Algaeorithm

FINAL RESEARCH REPORT

WRITTEN BY
Ashwin Mukherjee
Rohan Chanani

PROJECT ADVISOR
Jacob J. Valenzuela, Ph.D.

Abstract

Algaeorithm aimed to add to an educational ecosystem that empowers the next generation of scientists and engineers in leveraging algae biotechnology to build a sustainable future from the perspective of today's high school students. The approach is anchored by a comprehensive web application (algaeorithm.com) that includes a custom-designed neural network tailored to analyze microalgae growth phenotypes. In addition, our web application also hosts original multimedia content explaining algae's potential to disrupt the oil industry and how machine learning works within the context of our algorithms, among other relevant topics. In tandem with our application and content, we developed a comprehensive lab protocol, which standardizes Algaeorithm's usage in a classroom environment. The protocols can be used by teachers or students and detail every step of an experiment. For example, we provide a detailed list of materials to purchase for the cultivation and sampling of microalgae, as well as an overview of manual or automated growth characterizations with Algaeorithm. When developing new curriculum-based content, it is important to have stakeholders (i.e., High School Students) involved to ensure its practicality and efficacy for potential users. The Algaeorithm multimedia resources and classroom guide reside alongside our core analytical tool within the web application, thereby creating technology-focused content for algae experimentation and education for the high school classroom. In the coming months, our team will begin integrating Algaeorithm into the existing Bioengineering a Sustainable World curricula developed at ISB to increase its access to classrooms across the country.

Table of Contents

List of Figures

Introduction

The grand challenge of the AlgaePrize competition is to develop novel solutions to algae production and processing that ultimately lower the cost of bioenergy products. Our team has taken an unconventional yet innovative approach to address this challenge. We believe that significant changes to the algal bioenergy and bioproduct industry require a generational shift in how contemporary students become critical thinkers in regard to a future bioeconomy. This will mean increased accessibility to bioenergy information and a deeper understanding of the potential tools required to handle complex problems. For instance, machine learning will play a major role in how scientists unlock biological complexities to solve tomorrow's problems. Therefore, we have focused our efforts on developing and deploying a tool, Algaeorithm, that is a lens for high school students and teachers to explore the possibilities of the bioenergy industry and see the vision of BETO's education mission. Specifically, our project directly aligns with OPERATION BioenergizeME's mission to support formal and informal education by engaging future scientists, engineers, and entrepreneurs about the bioenergy industry, while introducing how a powerful tool like machine learning can help solve the world's energy problems. In furtherance of this mission, the unifying theme of our project is the creation of technology-focused content for algae experimentation and education within a standard high school classroom. We believe our outside-the-box approach and student perspective can have a major impact on the human capital that will enter the university systems and future workforce that will drive the bioenergy industry of tomorrow.

The Algaeorithm project originated out of the curiosity of two high school students, Ashwin Mukherjee and Rohan Chanani, who took the initiative to seek information about the algal biofuel industry and how they could incorporate machine learning into a useful tool for researchers. They sought out the expertise of systems biologist, Dr. Jacob Valenzuela, who has extensive experience in characterizing the physiological state shifts between growth dynamics and lipid accumulation (Valenzuela et al. 2012, 2013; López García de Lomana et al. 2015; Imam et al. 2015). Dr. Valenzuela has also developed an algal-based curriculum, Bioengineering a Sustainable World (BSW), as part of the Institute for Systems Biology (ISB) Systems Education Experiences (SEE) program. The curriculum teaches students how bioengineering can address the world's bioenergy and bioproduct challenges. Before the AlgaePrize competition, Ashwin and Rohan led the development of a prototype application that uses machine learning to quantify cell concentrations from microscope images. In the synopsis, we outline how we aim to use Algaeorithm to engage students, educate them on the algal bioindustry, and introduce them to machine learning concepts.

Our project's novelty originated out of its vision to invest in human capital at the student level to achieve the generational shift required to disrupt the current petroleum-based economy. The original working prototype of the Algaeorithm application proved its technical feasibility, and we have since expanded its versatility and refined its accuracy as described in the synopsis. In order to convert the novelty and functionality of our application into a tangible impact in the classroom, we have created an expansive suite of instructional videos and resources, including a detailed protocol that allows teachers to complete research-caliber lab procedures in the classroom without a large budget.

Through our application development, content creation, and wet-lab research, AlgaePrize has allowed us to create technology-focused content directed at high school students and teachers for algae experimentation and education within a classroom environment.

Methodology

In furtherance of our overall goal to create a web application and technology-focused content for algae experimentation and education in the classroom, we separated our goals, metrics, and milestones into three stages: The development of the Algaeorithm application, wet-lab research to verify and improve the application's technical feasibility, and content-based educational outreach to interface our research with students. Along with our goals, we also outlined specific technical components, research procedures, and production plans to assist ourselves in meeting these goals.

Application Development

The Need

Students can learn much about fundamental biology from growing microalgae and develop potential hypotheses about how algae are affected by their environment and what that might mean in relation to bioenergy. By characterizing algal growth patterns in the classroom, students can test their hypotheses and flex their research muscles.

Typically, in classrooms, students can count cells with a hemocytometer and microscope. This is an opportunity for students to learn how to use a microscope, quantify algae, and visually observe the algae. However, teaching how to count cells and having students count cells themselves takes a lot of classroom time. We feel that time could be better spent exploring the possibilities of algae or testing multiple hypotheses. Therefore, we aimed to streamline the cell counting process by using machine learning algorithms to calculate cell densities from uploaded microscope images taken from a standard smartphone.

Neural Networks

To computationally extract useful data from uploaded images, we sought to include multiple convolutional neural networks in our final application, improving upon those already included

in the application prototype. A convolutional neural network sequentially applies linear transformations to inputs, in this case, microscope images of algae cells, and minimizes the difference between the predicted output and the actual output, in this case, the locations and classifications of the algae cells.

