 'ষ', 'অ', 'ফি', 'সি', 'চ', 'তা', 'ধা', 'ৰা', 'দি', 'দী', 'ক', 'নি', 'মা', 'পি', 'ক্ষ', 'ক্ষ', 'ক্ষ', 'ম', 'মেটা', 'ক্ষ', 'মো', 'মী', 'গে', 'জে', 'খ', 'ঝ', 'ঝ', 'কো', 'শি', 'জা', 'শী', 'হী', 'প', 'ভি', 'ঢ', 'ঢা', 'ঢ়া', 'ঢ়া', 'ঢ়ে', 'শী', 'হু', 'দৌ', 'গ্রা', 'গ্রি', 'দে', 'দো', 'র', 'র', 'ক', 'পে', 'ব'.

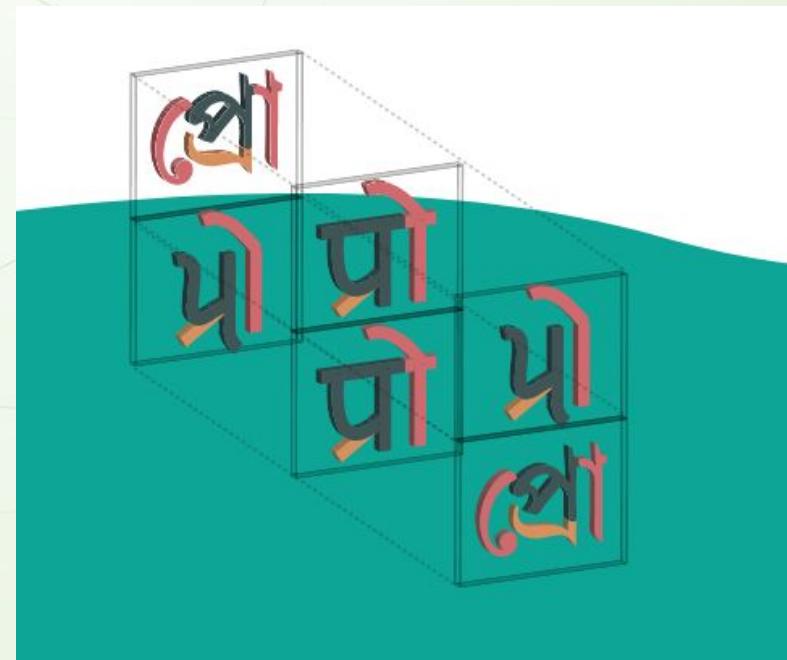
By contributing your handwriting you are consenting to be a part of a research being conducted by Bengali.AI (www.bengali.ai). The data will be anonymized and open-sourced for public use. Bengali.AI is a community initiative aiming to democratize AI research in Bangladesh.

Biggest Abugida* Grapheme Dataset

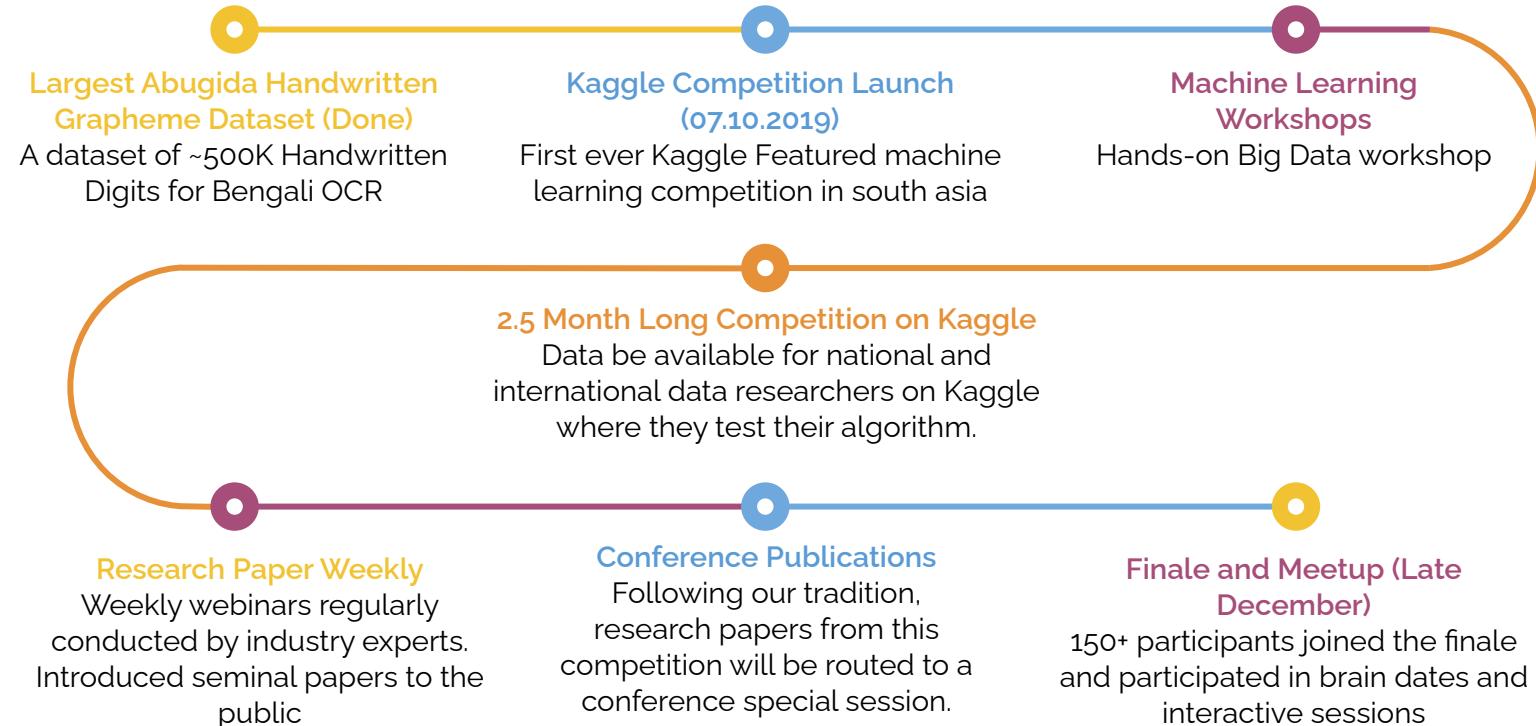
An **Abugida** is a [writing system](#) which is between [syllabic](#) and [alphabetic](#) scripts. In such systems, sequences of [consonants](#) and [vowels](#) are written as a unit: each unit is based on the consonant letter.

