

Masum Hasan

Research Assistant, Applied Machine Learning Lab, BUET
masum@ra.cse.buet.ac.bd | +8801684080299 | Homepage: masumhasan.net
Linked in: masum6 | GitHub: Masum06 | Skype: youngladesh@live.com

Research Interest

Natural Language Processing, Unsupervised and Low Resource NLP, Machine Learning for Source Code: Summarization/Synthesis/Search/Repair, Machine Learning, Deep Learning, Software Engineering, Human Computer Interaction, Artificial Intelligence.

Achievements

Best Student Paper Award (2018):

- ✦ International Conference on Bangla Speech and Language Processing – 2018
- ✦ Paper: “Recognition of Bengali Handwritten Digits Using Convolutional Neural Network Architectures”

Kaggle Contest Winner (2018):

- ✦ Bengali Handwritten Digit Recognition Competition
- ✦ <https://www.kaggle.com/c/numta/leaderboard>
- ✦ 99.3559% accuracy in NumtaDB, the largest dataset of Bengali Handwritten Digits

2nd Runner up at Robi Datathon (2019):

- ✦ First Data Science Hackathon in Bangladesh

NSysS Software Project Award (2016):

- ✦ <http://floaterfb.blogspot.com/>

Accepted Papers

CoDesc: A Large Code–Description Parallel Dataset. **Masum Hasan**, Tanveer Muttaqeen, Ishtiaq Niloy, Kazi Mehrab, Tahmid Hasan, Mahim Pantho, Wasi Uddin Ahmad, Rifat Shahriyar, Anindya Iqbal
[ACL Findings 2021](#).

Not Low-Resource Anymore: Aligner Ensembling, Batch Filtering, and New Datasets for Bengali-English Machine Translation. Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, **Masum Hasan**, Madhusudan Basak, M. Sohel Rahman, Rifat Shahriyar. EMNLP, 2020.
arXiv preprint: <https://arxiv.org/abs/2009.09359>

Recognition of Bengali Handwritten Digits Using Convolutional Neural Network Architectures. **Md Mahmudul Hasan**, Md Rafid Ul Islam, Md Tareq Mahmood. ICBSLP, Sylhet, Bangladesh, 2018.
DOI: 10.1109/ICBSLP.2018.8554753

Early Detection of Earthquake Using Satellite Based Quantum Computing. Akhter Al Amin, **Mahmudul Hasan**, Kazi Sinthia Kabir, Tanzila Choudhury, ABM Alim Al Islam. ICCSNT, 2015.
DOI: 10.1109/ICCSNT.2015.7490923

Completed Papers

Text2App: A Framework for Creating Android Apps from Text Descriptions. **Masum Hasan**, Kazi Sajeed Mehrab, Wasi Uddin Ahmad, Rifat Shahriyar. Submitted: [ACL System Demonstration Track 2021](#).
arXiv preprint + Code + Demo: <https://text2app.github.io>

Review4Repair: Code Review Aided Automatic Program Repairing. Faria Huq, **Masum Hasan**, Mahim Pantho, Sazan Mahbub, Anindya Iqbal, Toufique Ahmed. Rebuttal: [IST \(Journal\)](#), Impact Factor: 2.726
arXiv preprint: <https://arxiv.org/abs/2010.01544>

Using a Balanced Scorecard to Identify Opportunities to Improve Code Review Effectiveness: An Industrial Experience Report **Masum Hasan**, Anindya Iqbal, Amiangshu Bosu, Mohammad Rafid Ul Islam, A.J.M. Intiajur Rahman. Under review: [EMSE \(Journal\)](#)

BERT2Code: Can Pretrained Language Models be Leveraged for Code Search? Istiaq Niloy, **Masum Hasan**, Kazi Mehrab, Mahim Pantho, Tahmid Hasan, Rifat Shahriyar
Submitted: ICANN 2021

On-going Research

Sub-Tree Compression: A Novel Compression Algorithm for Source Code Data**

- ⊕ Compress repetitive code structures to reduce source code data length.
- ⊕ Lossless reduction of number of code tokens by 10~20%.
- ⊕ 10% faster training for state-of-the-art code summarization network with no performance reduction

GPT-Bengali**

- ⊕ Crawled and pre-processed 18GB Bengali Corpus
- ⊕ Trained a GPT2 small Language Model (110M parameters) in Bengali
- ⊕ New benchmark, datasets, evaluation metrics for Bengali Language models

What You See Is What You Learn: Mitigating Gender Bias in Natural Language Data by Encrypted Gender Training** [draft]

- ⊕ Gender debiasing LM's in 4 steps: Gender Encryption, Training, Generation, Gender Decryption
- ⊕ Technologies: NER, Char-LSTM, Coreference Resolution, GPT2 Language Model
- ⊕ Supervised by: [Dr. Md. Monirul Islam](#), CSE, BUET & [Nazmus Saquib](#), MIT Media Lab.

