Md. Shahidul Salim

Lowell, MA, USA

mdshahidul_salim@student.uml.edu Github Hugging Face Portfolio Linkedin Google Scholar

Job Experience

January 2024 – Now — **Research assistant,** University of Massachusetts Lowell, USA

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
  - Large language models
  - Transformer models (Translation, Summarization, Text generation, Conversational Question answering)
- Machine learning, Deep learning
  - Bioinformatics, Segmentation
  - Time Series
- Generative AI
  - Generative Adversarial Networks (GANs), Diffusion, Stable diffusion

Education

2016 – 2020 — **B.Sc. in Computer Science and Engineering,** Khulna University of Engineering & Technology (KUET), Bangladesh – **CGPA 3.86 out of 4.00** (4th position)

Research Publications

Journal Articles

1. Rahman, A., Zaman, S., Parvej, S., Shill, P. C., **Md. Shahidul Salim**, & Das, D. (2025). Fake news detection: Exploring the efficiency of soft and hard voting ensemble. *Procedia Computer Science*, *252*, 748–757. 4th International Conference on Evolutionary Computing and Mobile Sustainable Networks. doi:https://doi.org/10.1016/j.procs.2025.01.035
2. **Md. Shahidul Salim**, & Hossain, S. I. (2024). An applied statistics dataset for human vs ai-generated answer classification. *Data in Brief*, *54*, 110240. doi:https://doi.org/10.1016/j.dib.2024.110240
3. **Md. Shahidul Salim**, Hossain, S. I., Jalal, T., Bose, D. K., & Basher, M. J. I. (2024). Llm qa chatbot builder: A generative ai-based chatbot for question answering. *SoftwareX*.
4. Saad, A. M., Mahi, U. N., **Md. Shahidul Salim**, & Hossain, S. I. (2024). Bangla news article dataset. *Data in Brief*, 110874. doi:https://doi.org/10.1016/j.dib.2024.110874
5. Zubair, M., Mahee, M. N. I., Reza, K. M., **Md. Shahidul Salim**, & Ahmed, N. (2024). Climate data dynamics: A high-volume real world structured weather dataset. *Data in Brief*, 111156. doi:https://doi.org/10.1016/j.dib.2024.111156
6. Ashiqussalehin, M., Jahan, K. N., Rahaman, M. A., & **Salim, Md Shahidul**. (2022). Human abnormal behavior detection using convolution neural network. *Specialusis Ugdymas*, *1*(43), 4076–4083.

Conference Proceedings

1. Bose, D., & **Salim, Md. Shahidul**. (2024). Suggesting bengali words using masked language model. In *3rd international conference on computing advancements (icca)*.
2. Trisha, Shahid, **Md. Shahidul Salim**, Jeba, & Mahbub. (2024). Automated classification of gastrointestinal polyps from endoscopic images using a deep learning approach. In *2024 international conference on recent progresses in science, engineering and technology*.
3. Hossain, L., Hossain, I., **Salim, Md. Shahidul**, 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).
4. Nabil, A., Das, d., **Salim, Md. Shahidul**, 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).

5. Promi, R. T. H., Nazri, R. A., **Salim, Md. Shahidul**, & 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

6. **Salim, Md. Shahidul**, 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

7. **Salim, Shahidul**, 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

8. Ahmed, T., Hossain, S., **Salim, Md. Shahidul**, 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

9. **Salim, MD Shahidul**, 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

- **BConvQA: A Bangla Conversational Question-Answering Model Using Transformer-based Architecture(EMNLP 2024(Revision Completed))**
  – Developed a Bangla Conversational Question Answering (CCQA) system by creating a quality-controlled dataset using machine translation and LLM-based augmentation. Fine-tuned sequence-to-sequence models with contextual prompts to improve accuracy. Released dataset and testing scripts on GitHub, providing a foundation for future research in Bangla conversational AI.

- **Comparing Prompt Based and Standard Fine Tuning for Bangla Text Classification(Expected to submit in ACL)**
  – Conducted a comparative study on prompt-based fine-tuning versus standard fine-tuning for Bangla text classification, using five models across six datasets. Found that prompt-based fine-tuning improved accuracy by 10% on average, showing promise as a robust approach, especially in low-resource settings.
- **Deep learning models for dermoscopic skin lesion hair segmentation: An extensive experimental study**
  – Conducted an in-depth analysis of 10 deep learning models for hair segmentation in dermoscopic images to improve melanoma detection. Evaluated models on multiple metrics and computational complexity, identifying UNet++ with ResNet-50 as the most accurate. For low-resource environments, PSPNet or FPN with ResNet-18 are recommended, while UNet++ or Linknet with ResNet-18 offer a balance between accuracy and efficiency.

Awards and Projects

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

Projects

- **LLM based QA chatbot builder** A generative AI-based chatbot for question answering
- **Medical LLM Chatbot** - Chat with pdf using medical LLM 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** - Confidential Message Sharing
- **Statistics exam**-Design and Implementation of a Python/Flask-Based Randomized Statistics Exam Generator
- **Counterfeit note detection** - Fake Bangladeshi Banknote Detection using Convolutional Neural Networks (CNN)

Miscellaneous Experiences

2023
- Ibex(Supercomputer) fine-tune Mistral for medical data, Hajj data and natural SQL question-answering system
- Research paper reviewer: AAAI-2024, EICT-2024
- Undergraduate thesis coordinator

Technical Skills

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