Examples of Abugida language family is Bangla, Devanagari, Tibetan, Tamil, Gujarati, Thai, Javanese, etc.

The research conducted on this Bengali Handwritten Grapheme dataset will directly contribute towards standardizing handwritten OCR research for all these abugida languages due to the structural similarity.



Timeline for this competition



WHY SPONSOR THIS COMPETITION?

- >Bangla Handwritten OCR will drastically change the current scenario in medical, monetary and legal sector
- >There is no open source dataset and so researchers are not able to make progress
- >An Open source dataset and a competition will pave the way towards digitizing handwritten texts from any domain.
- >This dataset is the stepping stone towards that



WHY SPONSOR THIS COMPETITION?

- >This competition will facilitate 50,000+ hours of independent research on Bangla OCR
 - >Comparative analysis of the performance of different AI techniques for Abugida OCR
 - >AI is mandatory for scaling any business.
 - >Global giants like Facebook and Google are using AI for their products
- *Insert google camera Illustration

>Reach the largest and most diverse global data community with over a million registered users or Kagglers.

WHY SPONSOR THIS COMPETITION?

AI trailblazers, machine learning experts, forward thinking executives, data scientists, and product and engineering innovators will gather at the grand **finale** of our competition to share how to build **real-world AI solutions** as well as discuss the impact of AI in our society and life.

BE ESTABLISHED AS AN AI LEADER

Be recognized in front of an audience of top AI industry experts, startups, developers, corporations, and investors. Dedicated space will allow companies to display products/ services or recruit.



GET CONNECTED

Connect and network with our community. Scout innovation, hire new talent, create strategic partnerships, and learn about the latest technologies and industry trends.



CONTRIBUTE TO A NATIONAL GOAL

Support the AI community and the general population at large with knowledge, capital, and network.



WHAT IS BENGALI.AI?



Bengali.AI is a community founded by a group of academics and student researchers of Bangladesh with the goal of **democratizing Artificial Intelligence research** in the country

We focus on the three pillars of a digitalized Bangladesh- teaching **computers to read, speak and comprehend Bangla.**





WHAT BENGALI.AI Does



We create standardized datasets and open source them to enthusiastic **researchers** from all over the world.

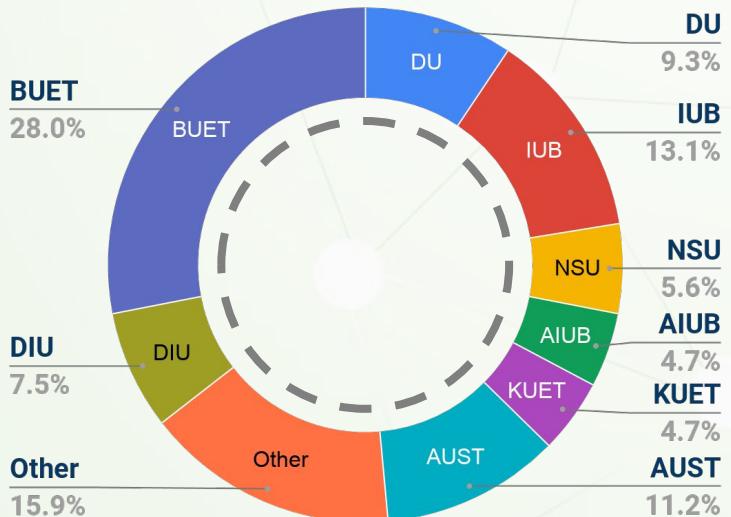
We arrange competitions to facilitate research on Bangla language



