

Dipayan Biswas

Houston, TX | (901) 549-5748 | dipayan1109033@gmail.com | [LinkedIn](#) | [Github](#) | [Website](#)

SUMMARY

PhD Candidate in Computer Science with 5+ years of research experience in computer vision and multimodal AI. Expertise in developing end-to-end ML pipelines that integrate vision, speech, and language with measurable improvements in summarization quality, object detection, and automated evaluation at scale.

EDUCATION

University of Houston

Ph.D. in Computer Science | GPA: 3.88

Houston, TX

Aug 2021 – Jul 2026 (Expected)

Khulna University of Engineering and Technology (KUET)

B.Sc. in Electronics and Communication Engineering | GPA: 3.88 | University Gold Medalist

Khulna, Bangladesh

Apr 2012 – May 2016

SKILLS

Languages: Python, C++, MATLAB, SQL, PHP, JavaScript

Frameworks: PyTorch, TensorFlow, scikit-learn, OpenCV, DeepEval, Hydra, Lightning

GenAI: HuggingFace, LangChain, Gemini API, Gemma, Qwen3-VL, ChatGPT API

MLOps: Git, Docker, Linux, Slurm, MLflow, W&B, MySQL, AWS EC2, AWS SageMaker

EXPERIENCE

University of Houston

Graduate Research Assistant

Houston, TX

Aug 2021 – Present

- Designed and led **multimodal research** for chapter-structured video summaries with aligned supporting visuals.
- Developed **multi-channel fusion workflows** combining speech, text, and visual signals for video understanding.
- Built automated LLM-based evaluation workflows with DeepEval and ChatGPT/Gemini APIs to assess summaries, improving factual faithfulness by 9.6 points and relevance by 5.2 points over unimodal baselines.
- Boosted visual detection accuracy (90.75% to 95.32%) using transfer learning and semi-supervised auto-labeling.
- Developed a novel graph-based algorithm to group meaningful lecture video objects, improving mAP by 15.8%.

Fellowship.AI

Machine Learning Fellow

Remote

Sept 2025 – Nov 2025

- Built a real estate data pipeline for listing scraping and comparative market analysis, estimating property values.
- Developed a multimodal recommendation system using Gemini APIs to jointly analyze listing images and text descriptions, delivering structured pricing and content optimization insights.

Khulna University of Engineering and Technology (KUET)

Assistant Professor / Lecturer

Khulna, Bangladesh

Feb 2017 – Aug 2021

- Developed a ResNet50 transfer-learning model for plant disease detection, achieving 99.80% accuracy.
- Built deep learning models for chest X-ray pneumonia and breast cancer prediction, achieving 99.28% accuracy.
- Mentored 10+ undergraduate students in ML research projects.

Samsung R&D Institute Bangladesh (SRBD)

Software Engineer

Dhaka, Bangladesh

Jul 2016 – Feb 2017

- Developed panoramic, 360-view, and little-planet visualizations for Samsung Gear 360 iPhone app in Objective-C.
- Optimized mobile image processing and rendering pipelines with OpenCV and OpenGL, improving visual quality and reducing rendering latency.

PROJECTS

Multimodal Video Summarization Engine

Dec 2025 - Mar 2026

- Designed a novel end-to-end multimodal summarization system for the NSF-supported VideoPoints platform, integrating slide visuals, ASR transcripts, and OCR text for chapter-structured summaries and navigation.
- Engineered multimodal fusion workflows using Whisper, Gemma 3/Qwen3-VL, and Gemini Flash with modality-aligned in-context learning, improving cross-modal synthesis quality.

LLM-Based Video Temporal Segmentation

June 2025 - Aug 2025

- Built a slide-anchored LLM chaptering pipeline aligning Whisper transcripts with slide transitions to segment long educational videos into coherent chapters.
- Applied boundary validation and correction logic to enforce contiguous, full-coverage chapter segmentation.

SOTA Object Detection Lab | [Code](#)

Dec 2024 - Feb 2025

- Developed a modular PyTorch framework for training and fine-tuning state-of-the-art object detectors on custom datasets with reproducible experiment workflows.
- Designed Hydra-based configuration with dynamic model loading and unified train/evaluate/predict pipelines, reducing experiment setup overhead.

Long-Distance Face Detection | [Code](#)

Feb 2023 - May 2023

- Proposed an IoU-aware RetinaFace extension with an auxiliary IoU head and multi-task loss for detecting small, distant, and occluded faces.
- Trained ResNet50 and MobileNet0.25 backbones on WIDER FACE, improving robustness at long-distances.

SELECTED PUBLICATIONS ([GOOGLE SCHOLAR](#))

D. Biswas, S. Shah, and J. Subhlok. “Visual Content Detection in Educational Videos with Transfer Learning and Dataset Enrichment”. In *IEEE MIPR*, 2025. [[Paper](#)] [[Code](#)]

D. Biswas, S. Shah, and J. Subhlok. “Lecture Video Visual Objects (LVVO) Dataset: A Benchmark for Visual Object Detection”. In *arXiv preprint arXiv:2506.13657*, 2025. [[Paper](#)] [[Dataset](#)]

D. Biswas, S. Shah, and J. Subhlok. “Identification of Visual Objects in Lecture Videos with Color and Keypoints Analysis”. In *IEEE ISM*, 2023. [[Paper](#)]

HONORS AND AWARDS

Ph.D. Research Showcase Award (Audience’s Choice) — University of Houston, 2025

Cullen Graduate Student Fellowship — Graduate School, University of Houston, 2021-2026

University Gold Medalist — Khulna University of Engineering & Technology (KUET), Bangladesh, 2018

LEADERSHIP

Team Lead, Real Estate Listing Optimization Project, Fellowship.AI

Sep 2025 - Nov 2025

Organizing Secretary, Graduate Student Association of Bangladesh, University of Houston

Aug 2022 - Jul 2025

Academic Advisor and Project Mentor, Khulna University of Engineering and Technology

Feb 2017 - Aug 2021

CERTIFICATIONS

[Generative AI with Large Language Models](#), AWS and DeepLearning.AI (Coursera)

[Machine Learning in Production](#), DeepLearning.AI (Coursera)

[AWS Cloud Practitioner Essentials](#), Amazon Web Services (AWS)

REFERENCES

[Shishir Shah](#), Chief AI Officer & Director of School of Computer Science, University of Oklahoma Email: sshah@ou.edu

[Jaspal Subhlok](#), Professor, Department of Computer Science, University of Houston Email: jsubhlok@central.uh.edu