

# Md. Shahidul Salim

📍 CSE, KUET, Khulna, Bangladesh

✉ ss@cse.kuet.ac.bd 🐙 Github 🌐 Huggingface 🌐 Portfolio 🌐 LinkedIn

## Job Experience

July 2022 – Present 📌 **Faculty member**, Department of CSE, Khulna University of Engineering & Technology (KUET), Bangladesh

April 2021 – March 2022 📌 **Faculty member**, Department of CSE, Uttara University, Bangladesh

## Research Interest

- 📌 Natural language processing - Stemmer, Transformer models (Translation, Summarization, Causal language model, Question answering), Fine-tuning LLMs (Mistral, Llama-2, Zephyr)
- 📌 Machine learning, Deep learning, Multivariate and univariate time series

## Education

2016 – 2021 📌 **B.Sc. in Computer Science and Engineering**, Khulna University of Engineering & Technology (KUET), Bangladesh  
– CGPA 3.86 out of 4.00 (4th position in my department)

## Research Publications

### Journal Articles

- 1 Ashiqussalehin, M., Jahan, K. N., Rahaman, M. A., & Salim, M. S. (2022). Human abnormal behavior detection using convolution neural network. *Specialusis Ugdymas*, 1(43), 4076–4083.

### Conference Proceedings

- 1 Hossain, L., Hossain, I., Salim, M. S., Raju, S. M. T. U., & Saha, J. (2023). A novel technique for classification of motor imagery eeg signal based on deep learning approaches. In *Proceedings of the 2nd international conference on big data, iot and machine learning (bim 2023)*. (Accepted).
- 2 Nabil, A., Das, d., Salim, M. S., Arifeen, S., & Fattah, H. M. A. (2023). Bangla emergency post classification on social media using transformer based bert models. In *6th international conference on electrical information and communication technology (eict 2023)*. (Accepted).
- 3 Promi, R. T. H., Nazri, R. A., Salim, M. S., & Raju, S. M. T. U. (2023). A deep learning approach for non-invasive hypertension classification from ppg signal. In *2023 international conference on next-generation computing, iot and machine learning (ncim)* (pp. 1–5). doi:10.1109/NCIM59001.2023.10212940
- 4 Salim, M. S., Murad, H., Das, D., & Ahmed, F. (2023). Banglagpt: A generative pretrained transformer-based model for bangla language. In *2023 international conference on information and communication technology for sustainable development (icict4sd)* (pp. 56–59). doi:10.1109/ICICT4SD59951.2023.10303383
- 5 Salim, S., Islam, T., Zannat, R., Mia, N., Fuad, M., & Murad, H. (2023). Towards developing a transformer-based bangla typing error correction model: A deep learning-based approach. In *2023 international conference on information and communication technology for sustainable development (icict4sd)* (pp. 75–78). doi:10.1109/ICICT4SD59951.2023.10303361
- 6 Ahmed, T., Hossain, S., Salim, M. S., Anjum, A., & Azharul Hasan, K. M. (2021). Gold dataset for the evaluation of bangla stemmer. In *2021 5th international conference on electrical information and communication technology (eict)* (pp. 1–6). doi:10.1109/EICT54103.2021.9733662
- 7 Salim Shakib, M. S., Ahmed, T., & Azharul Hasan, K. M. (2019). Designing a bangla stemmer using rule based approach. In *2019 international conference on bangla speech and language processing (icbslp)* (pp. 1–4). doi:10.1109/ICBSLP47725.2019.201533

## Under Review and Ongoing Research

### Under Review

- **An Applied Statistics Dataset for Human vs AI-Generated Answer Classification (Data in Brief)**
  - We have developed a dataset of 4231 question-and-answer pairs for Applied Statistics, which includes 116 questions selected by domain experts and answers provided by 100 students and 50 AI-generated responses using ChatGPT. Our dataset can be used to develop AI detectors for the Applied Statistics domain and to evaluate existing ones. Additionally, our data collection framework can be applied to other domains.
- **Agricultural Recommendation System based on Multivariate Weather Forecasting Model(Engineering Applications of Artificial Intelligence journal)**
  - This paper proposes a context-based crop recommendation system using a weather forecast model to improve farming practices in Bangladesh. The multivariate Stacked Bi-LSTM Network is used for accurate weather prediction, including rainfall, temperature, humidity, and sunshine. The system guides farmers in making informed decisions about planting, irrigation, harvesting, and more. It also alerts farmers about extreme weather conditions and provides knowledge-based crop recommendations for flood and drought-prone areas.
- **Detecting AI-Generated Assignments in Educational Evaluation: A Transformer-Based Approach**
  - This research work presents a transformer-based model to detect whether an assignment is AI-generated or human-written. The model was trained on a dataset of 5410 assignments, with 2742 being AI-generated and 2668 being human-written. Among the explored transformer-based architectures, DistilBERT provided the highest accuracy of 92%.
- **A Suffix Independent Algorithm for Stemming Bangla Words using Finite State Transducer (Expert Systems With Applications)**
  - This study proposes and evaluates a suffix-independent stemming algorithm for Bangla language using a finite state transducer (FST)-based framework. The algorithm creates a dictionary of root words implemented as an FST, achieving high speed and no memory usage for vocabulary keeping. A novel stemmer dataset was developed to evaluate the algorithm's performance, resulting in 96.58% detection accuracy and 96.34% stemming accuracy. The proposed scheme outperforms existing methods and demonstrates effectiveness through experiments.

### Ongoing Research

- Bangla context-based question answering with custom dataset using transformer model
- Automatic Text Summarization in Bengali using mT5 Transformer Model

## Awards and Projects

- Dean's Award by Faculty of Electrical & Electronic Engineering

### Projects

- **Medical LLM Chatbot** - Chat with pdf using LLM(LlamaV2) langchain and streamlit
- **KUET Chat Bot** - Information about KUET
  - Students can chat with the bot and get information about KUET.
- **Efficient Backlog Routine Generator** - Python and Flask
- **Anonymity-Preserving Post Web Application** - Implementing Python and Flask for Secure and Confidential Content Sharing
- **Statistics exam** - Design and Implementation of a Statistics Exam Generation System for Students using Python and Flask with Randomized Data Generation
- **Counterfeit note detection** - Fake Bangladeshi Banknote Detection using Convolutional Neural Networks (CNN)
- **Brain Tumor Detection using Convolutional Neural Networks (CNN)** - An AI-based Approach for Accurate Diagnosis

## Miscellaneous Experiences

- 2023
- Ibex(Supercomputer) fine-tune LlamaV2 for medical data
  - Research paper review (Worked as a reviewer for the conference AAAI-2024).

## Technical Skills

- Programming Languages - Python, HTML,CSS, C, C++,  $\text{\LaTeX}$  Java, Javascript
- Frameworks - Pytorch, Tensorflow, Scikit-learn, HuggingFace Transformer, Keras, Langchain, Streamlit, Gradio, Flask