ATTENDEE SNAPSHOT

From Last Year

Attendees by Affiliation

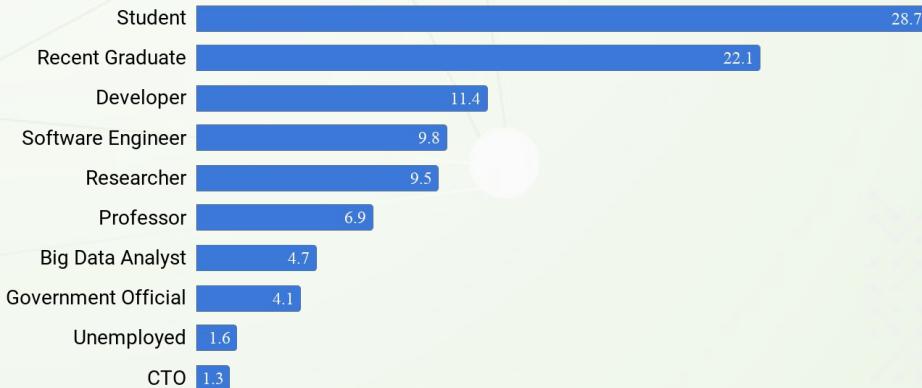


ML Practitioner Snapshot

- Amateur
- Intermediate
- Professional
- Expert



Attendees by Occupation



Frequently Asked Questions



What is Kaggle?

[Kaggle](#) is the Mecca for Machine Learning practitioners worldwide with more than 2 million registered members.

What is a Kaggle featured competition?

While competitions with limited scope can be launched on Kaggle for free (last year's [competition](#)), featured competitions are promoted worldwide by Kaggle themselves. While featured competitions require a hefty fee of **120K USD**, they are doing it for free for **this competition** considering the impact of our research.

Why are you open-sourcing and giving it out for free?

Why should I fund you, doesn't this also help my market competitors?

Apart from the altruistic concept, open-source makes sure that a lot of people are working on to find solutions to the same problem regardless of it being monetized or not. This parallelizes the process of finding out **what works and what doesn't**. Therefore one participant (regardless of their intent towards commercialization) can know what tools others are using to solve the problem and build on top of that. Open-sourced datasets and competitions (i.e. Imagenet) stand at the core of the Machine Learning tools we use today and have fueled the AI revolution so far. We are working towards revolutionizing AI research on Bangla language.

Why don't you develop such solutions and monetize them yourselves?

The team leading Bengali.AI are full-time researchers, with very little intent towards monetization of such algorithms. This is why we open-source the datasets to invite participants both from industry (with commercial intent) and academia (research intent). Only through commercialization can these services impact society. Therefore we open-source this for everyone, big and small, to develop/monetize/commercialize and bring societal impact.

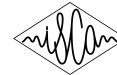


Who has Joined Us in the Past

Hosted On

kaggle

Sponsors and Supporters



Collaborators





JOIN THE ROAD TO THE FUTURE

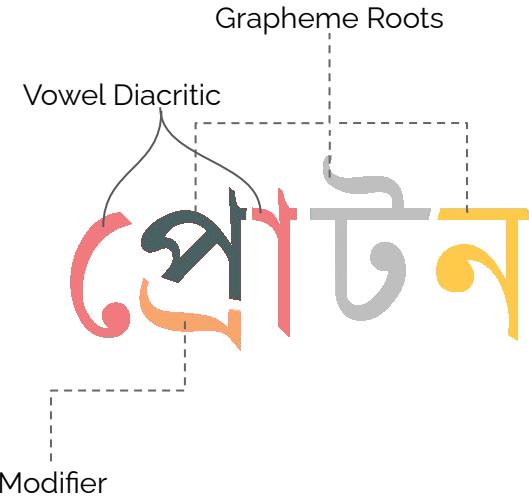
**Contact
us:**

cv19@bengali.ai

Inside the dataset: Anushilon

English Graphemes : Simple, comprised of alphabets

Bengali Graphemes: Complex, comprised of conjuncts and diacritics



In Bangla, every word is a combination of the basic grapheme forms.

Every grapheme has three attributes: i) Grapheme Root, ii) Vowel Diacritic & iii) Modifier