

# Sri Akash Kadali

240-726-9356 | [kadali18@umd.edu](mailto:kadali18@umd.edu) | <https://www.linkedin.com/in/sri-akash-kadali/> | <https://github.com/Akash-Kadali>

## Education

### University of Maryland, College Park, United States

CGPA/Percentage: 3.78/4

*Master of Science in Applied Machine Learning*

*Expected May 2026*

- **Relevant Coursework:** Scalable Machine Learning, Cloud Computing, Big Data Analytics, Distributed Systems, Data Engineering Pipelines, Database Management Systems, ETL Processes, DevOps for ML, High-Performance Computing, and Advanced Data Visualization.

### Indian Institute of Information Technology, Vadodara, India

CGPA/Percentage: 8.78/10

*Bachelor of Technology in Computer Science and Engineering*

*2020-2024*

- **Relevant Coursework:** Advanced Deep Learning, Computer Vision, Natural Language Processing, Reinforcement Learning, Data Structures and Algorithms, Statistical Machine Learning

## EXPERIENCE

### Research Intern - Generative Models for Hate Speech Detection

May 2023 – December 2023

*Indian Institute of Technology, Indore*

*Indore, India*

- Designed and implemented a DeBERTa-based architecture augmented with emotion embeddings and word-level attention mechanisms using Bi-LSTM for implicit hate speech detection, achieving a 5% improvement in F1-score over baseline models.
- Leveraged supervised contrastive learning to enhance feature representation, resulting in an 8% increase in classification accuracy on the IHSate and IHC datasets.
- Developed emotion synthesis pipelines incorporating sentiment features using the NRC Lexicon, contributing to a 6% boost in model precision.
- Applied extensive data augmentation techniques, including Replace Named Entities (RNE), Replace Scalar Adverbs (RSA), Back Translation (BT), and Generative Models (GM), leading to a 12% reduction in data sparsity.
- Conducted comprehensive model evaluations against state-of-the-art baselines like BERT, RoBERTa, and HateBERT, demonstrating a 10% higher recall in implicit hate speech detection.
- Utilized Azure ML and Databricks for scalable deployment, reducing experimentation time by 20% and ensuring reproducibility across diverse datasets.

### Research Intern, AI-Powered Transformers in Cancer Research

January 2024 – June 2024

*National Institute of Technology, Jaipur*

*Jaipur, India*

- Designed and deployed MaxViT-based models for histopathological image classification, achieving a 92% classification accuracy on large-scale medical datasets.
- Engineered Cascaded Deformable Transformer Layers (CDTL) to improve feature dependency modeling by 20%, optimizing AI workflows in medical imaging.
- Implemented Cascaded Deformable Self-Attention (CDSA) to enhance category-specific feature extraction by 18% across 10,000+ annotated medical images.
- Reduced model convergence time by 25% through integration of skip connections and deformable convolutions, enhancing computational efficiency for real-time applications.
- Developed classification pipelines for breast tumor analysis, achieving a 15% reduction in misclassification rates and an F1-score of 0.91, supporting accurate clinical decision-making.

### Research Assistant, Transformer-Based Graph Model for Bot Analysis

July 2024 – December 2024

*Indian Institute of Technology, Indore*

*Indore, India*

- Achieved a classification accuracy of **99%** on PAN and MiB datasets by effectively distinguishing bot-generated tweets from real ones, showcasing strong expertise in supervised learning techniques.
- Integrated **10+** user-level behavioral features into GraphSAGE and BERT models, significantly enhancing contextual representation and model performance.
- Engineered and processed features from over **500,000** tweets, optimizing model training pipelines and reducing evaluation latency.
- Designed and deployed a scalable system in just **2 months**, leveraging PyTorch and Hugging Face Transformers to meet high-performance demands.

## Teaching Assistant - Digital Logic Design

June 2022 – December 2022

Indian Institute of Information Technology, Vadodara

Diu, India

- Assisted in teaching core courses on Digital Logic Design, helping students understand fundamental concepts through interactive sessions and problem-solving.
- Provided guidance to students on assignments and projects, ensuring a strong grasp of key topics and fostering a collaborative learning environment.
- Conducted tutorial sessions for a class of 100+ students, improving their coding and analytical skills through hands-on exercises.

## PROJECTS

---

### Pre-training for Document Understanding | *Faster R-CNN, Tesseract OCR* August 2022 – December 2022

- Utilized pretrained BERT for language feature extraction and Faster R-CNN for spatial structure detection, enabling a comprehensive analysis of textual and layout information from documents.
- Deployed MFCNN (Multi-Feature Convolutional Neural Network) with pretrained BERT to enhance the representation of linguistic and layout data, achieving significant improvements in feature quality.
- Implemented a novel joint training mechanism that integrates 2D positional embeddings for layout modeling and multi-label classification to generalize across diverse document types.
- Designed a pipeline to preprocess and analyze thousands of documents, employing Tesseract OCR for text extraction and combining it with spatial features to enhance the system's accuracy and efficiency.
- Evaluated the model on benchmark datasets, achieving high accuracy in structured document understanding tasks, with potential applications in automated data extraction and document digitization workflows.

### Optimized Machine Learning for Edge Devices | *PyTorch, Model Quantization* June 2023 – August 2023

- Developed and deployed deep learning models optimized for Qualcomm's SOC compute and ML hardware accelerators using TensorFlow Lite.
- Integrated model compression techniques such as quantization and pruning to reduce model size by 40% while maintaining performance for real-time applications.
- Implemented Reinforcement Learning strategies to optimize dynamic resource allocation in edge computing scenarios.
- Collaborated on parallel computing frameworks using LLVM to improve inference speeds on embedded devices.

## Technical Skills

---

**Programming:** Python (Proficient), C++, C (1+ years), SQL (Postgres), R

**AI Expertise:** Deep Neural Networks, CNNs, RNNs, LSTMs, Reinforcement Learning, Neural Architecture Search

**Machine Learning Frameworks:** PyTorch, TensorFlow, TensorFlow Lite, Scikit-learn, Keras

**Optimization Techniques:** Model Compression, Quantization, Kernel Optimization, Federated Learning

**Computer Vision:** YOLO, Image Processing, Video Analysis

**Compiler Frameworks:** LLVM, GCC, TVM, XLA

**Parallel Computing:** CUDA, OpenMP, Multi-threading

**Tools and Platforms:** Databricks, Docker, Git, Azure ML, AWS, Google Cloud

## Leadership and Achievements

---

**Paper accepted at ICONIP 2024, Auckland: "CaDT-Net: A Cascaded Deformable Transformer Network for Multiclass Breast Cancer Histopathological Image Classification"** September 2024

**Research Paper accepted at IEEE Transactions on Big Data: "Supervised Contrastive Learning with Attention and Emotion Synthesis for Implicit Hate Speech Detection"** December 2024

**Gold Medal for Academic Excellence in B.Tech Program** November 2024

**Student Representative of IIIT Vadodara: Represented the student body, bridging communication between administration and students, ensuring effective resolution of issues, and organizing student-focused initiatives** December 2020 – May 2024

## Volunteering

---

**Volunteered at G20 Summit, Diu: Managed logistics for 50+ international delegates, fostering collaboration and ensuring smooth event operations.** May 2023