Rapid Prediction of Bacterial Growth Inhibition using Google's Coral AI Platform

DESIGN DOCUMENT

Team 18
Client: Meng Lu
Team Members:
Declan Costello
Shijin Huang
Wenjie Jin
Owen Schacherer
Ani Tangellapalli
Iris Top

sdmay21-18@iastate.edu
http://sdmay21-18.sd.ece.iastate.edu

Revised: 10/4/2020 V1

Executive Summary

Development Standards & Practices Used

- We will be using Google's recommended workflow & standards to develop the TensorFlow ML model

Summary of Requirements

- Portable, Hand-held device to collect bacteria video data
  - Must be usable in a lab setting
  - Must be able to collect video of E. coli samples
- Machine Learning model to detect and classify wild & anti-microbial E. Coli bacteria

Applicable Courses from Iowa State University Curriculum

List all Iowa State University courses whose contents were applicable to your project.

- COMS 227/228
- COMS 474
- CPRE 288
- CPRE 482x

New Skills/Knowledge acquired that was not taught in courses

- Understanding of Google Coral AI board
- Developing a new ML model
- CAD modeling using SolidWorks

Table of Contents

List of figures/tables/symbols/definitions

1 Introduction

1.1 Acknowledgment

We would like to thank our project advisor Dr. Meng Lu, Shirin Parvin, Rachel Shannon, and our team members.

1.2 Problem and Project Statement

Every year waterborne diseases cause a substantial economic burden, costing more than $2 billion in treatments in the US alone. Roughly 90 million patients fall ill per year to conditions such as Escherichia coli (E. coli). E. coli, one of the most common public health concerns, is spread through drinking water, contaminated food consumption, and contact with infected animals or people. Recently, certain strains have become immune to Penicillin, a common antibiotic. Therefore, it must be detected early to avoid any infections by the super disease.

Several E. coli detection methods exist, such as culturing samples on solid agar plates or in liquid media. The use of liquid growth media provides high sensitivity; however, it requires at least 18 hours for the final read-out. Solid agar plates are more cost-effective and more flexible but often take 24 to 48 hours to grow. It is also possible to use molecular detection methods to reduce the assay time to a few hours; however, the results lack the sensitivity of the tests mentioned previously. There is a strong need for an automated method that can achieve rapid colony detection with high sensitivity to accelerate the identification of dangerous diseases in a laboratory setting.

To provide a powerful alternative that can rapidly detect and classify resistant vs. non-resistant E. coli, we propose a system that will collect live growth data of E. coli with which it will use to classify the bacteria into the two required categories. The system will be composed of a physical device to collect the visual data and a software component to detect and classify the bacteria. The device will be capable of accumulating a video feed of E. coli samples. The video will be of sufficient length and quality to obtain the most accurate predictions. Due to restrictions in the lab, the device is small and portable. The software component is composed of a runner program and an ML model. The results from our system will accelerate the detection of resistant E. coli by many hours, which can help avoid many infections.

1.3 Operational Environment

During the fall semester, our ML experiments will be conducted using TensorFlow. However, factors such as environment and weather cannot be ignored. Therefore, in the final test, we will consider the growth rate of bacteria in different environments and whether the bacteria survive. For example, whether it is surface water or groundwater, rainwater, or snow water, there will be bacteria. According to the oxygen demand for bacteria, it can be divided into three categories: anaerobic bacteria, facultative anaerobes, and aerobic bacteria. Salmonella Enterica is one of the most common bacteria in water. Under normal circumstances, it can survive for 2-3 weeks, and it can survive for 3-4 months in the refrigerator. Its optimal breeding temperature is 37°C, and it can reproduce in large numbers above 20°C.

In the final test, we can study the growth rate and survival of bacteria at low temperatures. In addition, we can also compare the growth rate of different types of bacteria in different environments to determine which bacteria are the most threatening.

1.4 Requirements

Functional requirements

- The machine learning model must be able to detect resistive bacteria with an accuracy of 90%.
- The machine learning model must be able to analyze at least 10 minutes of video
- The mobile component should allow users to take and store video feed
- The whole system must be portable and be held and usable in the user's hands

Economic requirements

- The solution should be developed under a $500 budget

Environment requirements

- Keeping team members safe when working in the lab is our first priority
- Lab substances must be used and disposed of correctly
- Everyone should wear proper protective equipment and follow rules and instructions in the lab
- Everyone must take care of and be responsible for our lab equipment

1.5 Intended Users and Uses

Anyone whose job is related to dealing with the habitat of E. Coli could be the potential user of this project. Users could be from farmers to workers of the food industry and workers of the water purification market.

