

# CS CAPSTONE PROBLEM STATEMENT

OCTOBER 10, 2017

CS461 FALL 2017

## REAL-TIME SEED IDENTIFICATION

PREPARED FOR

OREGON STATE UNIVERSITY CROP SCIENCE  
DEPARTMENT

DANIEL CURRY

PREPARED BY

GROUP 11  
GREEN TEAM

OMAR ELGEBALY

### Abstract

The purpose of this project is to develop a solution that will accurately differentiate between good grass seed and bad grass seed. The OSU Crop Science department are given batches of 25,000 seeds at a time. These batches are received from various clients such as golf courses, gardens, etc. When the Crop Science departments receive these seeds, there are off-types within those packets that need to be removed. The reasoning behind this is because planting those seeds with the off-types would result in undesirable plants being planted. The Crop Science departments job is to pull out all the off-type seeds by hand through visual inspection. The seeds are magnified and the analysts differentiate between the good seed and the bad seed by looking at different markers such as shape, texture, color, etc. This is an extremely time-consuming process that is tedious and causes eyestrain, headaches, and backaches.

Our job is to utilize a machine learning tool that can identify the off-type seeds as off-types so they can be removed from the batch. The challenge is that it has to be perfect accuracy as any trace of off-type seeds could cause contamination and the client would discover that they have unintentionally planted undesirable seeds. If we cannot achieve perfect accuracy, the analysts would have to go through the batch by hand, which would make using a program useless.

**CONTENTS**

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Definition and Description of Problem</b> | <b>2</b> |
| <b>2</b> | <b>Proposed Solution</b>                     | <b>2</b> |
| <b>3</b> | <b>Performance Metrics</b>                   | <b>2</b> |

## 1 DEFINITION AND DESCRIPTION OF PROBLEM

The OSU Crop Science Department takes batches of 25,000 grass seed at a time from clients. The department is paid to sort through the seed and separate the desired seed from the off-types, which will unavoidably be mixed in. The reasoning behind taking out these off-types is because a client such as a golf course owner will desire a specific type of grass seed to be planted. If the off-types are planted with the grass seed, it could potentially result in contamination and the spreading of weeds. The Crop Science Department has analysts that perform this sorting job by hand. They sit for hours a day and differentiate between the off-types and the desired seed by looking at texture, shape, and color with the aid of a high-powered lens. This is a tedious process that causes eyestrain, headaches, and backaches. It is also inefficient because it is not only slow but also prone to human error, especially after sifting through seeds for hours on end.

## 2 PROPOSED SOLUTION

Our proposed solution involves using a high-powered lens camera to gather up as many images as possible of the known good seed. We will then train these images utilizing a machine learning tool such as Tensorflow. The goal is to train about 100,000 images to start with. Next, we will test it on a batch of images and note down the accuracy percentage. The goal here is to get as close to 100 percent accuracy as possible. This means when the software takes in an image of desired grass seed as input, it should do nothing. When the software takes in an image of an off-type grass seed as input, it should flag it as off-type and remove it from the stack. Our client said they are only really concerned with differentiating between desired seed and off-type but if we could do that successfully, he said they would also appreciate if we could categorize the different off-types. This could be accomplished by feeding the algorithm a batch of images and categorizing the good seeds in a folder and the off types in a folder. From there, we would need to sort through and train the machine learning tool on those specific off-types. The challenge in this is that it might require a person to sift through all the off-type seeds and figure out, which ones are the same. That way we could categorize the off-types in different folders and train the machine learning tool to place images in their corresponding off-type location.

## 3 PERFORMANCE METRICS

A complete version of this project would ideally be able to identify seeds of the desired grass type in real-time or more accurately identify what is not the desired grass seed type at 100 percent accuracy. It is imperative we reach perfect accuracy because the Crop Science Department will not risk losing their business if there is even the slightest chance that an off-type can slip through the cracks and contaminate the rest of the seeds. Our client mentioned they would rather throw away good seed than let off-types contaminate the batch. This means our leeway for false positives won't be as rigorous as I previously thought.

In addition to having perfect accuracy, the software needs to be able to identify the seeds in real time. Our client told us a mechanical engineering group was working on creating a conveyor belt with a high-power lens attached that is connected to the computer and will feed images of the seeds as they go to the program. The program needs to accurately take those images and categorize them based on the seed being a desired type or off-type. Since this is happening in real time, the GPU will likely have to be strong so the program is fast enough to keep up with the conveyor belt. A signal is sent from the computer to the conveyor belt when an off-type is detected. That off-type will be separated from the desired seed and placed into its own bucket.

There are many potential challenges with this project. Getting 100 percent accuracy should theoretically be doable but the seeds can be obscure that could result in lower accuracy. It also needs to run quickly enough to identify the seeds in real-time. This could be a challenge and we may not have the hardware to be able to achieve this.