Sahibnoor Singh

Phone: +49 1521-8729393 **Email:** SahibnoorSingh@hotmail.com

SUMMARY

Master's student in Statistics and Data Science at LMU Munich specializing in Computer Vision and Deep Learning. Developed an adaptive discrepancy-based reweighting framework (MDRIFT) and an Active Learning pipeline that cut manual labeling by 40%. Experienced in real-time image recognition with segmentation latencies between 15-50ms. Skilled in Python and PyTorch, with strong knowledge of C++, OOP, and Automated Testing. Dedicated to creating robust, high-performance algorithms for challenging scenarios and automated workflows.

EDUCATION

04/2023
Munich, Germany

Statistics and Data Science | Master of Science
Ludwig-Maximilians-Universität München

- Modules: Computer Vision, Deep Learning, Data Structures, & Stochastic Processes.
- Research: Developed and tested predictive ML pipelines; innovated data-driven methods for tech systems using OOP and Automated Testing in PyTorch.

10/2019 – 03/2022
Amritsar, India

Data Science | Bachelor of Science
Guru Nanak Dev University

- Modules: C++ Programming, OOP, Machine Learning, and Optimization.
- Achievement: Top 2% graduate; studied and enhanced algorithm efficiency for Big Data.

APPLIED PROJECTS

06/2026 – present

Research Project | ZEISS
GI Endoscopy Vision-Language Model (VLM)

• Developed a multimodal VLM pipeline combining GI endoscopy image classification, segmentation, and generative language modeling using SigLIP2 and Qwen.
• Fine-tuned Vision Transformer models on HyperKvasir, achieving ~91% classification accuracy, and developed a Kvasir-SEG segmentation model achieving ~0.90 mean Dice.
• Implemented and evaluated self-supervised representation learning with DINO, iBOT, and LeJEPA on a 1.1M+ image endoscopy dataset.
• Built scalable PyTorch training/inference pipelines with Hugging Face, CUDA, cloud GPUs, model checkpointing, evaluation, and interactive VLM deployment.

10/2025 – 03/2026

Master's Thesis | Department of Computational Statistics & Data Science
Machine Learning for Non-Stationary Time Series

- Innovated a reweighting framework (MDRIFT) that allows models to adapt to "shifting" data regimes, such as changing lighting or environmental noise, which often degrade standard ML performance.
- Enhanced an optimization scheme using Difference-of-Convex Algorithms (DCA) to align historical training data with current "target" distributions.
- Conducted extensive ablation studies comparing TRS and ABS formulations, providing actionable insights into balancing model adaptation with statistical stability.

04/2025 – 10/2025

Team Project Applied Deep Learning Lab
Accent-Aware Speech Recognition

- Evaluated the trade-offs of fine-tuning large-scale pretrained models on limited in-domain data, identifying critical failure points where error rates increased by 218%.
- Constructed and executed a contrastive learning pipeline to investigate how signal-based representations generalize across varied acoustic conditions.
- Refined training hyperparameters to mitigate the loss of generalizable patterns from pretraining, ensuring better performance in high-noise environments.

APPLIED PROJECTS

01/2026 – 02/2026

Cloud-Native RAG System on Azure

- Developed and deployed an end-to-end Retrieval-Augmented Generation (RAG) system that automates the extraction of actionable insights from large-scale technical document repositories.
- Pioneered a high-performance retrieval pipeline using Azure AI Search for vector-based indexing and FastAPI for low-latency query handling.
- Enhanced system reliability and portability by containerizing the entire workflow with Docker, ensuring seamless deployment across diverse cloud environments.
- Refined data quality by implementing advanced text chunking with overlap and embedding generation using state-of-the-art models to ensure grounded, source-cited responses.

SKILLS

– MACHINE LEARNING & MODELING

PyTorch | **Scikit-learn** | **XGBoost** | **Optuna**

Deep Learning | **Contrastive Learning** | **Representation Learning** | **Predictive Modeling**

Uncertainty-Based Sampling

– RESEARCH & EVALUATION

Experimental Design | **Benchmarking** | **Hyperparameter Tuning** | **Performance Evaluation**

Simulation-Based Analysis | **Statistical Modeling** | **Distribution Shift**

– PROGRAMMING & TOOLS

Python | **PyTorch** | **Pandas** | **NumPy**

scikit-learn | **Linux/Unix** | **Data Preprocessing** | **EDA**

MLflow | **Object-Oriented Programming (OOP)** | **Data Structures** | **Automated Testing**

– FRAMEWORKS & LIBRARIES

Docker | **FastAPI** | **Azure AI Search (Vector indexing)** | **CMake (Exposure)**

MLflow

LANGUAGES

– LANGUAGES

English
Proficient

German
Intermediate