For the cell counting algorithm, we aimed to have our neural networks serve the following purposes:

- Locating bounding boxes for all of the cells in each image, with the coordinates of each box for visualization purposes as well as the number of boxes and areas of each box for calculation purposes
- Classifying the species of each located cell to provide information about the algal population and indicate possible invasions from external species
- Quantifying visual metrics from individual cells and the collective population, including chlorophyll level, lipid count, and overall algae health

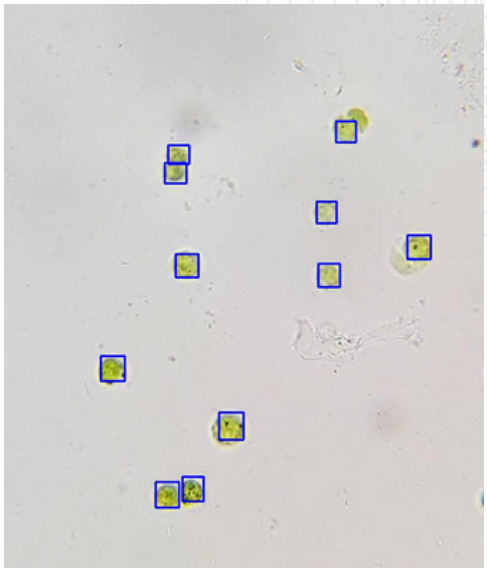


Figure 1. Neural network output

Our ultimate goal for the core algorithm was to use the number of pixels in the image, known measurements from the species classified in the image, and the sample depth to calculate the cell concentration in each image and return this concentration along with the corresponding cell count.

Analytical Processing

Our intended functionality for the application also included the computation of statistical metrics from the returned cell count and concentration to benchmark against existing data, which will come from both the Algaeorithm database of user-uploaded photos and from external databases. With time-sensitive data, the application's predictive models aimed to forecast growth curves and output metrics about the macrostate of the population, such as carrying capacity and growth rate.

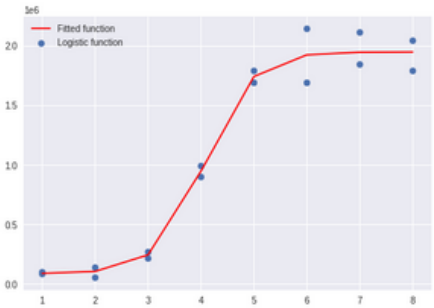
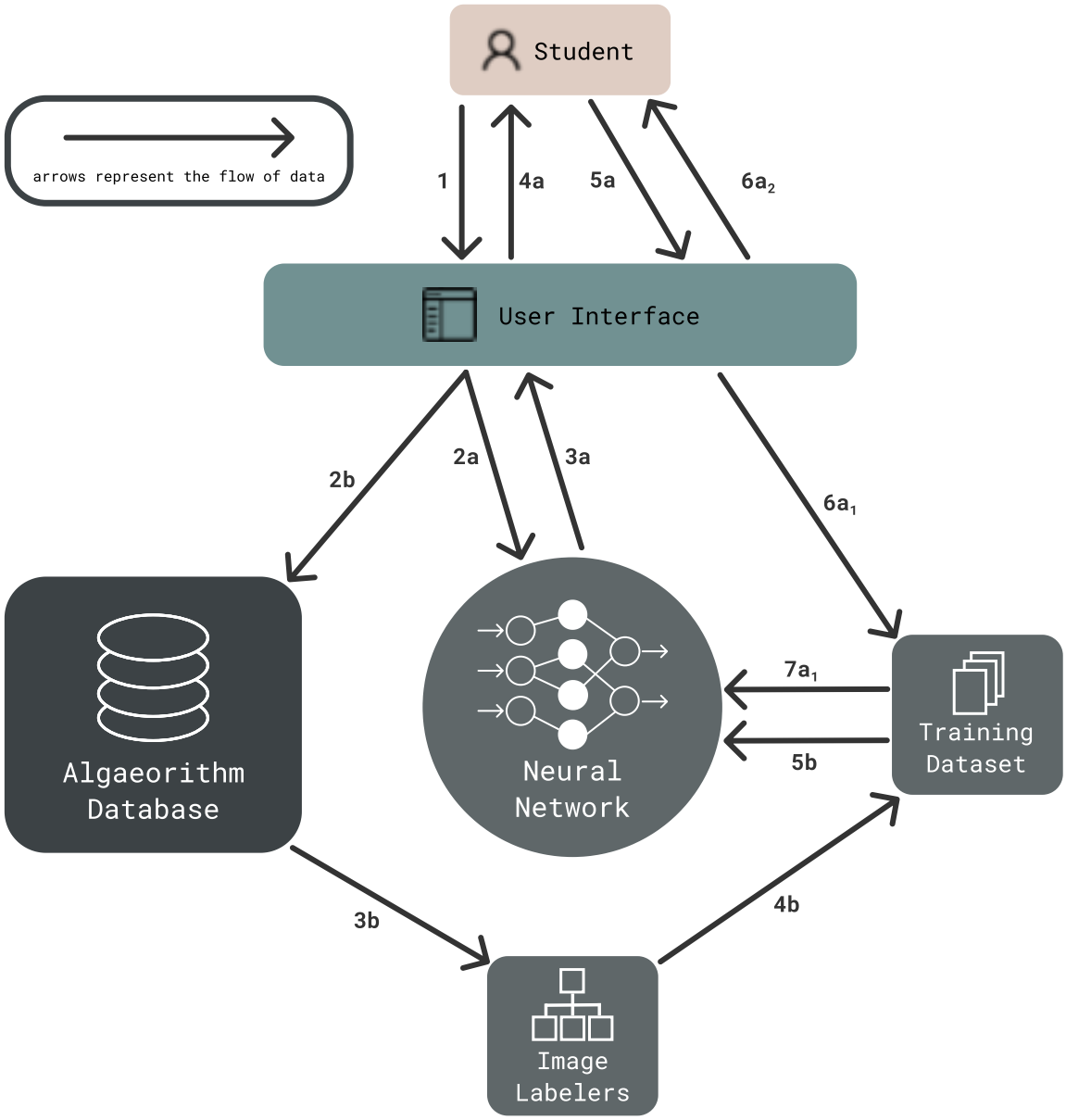


Figure 2. Sample growth curve

The proof-of-concept Algaeorithm application could potentially be extended to quantify other phenotypes associated with morphotypes that can be visually assessed with photos, i.e., chlorophyll content, lipid accumulation, and photosynthetic efficiency to provide a clear picture of the algae population state.

Technical Overview



Process Flow Diagram 1. Data Pipeline: After the user uploads an image (1), two pathways are initiated (a & b). First, the image is sent to the neural network (2a), which returns the corresponding output through the interface (3a) to the user (4a). The user can then modify the results and resubmit the edited output to the application (5a) to be used to both provide the user with more accurate information about their samples (6a2) and increase the training dataset (6a1), which will be used to retrain and improve the neural network (7a1). Second, the image is uploaded to the database (2b), where it will be made accessible to image labelers (3b). Each labeled image will then be added to the training dataset (4b), which will be used to once again retrain the neural network (5b).

