Research Interests and Background: My research interests lie at the intersection of Machine Learning (ML), Computer Vision (CV), and Human Computer Interaction (HCI). I am particularly interested in developing robust ML algorithms for effective diagnosis and monitoring of diseases. To this end, I am excited about mobile healthcare *i.e.,* utilizing the data from mobile sensors and using on-device machine learning to truly democratize healthcare. My prior research involves using machine learning, image processing and optical physics to tackle key challenges in computer vision and healthcare including (1) image compositing, (2) unsupervised representation learning, (3) image understanding with limited data, (4) anomaly detection, and (5) AI-based mobile medical diagnosis. These have resulted in publications at top conferences like **IJCAI, WACV, SIGIR, EMBC, and IMWUT/Ubicomp**. I wish to earn a Ph.D. to conduct research and build technological solutions that have real-world impact.

Computer Vision and Machine Learning: I delved into research in my sophomore year, when I started working with **Prof. P. J. Narayanan** at the Center for Visual Information Technology (CVIT) at IIIT-Hyderabad. Initially, I explored problems in the space of Computational Photography and Structure from Motion. For my first major research project, I worked on the task of **color-consistent background replacement**. The aim was to automatically replace a dull or overexposed sky in outdoor images with an aesthetically pleasing one. I proposed a data-driven approach that uses an ensemble of hand-crafted and deep-learned features to encode the image foreground and illumination. These features are used to retrieve candidates for sky replacement. We then use the segmentation masks and image blending to generate realistic composites. To add more diversity, the composites are re-ranked by optimizing a max-sum diversification objective. We validated our approach with thorough experiments and a user study. I received the **Dean's Research Award** for this project, and it was published at **MMM 2018**.

This motivated me to work on deeper problems and I got interested in **representation learning** to model the style and content in images. Style recognition and similarity are important measures for understanding abstract concepts like art, fashion, design, etc. However, the definition of style is contextual and vague. Deep neural networks (DNNs) perform well on image understanding tasks but require densely annotated large-scale datasets. Obtaining such annotations is expensive (in terms of time and money) and often infeasible. I developed a DNN based framework for learning a neural embedding that captures the 'look and feel' of an image. Unlike previous supervised learning methods, the proposed framework is **unsupervised** – it does not use categorical labels but uses a gram matrix feature-based clustering to get proxy labels for forming triplets of anchor, positive, and negative images. These triplets are used to train a siamese network with a triplet loss to learn an embedding useful for style-based search and retrieval. Despite being 16 times more compact than traditional representations, the embeddings achieved state-of-the-art results for style-based image retrieval and recognition on six datasets with different notions of style. Thus, the proposed method provides a general framework to learn style representations and can also be combined with other proxy measures for style-aware grouping. This work was published at **IEEE WACV 2020**.

My research journey at IIIT-Hyderabad culminated in the form of my Master's thesis. Working at CVIT introduced me to the fascinating world of research, where I developed a solid foundation in computer vision and machine learning, and skills to collaborate effectively and conduct principled research. The brainstorming sessions with my advisor and senior graduate students taught me to communicate well, ask questions, and develop methods for answering them. I left CVIT feeling optimistic about research and wished to explore similar problems further.

After completing my thesis, I joined the **Media and Data Science Research** lab at **Adobe** as an intern. Here I worked on building an efficient solution to detect and segment unseen (not seen during training) object classes in images. Image segmentation is crucial for visual understanding and is a common use case for software like Adobe Experience Manager – a digital asset management solution for brands like Audi, Nike, etc., to maintain their product images. However, supervised methods fall short because of limited availability of good quality annotated data. An automated solution to segment 'novel' object classes with minimal supervision can save a lot of time and effort. During my internship, I carried out an extensive literature survey of the domain and narrowed our focus to the **Few-Shot Image Segmentation** problem, *i.e.,* learning to segment a query image given only a few samples (1-5) as support for each unseen object class. I proposed a novel method that improves the similarity propagation between the support and query image features to get accurate segmentation. The method introduces two key components which provide a non-trivial improvement ($> 6\%$ mean-IoU) over prior methods. First, we jointly predict the support mask and the query mask to enforce self-similarity. Secondly, we introduce a novel foreground-background attentive fusion mechanism to utilize similarities in the background regions of the query and support images. The method achieved state-of-the-art performance on the one-shot and five-shot segmentation benchmarks. The work was published at **IJCAI 2020**, and a **US patent** has been filed for the same.

While working on this project, I observed that class-conditional similarity matching could only match pixels with a similar class mix between the query and the support images. Hence, I wish to explore introducing inductive biases via representations that capture the geometric structure and physics of the real-world to improve visual understanding. My internship experience enhanced my skills as an engineer and researcher. I learned to write well-documented and optimized code, and build applications to integrate my research project as a software tool.