1.6 Assumptions and Limitations

Assumptions:

- We detect and classify all bacterias and their rate of growth
- This research only used around Ames to help farmers.
- If the data we tested are not accurate or something goes wrong, we will retest them all until correct.

Limitations:

Limitation	Expression
Budget	No more than $500
Detection/Classification	All test will be on network and lab
Schedule	All tese will be end beofre 11/18/2020
Function	Mechine Learning model

Figure 1.6.1 - Limitations of the project

1.7 Expected End Product and Deliverables

Portable device (May 2021)

- The portable device will be created using off-the-shelf components. It will contain the Google Coral AI board, which is responsible for running the ML model, and a means of collecting videos of E. coli samples. The designed device will be handheld and portable to be used in a laboratory setting. The device will be running on a battery capable of powering both the Coral AI board and the video collection unit.

Machine Learning Model (January 2021)

- The ML model will be created using TensorFlow 2.0. The model will be capable of running on the Google CoralAI platform and it is responsible for locating and identifying the bacteria in the provided data. Specifically, the model will be capable of differentiating between wild and antimicrobial-resistant E. Coli bacteria with a high accuracy. The input data format must be an image or a single frame from a video.

2 Project Plan

2.1 Task Decomposition

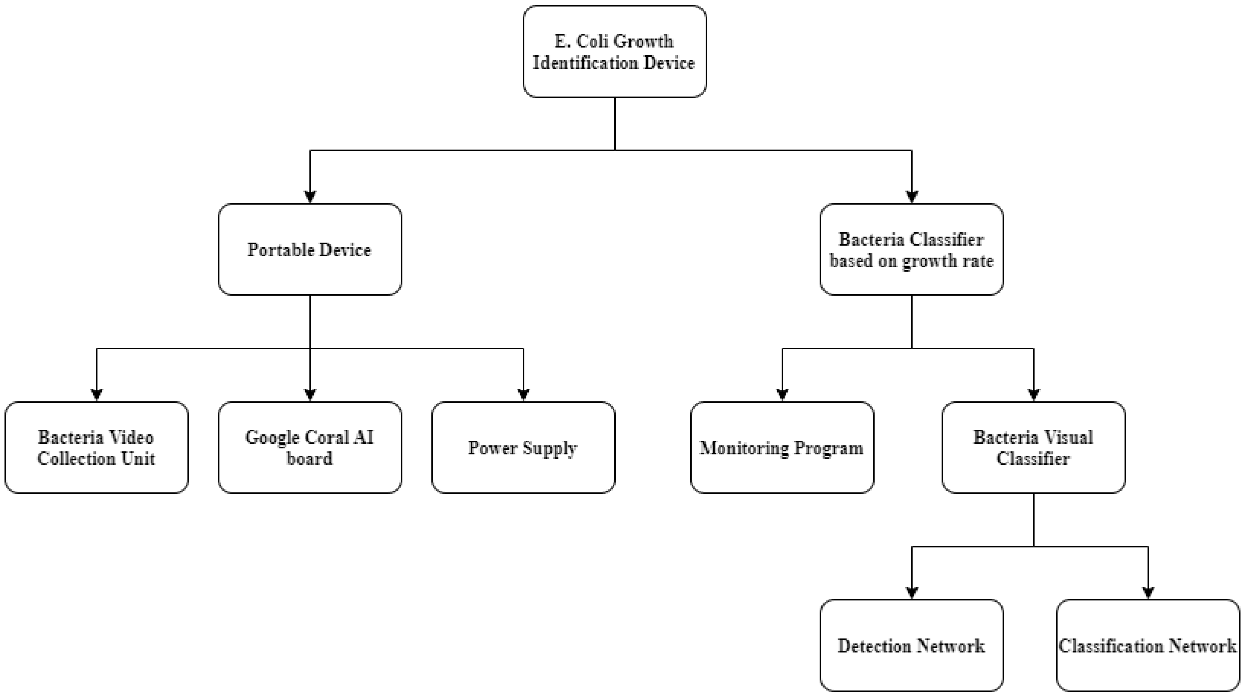

Figure 2.1.1 - Graphical representation of the tasks and their dependencies

- Portable Device
  - Houses the Google Coral AI board, the video collection unit and the power supply
- Bacteria Video Collection Unit
  - Capable of collecting video of E. coli samples
- Google Coral AI Board
  - Similar to a Raspberry PI; Responsible for running the ML model
- Power Supply
  - Responsible for powering the components on the portable device
- Bacteria Classifier
  - The full application including ML model and the runner program
- Monitoring Program
  - The runner program for the ML model and displaying results
- Bacteria Visual Classifier
  - The ML model capable of identifying resistant vs. non-resistant bacteria
- Detection Network
  - ½ of the ML model which locates any object in the given image
- Classification Network
  - ½ of the ML model which identifies the objects from the detection network into bacteria, dirt, etc.