Data Pipeline

In order to streamline the collection and analysis of data, we aimed to create an end-to-end data pipeline to accelerate the development of our application alongside its increasing usage. As soon as images entered our application, we planned to upload them to a central database where they would be labeled for training or compared with future user data.

User Interface

Through our continued development of the application, we aimed to concisely display the information computed by the algorithm through the following aspects of the user interface:

- Distribution graphs, such as box plots and line plots, to visualize and replicate data
- Growth curves to visualize time-sensitive data in multiple forms, including linear, exponential, and logistic growth
- For users that want to visualize their data, we also generate a data table, which can be downloaded as a CSV file for further analysis
- Specific algorithmic output for each image, with bounding boxes around the identified cells and labels for the corresponding species and visual metrics

We intended for the bounding boxes and labels on each image to be adjustable to allow users to manually improve the accuracy of their output data and to create additional training data for improving the core algorithm and neural networks. We also intended for users to be able to compare their images and statistics with the application's broad database of results

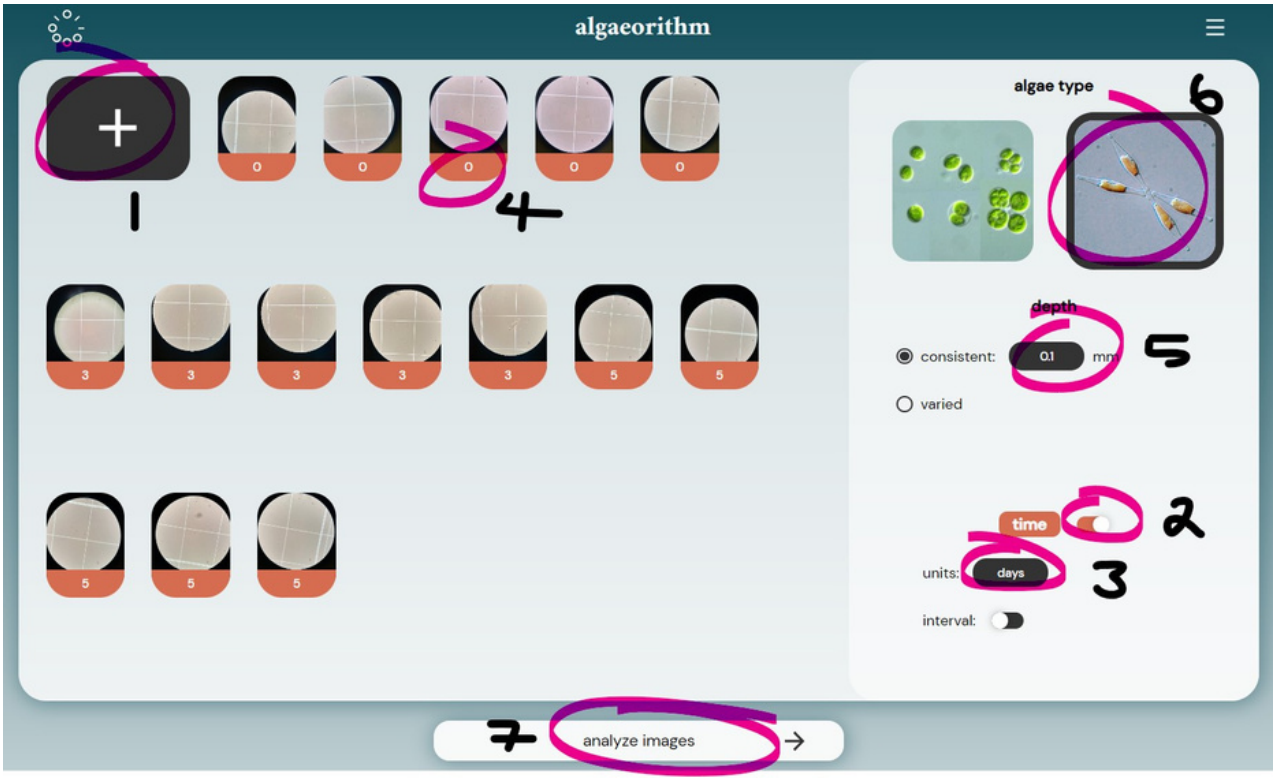


Figure 3. Algaeorithm User Interface: Students can upload images (1), toggle time-sensitivity (2), customize units (3), input time series (4), manually enter each sample depth (5), toggle algae species (6), and submit this data to the application (7).

Wet-lab Research

To test and improve Algaeorithm and its core algorithms, we aimed to cultivate algae and compare manually calculated results with the output of the application. Thus, we spent 2.5 weeks at the Institute for Systems Biology where we used large 1.8L photobioreactors to grow algae. We learned how to set up, sample, and quantify growth phenotypes from these custom reactors. The experience we gained from this authentic laboratory setting, directly fed into our development of classroom protocols.

Using the green microalgae *Chlamydomonas reinhardtii* and diatom *Phaeodactylum tricornutum* as sample species, our goal was to test the performance of the application with various types of microalgae. Algae cultivation required controlling the environment in which the algae grow and collecting data from the algal samples, which would be used to retrain the neural network.

Figure 4. Algae cultivation and analysis at ISB

Alongside our algae cultivation, data collection, and verification testing, we planned to create a comprehensive laboratory protocol that would allow for similar experiments and procedures in a high school classroom. At each step of our experimentation process, we aimed to document the materials and procedures required to reproduce similar results while also allowing for classroom creativity. The iterative process of cultivating algae, collecting and recording data, comparing the outcome with our intended results, and adjusting the existing instructions to improve the overall protocol, we were able to generate a Classroom Guide for constructing mini-photobioreactors for microalgae (Appendix B)