Applying ML, Vision, HCI to Healthcare: To further my research experience, I joined **Microsoft Research (MSR)** as a **Pre-doctoral Research Fellow** (residency program). I am currently working with **Dr. Mohit Jain** and **Dr. Nipun Kwatra** on developing AI-based low-cost diagnostic solutions. Over the past year, I have worked on building a smartphone-based corneal topographer to diagnose keratoconus – a severe eye disease. Such a device can be highly useful in rural and remote locations where modern medical facilities are inaccessible. A commercial corneal topographer is an expensive (∼$10,000) device with proprietary hardware and software. It projects a concentric ring pattern on the eye and reconstructs the corneal surface using ray-optics and 3D geometry principles. The generated topography maps highlight any deformities in the cornea. I directed this project from end-to-end, and after months of consistent efforts we built a working prototype. Building a low-cost device with the same fidelity as a commercial topographer required several key design and technical innovations. First, I designed a 3D printed conical attachment to project a concentric ring pattern on the eye. Second, I developed an AI-assisted smartphone app to ensure proper alignment and image quality with a hand-held device. Third, to process the captured images I implemented an analysis pipeline that involved image processing and a 3D reconstruction algorithm to generate the corneal topography. And to automatically detect keratoconus, we trained a CNN-based classifier on the output maps. Further, due to lack of hardware sensors on smartphones I developed an approximate vision-based algorithm to estimate the working distance (between camera and eye), which is required for reconstruction. We conducted an extensive clinical evaluation of our system (at Sankara Eye Hospital) and found that it performed at par with the gold-standard medical device 'Optikon Keratron'. This project led to a publication at **ACM IMWUT/Ubicomp 2021**. The proposed attachment costs **$33** (**303x** less than Keratron) and weighs just **140 grams**, making it ideal for use by community health workers. We are now working towards deploying the device at Sankara clinics across India for mass screening of keratoconus, through which we envision early interventions and timely treatment.

Automatic detection of diseases can aid in early interventions and prevent delayed treatment. In this context, I worked on the task of **anomaly detection** from auscultation signals. I developed a "ResNet34" based lung sound classification model to identify abnormalities like crackling and wheezing. Interestingly, a simple CNN model worked best for this task compared to Hybrid CNN-LSTM networks and Transformer networks, which had been the focus of prior work. We proposed and combined 'novel' data augmentations and transfer learning techniques to achieve top scores on a small-sized dataset. The work was accepted at **EMBC 2021**. This reinforced my motivation to study machine learning as a continuous process that builds effectively on past knowledge and adapts to new tasks with minimal supervision, especially in medical domain where there is lack of labeled data.

My experience at MSR broadened my perspective on how technology can have an impact beyond the realm of computer science as a field. Thus, I wish to channel my efforts towards using machine learning to solve problems in healthcare and medicine. Through my work, I noticed that there exists a huge gap between research and deployment. Once the system is deployed, many unforeseen scenarios can arise – system crashes, unexpected data (due to changing environment), and edge cases. These unexpected scenarios are not accounted for when working in a controlled lab setting. This emphasized the importance of building robust and generalizable systems.

Ph.D. and Future Goals: After completing my Ph.D., I wish to continue in academia. I have found learning about new problems, conducting research, collaborating with fellow researchers, and teaching – fulfilling and enriching. A Ph.D. would help me build a strong foundation and depth to achieve this goal. This decision is informed by my research background and positive experience working as a teaching assistant for graduate and undergraduate level courses. My experience as a TA improved my understanding of multiple core subjects, and it was gratifying to see students indulge in critical thinking and ask challenging questions. During my Ph.D. I wish to continue working on building efficient and trustworthy machine learning models to aid in healthcare. Modern healthcare today is expensive, requires sophisticated medical devices, and is inaccessible to a large fraction of the population. The synergy between human and artificial intelligence can help build more accurate, reliable and affordable diagnostic systems. Building end-to-end AI systems that can be deployed in the real-world require several components: efficient data-collection pipelines, robust machine learning algorithms, interpretable outputs, and exhaustive testing protocols to guarantee reliability. I wish to contribute towards building such systems through my research.

At **MIT Media Lab**, I am interested in working with **Prof. Fadel Adib** on his research efforts towards building wireless health sensing systems, and developing robust visual systems to aid in robotic perception. My background in applying machine learning and computer vision to building and deploying healthcare systems aligns well with his current projects. I am also excited by the work done in **Prof. Ramesh Raskar's** Camera Culture group on building imaging systems for healthcare applications. I believe that my current experience at MSR in developing end to end AI-based healthcare systems will add value to such research. Additionally, I am also interested in collaborating with **Prof. Pattie Maes** on projects related to health and well-being. I especially appreciate her work on building wearable technology to transform human life. MIT has a strong and diverse faculty body, close collaborations across research groups, and hence, is ideal for conducting fundamental and inter-disciplinary research. I believe my strong background in research, software engineering experience, and alignment of interests with faculty makes me a good fit for MIT Media Lab's graduate program.