2.2 Risks And Risk Management/Mitigation

Risks for our project include scope, hardware, and COVID. When gathering training data, uncontrolled changes and continuous growth of the scope of our project can occur. As we collect training data, we can sample out valid data at the cost of time. Another risk for our project is the hardware and software malfunctioning. Malfunctions can be mitigated by investing in better equipment as well as trying other variations of equipment. Another risk is COVID in general. COVID can make it hard to keep up with the current restrictions put on campus to go and physically collect our sample data. This can be mitigated by overcommunicating with our supervisors when talking about the current precautions. COVID can also harm our group's availability to meet. This can be mitigated by using better software to meet and communicate.

2.3 Project Proposed Milestones, Metrics, and Evaluation Criteria

Some key milestones in our proposed project include mastering TensorFlow, choosing a machine learning algorithm, and choosing the correct metrics to measure our project. More milestones include collecting up to about 12,000 valid training data sets and revisiting and optimizing past tasks. These soft goals will help us to reach our hard goals of raising our machine learning models to 80% accuracy. This agile project will grow with iterations as we go back and optimize different past tasks and collect more sample data.

2.4 Project Timeline/Schedule

A Gantt chart has been created in google sheets for the team to use as a project timeline tracker. An overview of this gantt chart can be seen in figure X-1 and the whole gantt chart can be seen here. Our goal this semester is to design the physical components, create the needed embedded software to run the physical component, and train the bacteria detection models in TensorFlow. The goal next semester is to build the prototype and combine all the components into one and test.

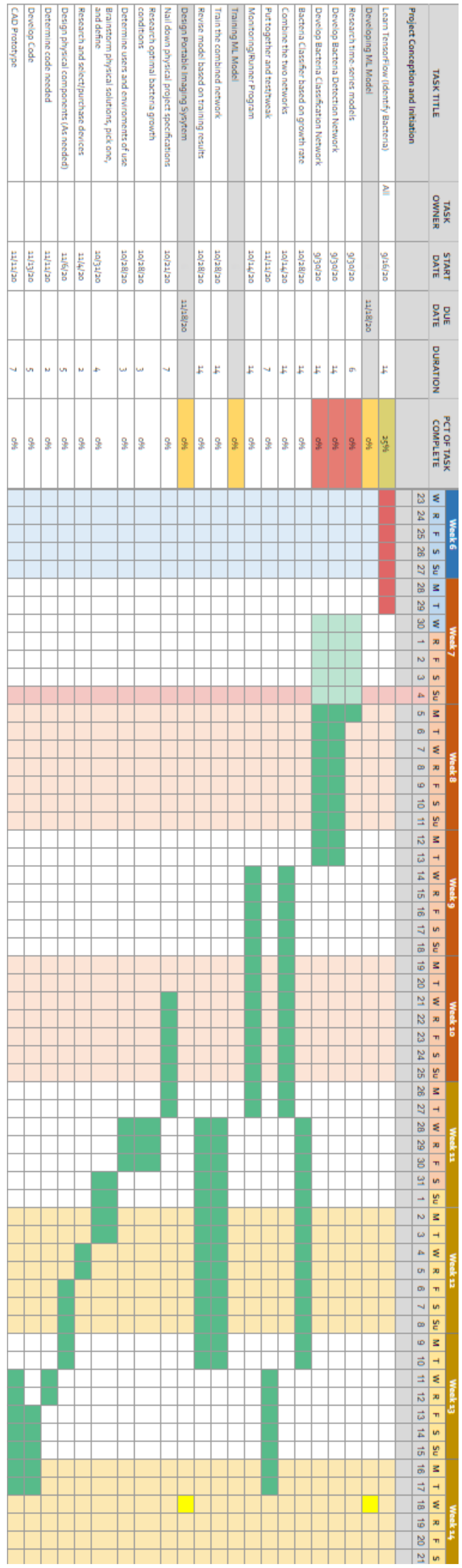

Figure 2.4.1 - Overview of the project timeline

This semester can be split into two major components both of which we plan to complete by November 18th. The first of which is a machine learning model that can accurately predict if bacteria is resistant (identification and classification) based on its initial growth. This can be broken down into 6 different sub components/tasks: bacteria detection DNN, bacteria classification DNN, bacteria classification based on growth rate, combining the networks, monitoring the DNN, training, and revisions. The timing of each of these tasks can be seen below in figure X.