Education Outreach

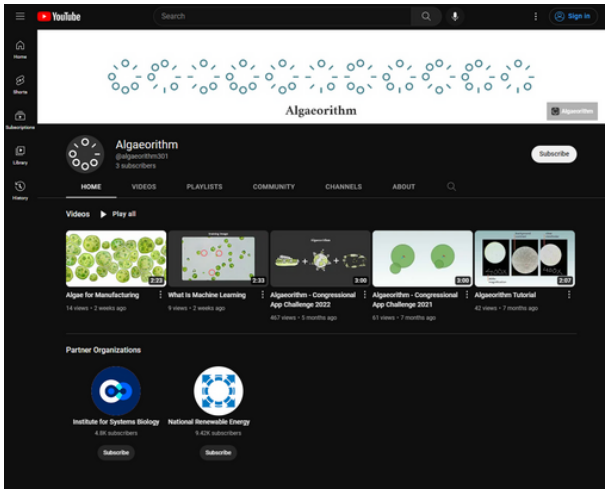


Figure 5. Algaeorithm's YouTube channel

To the best of our knowledge, there is not a lot of content on the intersection of algal bioenergy or bioproducts and machine learning for high school students. In comparison with nuclear, solar, and wind energy, very few students are familiar with bioenergy, and most students know little about machine learning beyond digital assistants and pop culture depictions of artificial intelligence. Our project aimed to use our unique perspective as high school students to help remedy these issues with

the following initiatives along with the actual Algaeorithm tool:

- Video tutorials and demonstrations that allow middle and high school students of all experience levels to use our application and gain valuable bioenergy experience working with algae
- Educational content about the basics of machine learning and how we use it in our application to expose students to this field
- A live, interactive workshop with students and teachers to teach them how to use the application in the classroom
- If applicable, collaboration with the BioenergizeME team to integrate some of Algaeorithm's proposed content into BETO's Bioenergy Education Resources

Challenges

To ensure our algorithm is consistently accurate, we needed to minimize any variation between the types of photos submitted by students. One thing our team quickly realized while we were taking our own pictures was everyone's hands, phones, microscopes, and environments contribute to wildly different images. Often, the differences in color, lighting, or even clarity lead to meaningfully less accurate analytics. We then developed multiple strategies to address this issue. First, a low-cost attachment for phones to clip into the microscope lens and eliminate shaky pictures. Second, a calibration microscope slide that includes standardized size and color markings. Students could submit a picture of this slide using their setup in order for the algorithm to calibrate accordingly and account for differences in color and lighting. The size markings are another way we could improve accuracy, especially in the context of cell concentration.

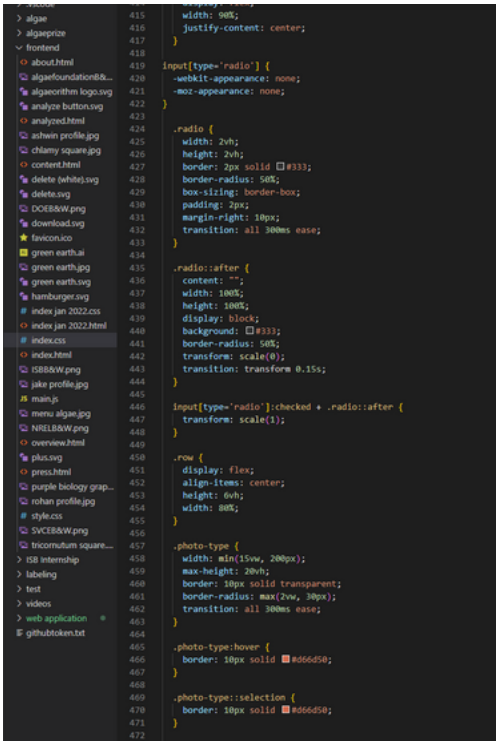


Figure 6. Subset of the Algaeorithm codebase

In order to create the data pipeline, a major challenge was creating smooth transitions between platforms in order to transport the images in an efficient manner. We largely addressed this challenge by using cross-platform tools like boto3, but the cost of proprietary services like S3 and Colab remains somewhat limiting.

For the algorithm, a major challenge was tuning the model's hyperparameters to prevent it from diverging during training or overfitting to the training dataset. In order to address this challenge, we iteratively completed several training batches with adjustments in between, but as our training dataset changes, we will need to continue optimizing these criteria.

One of the biggest challenges is managing time; our team consists of two high school students with obligations in school, extracurricular commitments, and college applications, all while being in three different states. We decided that during this busy period, we would focus on the "scaffolding" of our deliverables: creating outlines and drafts for new content that we can develop in the coming months. Moreover, we leaned into our past experiences creating videos to streamline the development process.

We quickly realized the challenges of growing algae in the classroom as it is not like the highly-equipped labs at the Institute for Systems Biology. Algaeorithm itself is complete, with room for improvement on the front end, particularly instructional content. It's important that we thoroughly test Algaeorithm, including the pre-analysis steps of algae cultivation and sampling, from the perspective (and equipment availability) of classrooms. To do so, our team has compiled a list of the relevant equipment for the mini-photobioreactors similar to those utilized by ISB, along with other supplies needed for algae cultivation.

Figure 7. Mini photobioreactors

Results

Including the challenges listed above, we ultimately achieved most of our goals in technical development, wet-lab research, and content creation.

Technical Development

While working at ISB over the summer, we successfully created a streamlined data pipeline, which now allows us to go from data collection to training the model as quickly as possible. Using the AWS Sagemaker Ground Truth tool for labeling images as well as a neural network architecture built from scratch using the python libraries Jax and Flax, we made it easy to efficiently label images and add them to our compiled training dataset. This way, we'll be able to quickly use the data we collected to make swift improvements to our algorithm. Not only can we quickly add additional labeled images of existing species to our dataset, but we'll also be able to add new species much more efficiently.

From the front-end perspective, we were able to design a layout for inputting algae sample data that we think will be extremely useful for students, particularly those without prior lab experience. This new layout emphasizes familiar geometric components and colors to curate a more simple user experience for students. This is just one feature in the broader frontend rollout that also includes new pages that host educational content and our team's mission.

For the core algorithm, we used Google Research's Jax and Flax libraries to create a custom implementation of the You Only Look Once version 1 (Yolov1) object detection neural network. Jax is specifically designed for accelerated computing on graphic processing units (GPUs) and tensor processing units (TPUs), although it runs rapidly on central processing units (CPUs) as well. It can compute graphs of linear algebra computations on matrices at incredibly fast speeds, which makes it very useful for building and training neural networks. Flax is a neural network library built on top of Jax that is designed for flexibility in creating models and takes advantage of Jax's accelerated capabilities. Yolov1 also prioritizes speed, and it completes both training and prediction at a rapid rate compared to other object detection architectures while retaining a similar level of accuracy. Because our application is designed to process a large volume of algae images for data extraction, these traits made the Yolov1 architecture ideal for our use case.

