Statement of Purpose

Shreyas Kalvankar *M.S. Applicant*

My research interests lie in uncovering the mathematics that drives Machine Learning and deep learning and their theoretical nature. What are the theoretical foundations explaining the success of deep networks? Are there inherent limits to their expressiveness? What are the potential applications of learned representations in various domains, such as computer vision, natural language processing, and speech recognition? How do we address the challenges of interpretability and explainability in these learned representations? I aim to explore these problems to develop an understanding of the theoretical limits of Machine Learning. I am particularly interested in understanding and applying these algorithms across engineering and natural sciences to build intelligent systems and conduct research in physics and mathematics to solve problems such as approximating numerical computations using neural networks, astronomical simulations, assisted theorem proving, etc. My unwavering passion for engineering and research is evident in my academic achievements during my undergraduate years, reflected in my transcripts and my active involvement in extracurricular activities. My journey has underscored the importance of further specialization and enrichment, which I believe a master's program can provide, laying the foundation for a future Ph.D. pursuit. In this regard, I am convinced that North Carolina State University is the ideal place to nurture my potential.

During my bachelor's I was extremely fascinated by astrophysics and its intersection with computer science. In my junior year, I led a team of three and delved into the research on neural network applications in galaxy morphology classification. Identifying galaxy morphologies has important implications in many astronomical tasks, e.g., studying galaxy evolution. The recent data influx in astronomy necessitates a robust and automated system for processing large amounts of images. We aimed to set up a 7-class classification system that classified galaxy images using CNNs, which can surpass existing benchmarks. I was responsible for establishing a robust training pipeline and optimizing various neural network architectures from the ground up. Our collective efforts outperformed the second-best submission on the Kaggle Public Leaderboard. I gained a profound understanding of the intricacies of applying deep learning to various tasks and explored aspects such as hyperparameter tuning, debugging, troubleshooting, and custom model design—each underscored by a need for a profound grasp of the underlying theory. My role as the project lead was a significant learning experience that taught me various aspects of management, communication, and project planning. Through this project, I honed my research skills, nurtured my critical thinking abilities, and sharpened my decision-making capabilities.

For my bachelor's thesis, I worked on utilizing Conditional GANs for astronomical image colorization. Space archives are filled with large amounts of low-quality, greyscale images, many of which go unnoticed. I was interested in utilizing generative models for creating aesthetically pleasing representations of celestial scenes. Generative models are attractive because they enable us to better understand, create, and work with data, having many practical and theoretical implications. In my work, I focused on tasks like image-to-image translation and style transfer, where I realized the need for distributional robustness in my application to maintain quality in my output images, even in varying conditions. GANs often deal with high-dimensional data and are used for various conditional image generation tasks, hence, it is important to study their optimization landscape and convergence guarantees. In my ongoing academic journey, I'm equally interested in gaining a theoretical understanding of how various neural network models generalize, especially in complex, high-dimensional spaces when provided with different conditions or inputs. Undergraduate studies seldom comprehensively cover advanced machine learning theory. It is necessary to understand the generalization bounds of different learning algorithms to prepare for a successful research career in ML. I wish to study these topics during my graduate studies at NCSU as a part of the Special Topics in Computer Science course (CSC-295 & CSC-495) and further my understanding.

The Galaxy Morphology Classification project proved to be a stepping stone in connecting astrophysics to computer science. Soon after that project, I began contributing to EinsteinPy, an open-source Python package designed to address issues in General Relativity and gravitational physics. My work specifically involved incorporating various symbolic computations, such as the Reissner–Nordström metric and calculations for the event horizon and ergosphere of a Kerr-Newmann black hole. While symbolic computations are valuable in controlled environments, real-world problems often necessitate approximate rather than exact solutions. For this reason, the potential for deep learning in simulating such complex system behaviors is vast, e.g., in surrogate modeling where we can use deep learning to approximate behaviors in complex systems allowing for faster evaluations when original simulation using numerical methods is computationally expensive. In simulations, gravitational interactions are currently computed using hierarchical tree algorithms like the Barnes-Hut & Fast Multipole Method. I am interested in leveraging AI and Machine Learning for this task in the long term. One possible approach is to model the gravitational interactions of various particles as nodes of a graph and learn the patterns as they evolve over time by using Graph Neural Networks. I believe Prof. Xiaorui Liu's expertise in Graph Neural Networks can help me immensely in my future research pursuits. Owing to an overlap of interests in large-scale optimization, ML with graphs, and robust & certifiable ML, he would be a compelling choice as a master's advisor.

After completing my degree, I embarked on my professional journey, starting as a Machine Learning Engineer and later transitioning to a Machine Learning Consultant role at the startup Relfor Labs. My primary focus revolved around applying deep learning to audio data analysis and classification by devising novel architectures. I worked on designing a system for pre-processing large amounts of raw audio data by converting it into mel spectrograms, refining the labels. I set up an end-to-end training pipeline to streamline experimentation. Throughout my work on this and various other ML and data-intensive projects, I have been interested in setting up self-sustaining models which can keep learning in the long term. However, I have faced several challenges in my attempts to achieve this which include data quality and consistency issues, model drift, and version control. My experience in developing ML models and designing ML pipelines, along with the challenges I faced make the prospect of working with Prof. Jung-Eun Kim to build sustainable and autonomous systems extremely fascinating.

The scientific landscape has been consistently evolving, but many longstanding problems have eluded complete solutions for many years. Looking into the future, I wish to make a difference in this significant era. My determination to pursue a research career is unwavering, and I aspire to work in an academic setting and eventually pursue a Ph.D. The vibrant community of students and researchers at NCSU is the perfect place for me to nurture myself and further my commitment to help expedite this scientific advancement. Thank you for considering me as a prospective student at your university.