TASK TITLE	START DATE	DURATION	PCT OF TASK COMPLETE
Developing ML Model			0%
Research time-series models	9/30/20	6	0%
Develop Bacteria Detection Network	9/30/20	14	0%
Develop Bacteria Classification Network	9/30/20	14	0%
Bacteria Classifier based on growth rate	10/28/20	14	0%
Combine the two networks	10/14/20	14	0%
Put together and test/tweak	11/11/20	7	0%
Monitoring/Runner Program	10/14/20	14	0%
Training ML Model			0%
Train the combined network	10/28/20	14	0%
Revise model based on training results	10/28/20	14	0%

Figure 2.4.2 - Initial Timeline for training of the machine learning model

The second major component we will be working on this semester is the physical device and the code that will run this device. We will be following the design process to create this component to ensure we create the most viable product. It can be broken down into nailing down the project specifications, researching bacteria growth (safety, optimal conditions, sizing, etc), determining the users and the environments, brainstorming (and direction selection), selecting the store bought components (researching), design the physical components, determining the code required and writing it, and creating the final CAD prototype. The timing for each step can be seen in figure X+1 below.

TASK TITLE	START DATE	DURATION	PCT OF TASK COMPLETE
Design Portable Imaging Sysytem			0%
Nail down physical project specifications	10/21/20	7	0%
Research optimal bacteria growth conditions	10/28/20	3	0%
Determine users and enviroments of use	10/28/20	3	0%
Brainstorm physical solutions, pick one, and define	10/31/20	4	0%
Research and select/purchase devices	11/4/20	2	0%
Design physical components (As needed)	11/6/20	5	0%
Determine code needed	11/11/20	2	0%
Develop Code	11/13/20	5	0%
CAD Prototype	11/11/20	7	0%

Figure 2.4.3 - Initial timeline for the design of the physical prototype

The goal of the second semester is to take the components we have already created, combine them together, and test. This will be accomplished by uploading our trained model on to the Google Coral board and have it analyze the real time video from the physical system.

2.5 Project Tracking Procedures

We will be using a variety of softwares to track our progress and communicate on this project. We will be using Git & Gitlab as our version control tool, Microsoft Teams to communicate, a shared Google drive to store documents, and a Google Sheets document as a Gantt chart to keep track of our progress.

2.6 Personnel Effort Requirements

The textual reference for this work table will be the gantt chart detailed in the above section (2.4).

A day's worth of projected effort will be estimated 30/6/5 hours (1 hour) per person.

Task	Assigned (if it has yet)	No. Days	Projected Person-hours
Learn TensorFlow (Identify Bacteria)	All	14	84
Developing ML Model			83
Research time-series models		6	6
Develop Bacteria		14	14

Detection Network			
Develop Bacteria Classification Network		14	14
Bacteria Classifier based on growth rate		14	14
Combine the two networks		14	14
Put together and test/tweak		7	7
Monitoring/Runner Program		14	14
Training ML Model			28
Train the combined network		14	14
Revise model based on training results		14	14
Design Portable Imaging System			38
Nail down physical project specifications		7	7
Research optimal bacteria growth conditions		3	3
Determine users and environments of use		3	3
Brainstorm physical solutions, pick one, and define		4	4
Research and select/purchase devices		2	2
Design physical components (As needed)		5	5
Determine code needed		2	2
Develop Code		5	5
CAD Prototype		7	7

Figure 2.6.1 - Breakdown of each task and approximate effort required

Total number of projected person-effort hours: 335

2.7 Other Resource Requirements

Resources we will be using throughout the semester to complete our project are listed below:

- Lab via Client
- Microsoft Teams
- Google Coral A.I. Hardware
- Google Drive
- GitLab
- TensorFlow
- Python
- Parts for physical system from vendors
- Solid Works
- Powerful computer for DNN training

2.8 Financial Requirements

We will be allotted a total of $500 for this project. The only financial expenses will be from the creation of the portable system's prototype.