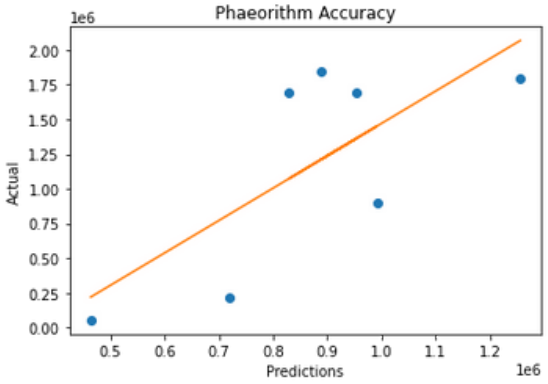


Figure 8. Phaeorithm Accuracy: Correlation based accuracy metric for *Phaeodactylum tricornutum*.

We also built a novel data pipeline to make it as easy as possible to transition from data collection to data analysis. The pipeline begins with images taken by a mobile phone through a microscope lens. These images are then uploaded into a publicly available Google Drive folder. Once the images are in the folder, we use Google Colaboratory to preprocess the images and upload them into an AWS S3 bucket. Once the images are in the bucket, we use AWS Sagemaker's seamless integration to directly label the images, thereby generating the aforementioned training dataset. This dataset can then be accessed from Google Colaboratory and used to train the Yolov1 model, which improves our core algorithm.

Our implementation achieved over 90% mean average precision on training datasets from both Pascal Visual Object Classification (VOC) and *Phaeodactylum* images we collected ourselves, which is a significant improvement over the neural network we were using previously. Both the labeled images produced by the model and a compilation of annotations from the model demonstrate the success we were able to achieve with the new architecture.

Using the new data pipeline, we were able to label over 1000 images of *Phaeodactylum tricornutum* cells, which allowed us to create a substantial training dataset of images to further refine the model.

Due to the increased convenience as a result of the data pipeline and the reduced training/prediction time as a result of the new algorithm, our overall iteration time has been reduced by over 50%.

Content Creation

After curating a list of equipment for classroom activities, we consolidated this information, as well as assembly and setup instructions, into an initial draft for classroom testing of our laboratory protocol. The draft allowed us to iteratively change the protocol as we tested it ourselves and to receive feedback from teachers and students.

We also created instructive and expository video content to integrate into the application itself. The videos are as follows:

- "What is Machine Learning" (explained using our application's core deep learning functionality)
- "Why Algae is a Sustainable Alternative to Petroleum for Disrupting Oil/Gas"
- Algaeorithm Tutorial

This combination of general videos related to the biofuel and machine learning industry as a whole as well as topics specific to Algaeorithm itself ensures that using our application and content will prepare students for careers in STEM fields.

Our team submitted Algaeorithm to the Congressional App Challenge, where Congresswoman Anna Eshoo (CA-16) awarded Algaeorithm Special Congressional Recognition.

Wet-lab Research

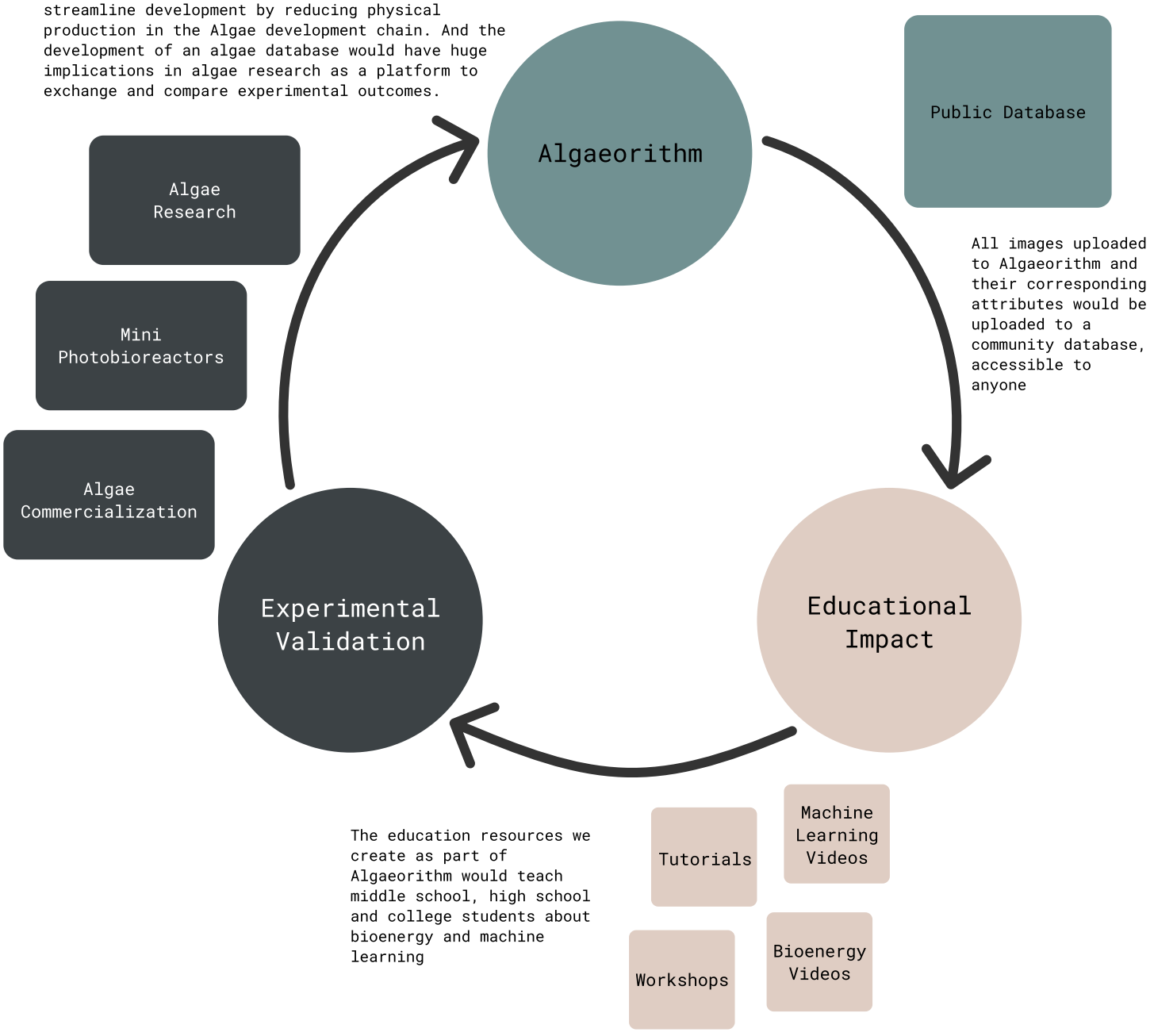
Working in the Baliga lab, we collected an array of data relevant to our project. First and foremost, we began growing *Phaeodactylum tricornutum* across two large custom photo-bioreactors. We went through the entire growth process, which included the collection of samples for microscopy, characterizations of environmental parameters (e.g., pH and nutrient profiles), and photosynthetic efficiency.

