
Capstone Faculty-Sponsored Project (FS#7)-- Fall 2019

Faculty Sponsor(s):

- Tal Danino, Assistant Professor, Biomedical EngineeringFu Foundation School of Engineering and Applied Science;Email:td2506@columbia.edu; Lab: <http://daninolab.nyc/>
- Svebor Karaman, Postdoc, Electrical Engineering ; Email: sk4089@columbia.edu

Project Description: 2. MOTIVATION, BACKGROUND AND OVERVIEW

Bacteria-related illnesses – caused by strains of Escherichia coli, Salmonella enterica, Listeria monocytogenes, Vibrio cholerae, Corynebacterium diphtheriae, methicillin-resistant Staphylococcus aureus (MRSA), and Pseudomonas aeruginosa—are responsible for approximately 5 million deaths per year worldwide. Many of these infectious diseases are not only prevalent in the developing world, but are becoming more common in the developed world with increasing instances of food contamination, hospital-acquired infections, and bioterrorism. Identification and monitoring of bacteria outbreaks and antimicrobial resistance is critical for tracking health developments and is recommended by the WHO. Current microbial identification approaches can be extraordinarily time consuming, laborious, and expensive. Specifically, in the case of gold-standard clinical microbiology assays, such as staining and microscopy following culture on solid media in Petri dishes, the assistance of experts for morphological identification is required. Furthermore, these identification methods are often sample destructive. Thus, the development of a rapid, automated process to identify and characterize bacterial species from environmental and clinical samples would be a major health innovation. The aim of this proposal is to develop a machine-learning based approach for the identification and characterization of bacteria via analysis of photographs of bacteria colonies from environmental samples grown on solid media in Petri dishes. The aim would be to build towards machine learning algorithm which could classify colonies on a plate into species, based on aspects such as colony morphology, color, and other features.

Dataset Description: Name: Images of bacterial colonies from soil samples Description: Scanned images of Petri dishes with well-separated colonies from a single species per plate (for training) and scanned images of Petri dishes, captured by Epson scanner at resolution ranging from 400 to 3200 dpi. All jpeg format. Additional Petri dishes from different species on the same plate (goal is to classify these).

Data format: Raw images

Project Deliverables: The goal of this project is to build an algorithm that analyzes an image of a Petri dish with colonies from different bacterial species and classifies them by species. One such deep learning classifier has already been developed to classify colonies from two species, K. pneumoniae and E. coli, which could be extended for the more varied dataset. However, there are a variety of approaches and the research goal of this project is to determine which types of algorithms for this task can perform best.

Possible deliverables:Software

Required background and skill sets: See below

Learning opportunities: Project planning and scoping, Data cleaning, Exploratory data analysis and visualization, Supervised modeling, Unsupervised modeling, Establishing evaluation metrics, Working with image data, Working with time series data