NMT for Endangered Indigenous Languages in Bangladesh*

- ⊕ Voluntary research with non-profit organization "[Beyond The Hills](#)" to break the language barrier between mainstream Bangladeshi people and indigenous minorities in Bangladesh.
- ⊕ Building NMT models for Chakma Language that has parallel in few thousands.

* Project Lead ‡ Research Initiator

Projects

[GenderViz: Visualize Gender Interactions from Text \(2019\)](#)

- ⊕ Generating **Semantic Graph** of gender interaction from any given text.
- ⊕ Made with Character-LSTM gender classifier with 3 million Bangladeshi names
- ⊕ Named Entity Recognizer, Information Extraction, Semantic Triples

[Game of Thrones Nudity Filter \(2018\)](#)

- ⊕ Created nudity filter with **Tensorflow Object Detection API** so that my niece could watch GoT.
- ⊕ Trained with 1600 boundary boxed images for identifying 5 types of obscene parts.

[Floater fb \(2016\)](#)

- ⊕ A small beautiful desktop UI to use Facebook that helps users to focus.
- ⊕ L-2,T-1 Sessional Term Project. 700+ downloads. Received NSysS Software Project Award – 2016

[Project VEX: Virtual Existence \(2016\)](#)

- ⊕ You can remotely see and control a paintbrush with your hand gesture. At KolpoKoushol.

Professional Experience

Research Assistant (RA), Applied Machine Learning Lab, BUET

Nov 2018 to Present

- ⊕ Supervised by, [Dr. Rifat Shahriyar](#), [Dr. Anindya Iqbal](#)
- ⊕ ML for source code: Source Code Summarization/Synthesis/Search/Repair/Compression
- ⊕ NLP for Bengali: GPT2, BERT, Bengali-English NMT

Deep Learning Research Intern, Gaze Technologies

Jun 2018 to Aug 2018

- ⊕ Object Detection, License Plate Recognition
- ⊕ Tensorflow Object Detection API

R&D Intern, REVE Systems

Mar 2018 to Jun 2018

- ⊕ Optical Character Recognition
- ⊕ Industry Standard Coding Practices

Cofounder and CEO, Bizzy Ltd.

Dec 2016 to Jul 2017

- ⊕ [BizzyBd](#) started as an easy-to-use website building tool for non-programmer professionals.

Education

Incoming PhD in Computer Science (CS)

University of Rochester (NY)

- ⊕ Starting from Fall (August) 2021
- ⊕ Human Computer Interaction (HCI)

BSc. In Computer Science and Engineering (CSE)

Bangladesh University of Engineering and Technology (BUET) ~ Oct 2018

- ⊕ Thesis: Mythbusting Gender Stereotypes Using Natural Language Processing

Standardized Tests

Graduate Record Examinations (GRE)

- ⊕ Analytical Writing: 4.0, Verbal Reasoning: **154**, Quantitative Reasoning: **166**. **Total: 320**

Test of English as a Foreign Language (TOEFL)

- ⊕ Reading: **28**, Listening: **29**, Speaking: **24**, Writing: **24**. **Total: 105**

Technical Skills

Programming Languages: Python, Java, C, C++, JavaScript, MATLAB, HTML, SQL, MySQL, LaTeX, Arduino, AVR Programming.

Library/Framework/Others: Numpy, **Keras**, **PyTorch**, Tensorflow Object Detection API, Django, **OpenNMT**, **HuggingFace**, JQuery, Google SyntaxNet, Bootstrap.

Online Courses

- ⊕ [Machine Learning](#), Coursera.
- ⊕ [Deep Learning Specialization](#), Deeplearning.ai
- ⊕ [Practical Deep Learning for Coders](#), Fast.ai
- ⊕ [Book] [Neural Network and Deep Learning](#), Michael A. Nielsen
- ⊕ [Partial] [CS224n: Natural Language Processing with Deep Learning](#), Stanford University
- ⊕ [Essence of Linear Algebra](#), 3blue1brown

Blogs and Articles

Machine Learning:

- [Absolute Beginner's Guide to Machine Learning and Deep Learning](#)
- [Setup and Run fast.ai in Amazon AWS](#)

Graduate Admission:

- [How I got 4 Ph.D. offers in the US with a CGPA 2.79 — and what you can learn from it](#)

Extracurricular Activities

- ⊕ [I preach AI on YouTube!](#), also [physically](#)
- ⊕ Speaker at Bengali.AI Research Paper Weekly Webinar
- ⊕ Volunteer at InclusionX at EMK Center, Dhaka
- ⊕ Taekwondo at BUET Self Defence Club
- ⊕ Volunteer, EducationUSA at American Center, Dhaka (2013-2015)
- ⊕ Participant, [KolpoKoushol Workshop](#) (supervised by. Nazmus Saquib, MIT Media Lab), 2015, 2016
- ⊕ Former Vice President, ESKAP (Ex-students association of Kalia Pilot high school) 2013-2015