Given that the central feature of our application is the image-analysis tool, we wanted to obtain as many algae images as possible. We're also targeting students in classrooms, so we tried to mimic accurate conditions; we used microscopes available in most high schools, took pictures directly from our phones, and included many blurry and otherwise distorted images. These images were then added to our training dataset, labeled, and used to further refine our algorithm.

The overall lab procedures were also extremely helpful for our team to identify additional opportunities for our application. Beyond the calculation of cell concentration using a hemocytometer, our sample analysis included measuring quantum yield and fluorescence. These measurements are something we hope our algorithm can compute in the future through relative comparison and color analysis.

Project Framework

With the help of experimental validation with mini photobioreactors, Algaeorithm could be expanded for use in Algae Research and Commercialization. More accurate predictive models could be used to streamline development by reducing physical production in the Algae development chain. And the development of an algae database would have huge implications in algae research as a platform to exchange and compare experimental outcomes.



Process Flow Diagram 2. Value Creation Cycle: The project consists of three elements, which together form a robust cycle. The application itself will contribute to educational impact through its use by students and teachers. Students and teachers using the application will contribute to the experimental validation by providing images for the training dataset. Finally, the experimental validation will contribute to the application by improving the algorithm with additional data. As depicted in the diagram, each element will also generate various offshoots which add to the algae value chain.

Throughout our time at ISB, we recorded cell concentration, quantum yield, and fluorescence. Doing so allowed us to plot these values over time and compare them to the growth models generated by Algaeorithm. Another feature we hope to eventually implement is a far more robust growth model that also relies on a neural network for more accurate predictions.

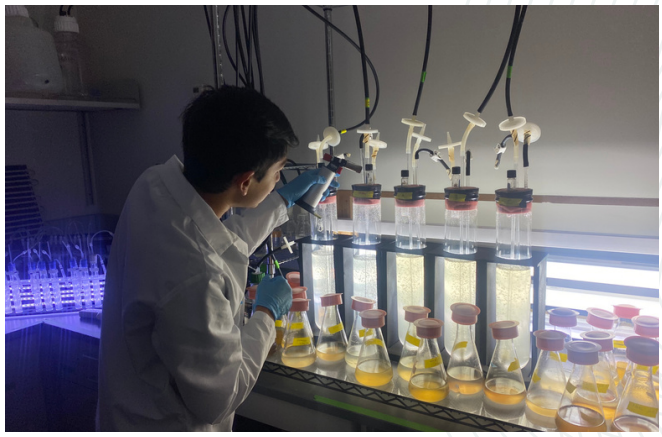


Figure 9. Sampling from bioreactors at ISB

We also performed a Nile Red assay over the course of our time at ISB, which allowed us to measure lipid accumulation as a result of nitrogen starvation. Because lipid droplets can also be observed in a microscope image of algae cells, we hope to eventually be able to use the data we collected to measure the accumulation of lipids in a microscope image. This is an extremely relevant metric for the study of algae in biofuel research, so we would love to bring this to students as well.

Deviations

Technical Deviations

We primarily deviated from our technical goals in the process of adding additional features beyond cell count and cell density.

Although we successfully added accurate analysis for an additional species (*Phaeodactylum tricornutum*), we did not add a classifier to differentiate between multiple species in the same image. While working in the lab, we found that classrooms almost always worked with samples containing only one species, and the loss in accuracy resulting from the addition of more than one species in the same image outweighed the benefit of using mixed samples. Instead, we added an option on the application to manually toggle between species, which allows for a seamless alternation between different species.

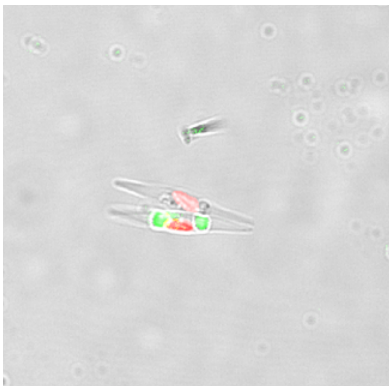


Figure 10. *Phaeodactylum tricornutum* lipid content

Through our manual collection of data for fluorescence, lipid content, and photosynthetic yield, we found that ascertaining this information from a microscope image is not currently technically feasible with the technology available in a classroom. This type of analysis will take more work piloting how this can be achieved, so we decided to shift our focus to metrics that could be determined visually.

Content Deviations

Due to the logistical challenges of working directly with students, we shifted our educational content focus to the creation of standalone material, which meant temporary sessions with students will have to wait. Through the process of scheduling and conducting pilot testing of our application and potential workshops with classrooms across the country, we found that the most effective use of our time and resources was to rely on existing student infrastructure for the last-mile delivery of our content and technology and to focus on the creation and improvement of the application itself instead.

Discussion & Next Steps

Commercial Applicability

Algaeorithm expands access to and awareness of the bioeconomy, enabling collaborative academic endeavors and drastically reducing R&D resource usage. We anticipate the largest impacts in education by potentially allowing thousands of students to participate in real-world algae research at the high school level. Additionally, students would be part of a rich network that could use our application as an organizational framework to collaborate on large-scale projects. In ISB's words, these projects help "bridge the distance between the professional lab and the classroom."

In aiming to achieve cost competitiveness with fossil fuel markets, far more students will need to become engaged with bioenergy and develop career-relevant experience. Our project contributes to this labor pool by introducing students to the field of bioenergy and providing them with the knowledge, ability, and interest necessary to see the potential of a bioenergy career while they are still in high school and developing their interests (Figure 8). From industries ranging from agriculture to financial services, machine learning models have helped create more lean processes that increase profitability and cost competitiveness.

