

# ABHAY SHASHIDHARA

Minneapolis , MN , USA

[abhayshashidhara6@gmail.com](mailto:abhayshashidhara6@gmail.com) | +1 (763)-344-6520 | [abhayshashidhara.github.io/](https://abhayshashidhara.github.io/) | [linkedin.com/in/abhayshashidhara](https://linkedin.com/in/abhayshashidhara)

## ACADEMIC QUALIFICATION

**Master of Science in Data Science , University of Minnesota - Twin Cities , MN , USA | 2025-2027(Expected).**

*Course work: Machine Learning Fundamentals ,Principles of Database Systems, Introduction to Data Mining, Natural Language Processing , Spatial Data Science Research.*

**Bachelor of Engineering in Biotechnology, RV college of Engineering, Bengaluru, India | 2021–2025.**

**Minor in Artificial Intelligence , Indian Institute of Technology, Ropar | August 2024 - August 2025.**

## TECHNICAL SKILLS

- **Programming:** Python, C, C++, R
- **Developer Tools:** GitHub, Google Colab, Jupyter Notebook, VS Code, Google Cloud Platform , Gradio, Streamlit, Kaggle, PostgreSQL.
- **Data Analysis:** MySQL, Excel, Power BI , Tableau.
- **Libraries:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PyTorch, Hugging Face Transformers, OpenCV, PyPDF2,FAISS, SentenceTransformers.
- **Soft Skills:** Teamwork, Time Management, Communication, Problem Solving.

## WORK EXPERIENCE

**Octakagon Bock Pvt Ltd.(BOCK AI) , Bengaluru, India**

December 2024 – June 2025

### AI and Automation Intern

- Curated and processed **10,000+ medical text samples** from clinical textbooks and public health sources using **PyMuPDF** and Python, converting unstructured content into clean **JSON** datasets for LLM fine-tuning.
- Fine-tuned a **DistilGPT-2** model on **domain-specific medical data**, achieving **~82–86%** diagnostic accuracy and improved consistency in symptom-to-diagnosis text generation.
- Built and deployed a **scalable LLM fine-tuning pipeline on Google Cloud Vertex AI**, reducing experiment iteration time by **~30%** and enabling rapid retraining across medical sub-domains.
- Collaborated with a team to build a prototype **text-to-video generation pipeline** leveraging **transformer-based prompt encoding** and **diffusion-based video synthesis models** for generating medically relevant visual sequences from clinical instructions.

**Center of Excellence for Computational Genomics, RVCE , Bengaluru, India**

November - December 2022

### Project Intern

- Implemented a **Generative AI-based equivariant diffusion model (DDPM)** for structure-based drug design targeting the **SARS-CoV-2 spike protein**, generating **196 de novo compounds** with zero matches in major chemical databases.
- Leveraged the framework with **conditional inference and inpainting**, integrating **scoring-based filtering pipelines** to identify **15** high-confidence candidates, including a top-ranked compound with a **-6.93 kcal/mol** binding score and stable dynamics over **100 ns** molecular dynamics simulations.
- Accelerated early-stage drug discovery through **ML-driven in-silico candidate generation**, reducing wet-lab search space and screening costs, and resulting in a **peer-reviewed publication**.

## ACADEMIC PROJECTS

### A machine learning-based diagnosis tool for Parkinson's disease.

- Developed a **multimodal diagnostic system** combining structured patient metadata (e.g., age, comorbidities, symptom onset) and spiral test images for early-stage Parkinson's detection.
- Built a **Retrieval-Augmented Generation (RAG) pipeline** using **OpenChat 3.5** and domain-specific transformers like **PubMedBERT** and **BioClinicalBERT** to generate clinician-style diagnostic responses.
- Integrated **ResNet-18 CNN image features** with RAG-generated clinical text representations in a multimodal diagnostic pipeline, achieving **96.25% accuracy** on Parkinson's disease detection.
- Prototyped the full system using **Gradio** to support real-time clinical inputs and generate interpretive text-based summaries from combined image and metadata sources.

### A Web-Based Platform for Storing and Accessing Genomic, Diagnostic, Proteomic, and MRI Data of Alzheimer's Disease.

- Developed a **Flask-based web application** to centralize **genomic, diagnostic, proteomic, and MRI** data related to Alzheimer's .
- Created **RESTful APIs** with JSON responses and used Python's **OS** module for structured data retrieval and file management.
- Built a responsive and user-friendly frontend using **HTML, Bootstrap, and JavaScript** to enable smooth navigation and real-time data access for researchers and clinicians.

## PUBLICATIONS

- De novo Design of Anti-COVID Drugs Using Machine Learning-Based Equivariant Diffusion Model Targeting the Spike Protein (Published: MDPI, 12 May 2023)
- Impact of Nitrogen Deficiency on Bacopa monnieri: A Computational and Experimental Study (Published: Metszet, Volume 9, Issue 9, September 2024).