

# Sripad Karne

858-275-3303 | [sk5695@columbia.edu](mailto:sk5695@columbia.edu) | [linkedin.com/in/sripad-karne](https://www.linkedin.com/in/sripad-karne) | US Citizen

## EDUCATION

### Columbia University

Master of Science (M.S.) in Data Science

New York City, NY

*Expected Dec. 2026*

### University of California, San Diego

Bachelor of Science (B.S) in Cognitive Science w/ specialization in Machine Learning

San Diego, CA

*Sep. 2021 – June 2025*

## EXPERIENCE

### Data Science Intern

*UC San Diego Health*

May 2025 – Sep. 2025

*San Diego, CA*

- Supporting clinical research at UC San Diego Health by processing and analyzing large-scale EHR datasets (>10 million patients) on Databricks and GCP using Spark SQL, helping integrate data across hospitals using standardized clinical terminologies (ICD-10, CPT, etc.)
- Conducting advanced exploratory data analysis in Python to identify key clinical and demographic risk factors for chronic conditions such as lung cancer and CKD, informing early detection and personalized risk stratification
- Developing and optimizing machine learning models, such as XGBoost and Random Forest, to model disease progression and generate individualized risk trajectories, contributing to an upcoming manuscript on ML-driven CKD prediction
- Designing visualizations and presenting data-driven insights to key partners, including LifeGuard Health Networks and Stand Up To Cancer, to support clinical research and decision-making initiatives

### Instructional Assistant for Advanced Machine Learning Methods

*University of California, San Diego*

Mar. 2025 – June 2025

*San Diego, CA*

- Conducted weekly office hours and sessions to deepen student understanding of ML methods, from regression and classification to Generative Modeling, Transformers, and LLM models
- Provided individualized guidance on complex topics, offering debugging support, and clarifying assignments

### Computer Vision Intern

*Machine Learning, Perception & Cognition Lab*

June 2024 – Mar. 2025

*San Diego, CA*

- Designed a novel controllable image generation architecture, leveraging cutting-edge Diffusion Transformers and state-of-the-art deep learning models
- Integrated LLMs into data pipelines for processing multi-terabyte video datasets, using them to extract and refine open-vocabulary object prompts and dynamically guide bounding box detection at scale
- Conducted in-depth model testing and evaluation, applying various performance metrics and validation techniques to ensure the robustness, accuracy, and efficiency of the model

### Data Analyst Intern

*Neural Engineering and Translation Labs*

Jan. 2023 – Apr. 2024

*San Diego, CA*

- Developed and implemented machine learning models to analyze behavioral and questionnaire data, generating personalized recovery plans for patients displaying signs of depression and other mental health conditions
- Contributed to the design and implementation of efficient data pipelines to streamline data flow and enhance data accessibility for the research teams, as well as doing rigorous data exploration, cleansing, and preprocessing to ensure data integrity

## PUBLICATIONS

Hwang, H., Yoon, S., **Karne, S.**, Boussina, A., & Sitapati, A. (in progress). *Machine learning-derived risk trajectory of chronic kidney disease and clinical implications after disease onset.*

## TECHNICAL SKILLS

**Languages:** Python, SQL, R, MATLAB

**Frameworks & Libraries:** PyTorch, TensorFlow, AWS, GCP, Snowflake, Databricks, NLP, LLM, Docker, Apache Spark, Pandas, NumPy, Scikit-learn, XGboost, RandomForest, PySpark, Excel, Tableau, Git, Linux, LangChain

**Relevant Coursework:** Advanced Machine Learning Methods, Modeling and Data Analysis, AI Algorithms, AI Engineering, Algorithms for Data Science, Neural Networks and Deep Learning, Exploratory Data Analysis and Visualization, Business Analytics