Our concentration on education allows our team's work to naturally incorporate diversity, equity, and inclusion. Despite its clear benefits, scientific research has traditionally been off-limits to millions of students across the country, excluding them from countless educational benefits. By specifically targeting high school classrooms across the country, our application fosters a demographic makeup of users that more accurately reflects the diverse makeup of the nation.

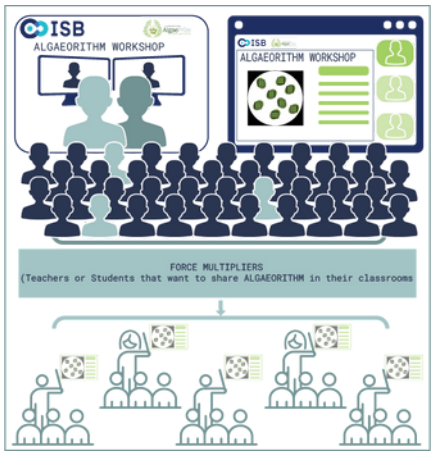


Figure 11. Algaeorithm Student Impact: Algaeorithm creates force multipliers who can share this information with their communities.

Lessons Learned

Over the nearly 2 years our team has spent working on Algaeorithm, we've had successes and failures and learned from them every time. For one, our efforts achieve the highest ROI when our team's working on software. This isn't totally surprising; coming into the project, Rohan's background was in data science and machine learning while Ashwin had experience in design and content creation. We figured out very early in the project that having Rohan focus on algorithmic improvements and backend mechanics in parallel with Ashwin developing a frontend and educational content resulted in our most efficient progress. As such, the software component of our project (algaeorithm.com) is the most ironed out and will have the most long-lasting use by students. As we tried to pivot to hardware and outreach, we found that, with our small team of two busy high schoolers, it was difficult to make progress toward some of our more ambitious goals.

For example, in our content creation endeavors, putting out one original video can demand dozens of hours stretched over weeks. However, we were still able to create five videos, each 2-4 minutes long, that explain why we developed algaeorithm, and how it works, among other brief explainers on bioenergy and machine learning. Though these videos touch on important ideas related to algae biotechnology, with more time we may have been able to create additional content or refine some of our work more.

Venturing into the education outreach aspects of our project, we found similar difficulties. In testing our application, we were able to iron out many of the issues in our code through self-tests but knew that to truly create a fluid user experience we needed beta-testing in a classroom environment. So, with help from the Institute for Systems Biology (ISB), our application was rolled out in two different Washington high schools. Given the progressive loop of Algaeorithm's neural networks, the data we received from these tests were tremendously helpful in improving our algorithms, specifically for low-resolution images. However, given the size of our team and our locations in California and Georgia, and couldn't visit the high schools in person, it was difficult to observe student use and thus identify shortcomings in the user experience. In addition, we found it tricky to communicate with teachers about how the experience was progressing, and potential strategies for implementing our tool into existing curricula.

From this experience, we derived the second important lesson of our project: leave the last-mile delivery to experts. In our case, the "last-mile delivery" consists of getting the tools and content we produce (i.e., Algaeorithm web application, videos, and protocol) into the hands of students. We simply ran out of time and could. As students, we appreciate the importance of this step; often, the way in which content and activities are delivered in class can make or break the enjoyment and ultimate pursuit of a specific topic. In recognizing this, we've partnered with the seasoned professionals at ISB's Systems Education Experiences (SEE)

program. The SEE program has extensive teacher networks across the country and has both created novel science curricula and integrated concepts into existing systems used by countless students. Therefore, moving forward our team will partner with SEE to hopefully get Algaeorithm into 100 classrooms by the year 2025.

On a broader level, managing our project as full-time students honed our time management and project management skills. At the beginning of our project, we had lots of ideas about the myriad analytical tools, student-focused features, and intricate designs we could build into our application. However, we needed to be realistic about what we could achieve. This ultimately led to our parallel development approach where each of our two team members worked on different "halves" of a component and then, every week or so, collaborated to assemble the components. Furthermore, we learned the necessity of proposal writing and managing budgets in relation to scientific projects. We had no grasp of the amount of time dedicated to writing about our work in a high-level scientific competition like AlgaePrize. Additionally, we needed to compile expenses whenever we raised money from external sources like Silicon Valley Clean Energy. Competing in AlgaePrize has provided invaluable experiences in both these areas and we're grateful for how much we've learned outside the core skills needed to build Algaeorithm.

Reflecting on our work and the lessons we've learned along the way, it makes most sense for our team to continue focusing on end-to-end and vertical integration of algae research tools within classroom environments. The existing application has already merged a discrete set of time-consuming tasks into one continuous process. After imaging algae samples, Algaeorithm swiftly moves through counting cell concentrations, modeling growth curves of best fit, and providing macro-level statistic analysis between samples. We'd like to continue building out features that cut out the number of separate processes necessary for classroom science research and education. For one, we hope to create Algaeorithm apps for both iOS and Android platforms. This cuts out the photo-transfer step, enables algae research for the 93% of high schoolers with smartphones, and allows us to capitalize on built-in cameras and sensors to provide new analytical insights. We also plan on further developing our protocol to provide students with standardized experiments and data. These sustained efforts, along with the work we've already completed, are helping us cultivate an Algaeorithm ecosystem of tools, content, and guides that work together in introducing algae biotechnology to students across the country.

Our team is immensely proud of the work we churned out over the course of the AlgaePrize competition, and we're optimistic for Algaeorithm's ability to make a meaningful impact on the next generation of scientists who will help lead the US into a clean future.

Citations

Gottschalk, U. (2011). *Downstream processing*. Downstream Processing - an overview | ScienceDirect Topics. Retrieved April 4, 2022, from https://www.sciencedirect.com/topics/engineering/downstream-processing

Institute for Systems Biology. (2022, March 2). *Home*. Baliga Systems Education Experiences. Retrieved April 4, 2022, from https://see.isbscience.org/

Institute for Systems Biology. (2021, May 28). *Curriculum modules*. Baliga Systems Education Experiences. Retrieved April 4, 2022, from https://see.isbscience.org/modules/

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: A System for {Large-Scale} Machine Learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).

Pew Research Center. (2021, November 23). *Mobile fact sheet*. Pew Research Center: Internet, Science & Tech. Retrieved April 4, 2022, from https://www.pewresearch.org/internet/fact-sheet/mobile/

Pew Research Center. (2021, November 23). *Internet/broadband fact sheet*. Pew Research Center: Internet, Science & Tech. Retrieved April 4, 2022, from https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Renotte, N. [Nicholas Renotte]. (2021, April 9). *Tensorflow Object Detection in 5 Hours with Python | Full Course with 3 Projects* [Video]. YouTube.

Valenzuela, J., Mazurie, A., Carlson, R. P., Gerlach, R., Cooksey, K. E., Peyton, B. M., & Fields, M. W. (2012). Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum. Biotechnology for biofuels, 5(1), 1-17.

Valenzuela, J., Carlson, R. P., Gerlach, R., Cooksey, K., Peyton, B. M., Bothner, B., & Fields, M. W. (2013). Nutrient resupplementation arrests bio-oil accumulation in Phaeodactylum tricornutum. *Applied Microbiology and Biotechnology*, *97*(15), 7049-7059.

López García de Lomana, A., Schäuble, S., Valenzuela, J., Imam, S., Carter, W., Bilgin, D. D., ... & Baliga, N. S. (2015). Transcriptional program for nitrogen starvation-induced lipid accumulation in Chlamydomonas reinhardtii. *Biotechnology for biofuels*, *8*(1), 1-18.

Imam, S., Schäuble, S., Valenzuela, J., López García de Lomana, A., Carter, W., Price, N. D., & Baliga, N. S. (2015). A refined genome-scale reconstruction of Chlamydomonas metabolism provides a platform for systems-level analyses. *The Plant Journal*, *84*(6), 1239-1256.

Acknowledgments

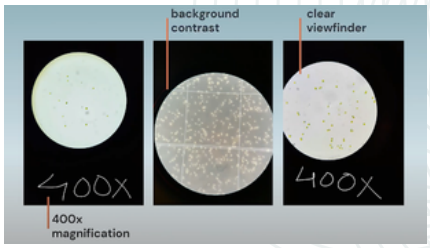
We extend our appreciation to the Institute for Systems Biology, Silicon Valley Clean Energy, the National Science Foundation, the U.S. Department of Energy Bioenergy Technologies Office, the Algae Foundation, and the National Renewable Energy Laboratory for supporting Algaeorithm throughout the 2022-2023 AlgaePrize competition.

Appendix A

Multimedia Resources

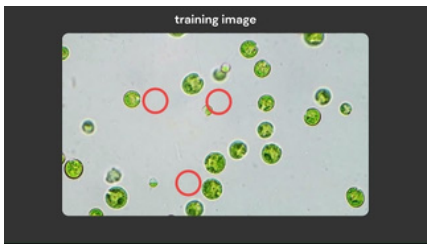
Algaeorithm Tutorial

A tutorial on how to use Algaeorithm to upload & analyze images of *Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum* algae.



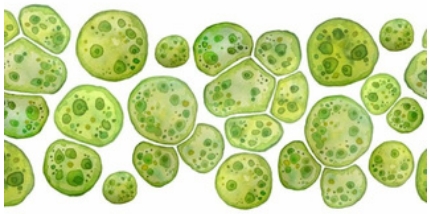
What Is Machine Learning?

A simple explanation of what machine learning is and how it works, explained through the lens of Algaeorithm.



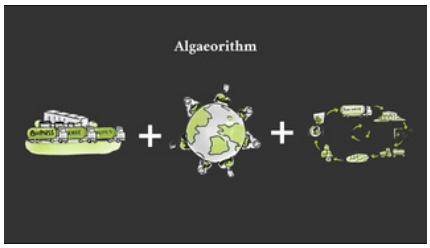
Algae for Manufacturing

A brief introduction to algae's potential for manufacturing, from biofuels to cosmetics.



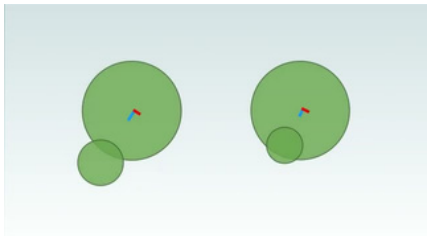
2022 Explainer

A demo of the Algaeorithm web application we made for the 2022 Congressional App Challenge.



2021 Explainer

A demo of the Algaeorithm web application we made for the 2021 Congressional App Challenge.



Appendix B

Classroom Guide

Materials List

1. Microscope (1ct.)
2. Hemocytometer (1ct.)
3. Glass Tubes (12ct.)
4. Air Pump (1ct., 2 output tubes)
5. Concentrated Media (1ct., makes 25 liters)
6. Freshwater Media (1ct., makes 1 quart)
7. Chlamydomonas, living (1ct., enough for 30 students)
8. Tube Rack (1 ct., holds 6 glass tubes)
9. Gang Valves (1ct., 5 output tubes)

1. Prepare photobioreactors

a. As shown in Figure B1, place the glass tube rack (Item 8) and UNPLUGGED aquarium pump (Item 4) on a table or flat surface where they will not be disturbed

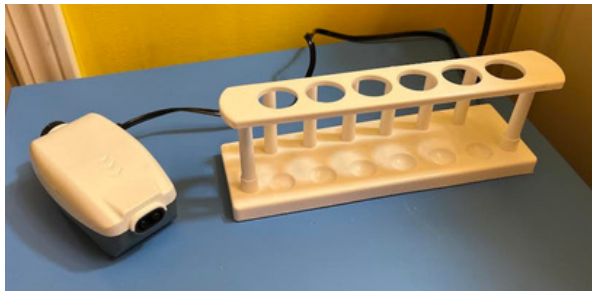


Figure B1. Aquarium pump and test tube rack

b. Repeat the following procedure for each reactor you plan to prepare:

   i. Remove the cork from one glass tube (Item 1) and set the tube aside.

   ii. Drill two holes in the cork using a drill bit. The first should be at least 1/4" in diameter, and the second should be equal or lesser in size. Make sure that the holes are separate from each other and separate from the edge of the cork. The completed cork should look similar to Figure B2.

Figure B2. Completed test tube cork

   iii. Thread a section of aquarium tubing through the larger opening in the cork. Do not cut the tubing until you are sure it will be long enough to connect the pump to the reactor. As shown in Figure B3, the amount of tubing exiting the cork on the bottom should be less than an inch from the bottom of the tube when the cork is placed back onto the tube.

Figure B3. Tubing length for bioreactor

   iv. Remove the cork with tubing from the test tube and set it aside.

v. Using the dropping pipette provided with the hemocytometer (Item 2), pipette approximately 35 mL of the growth media (Item 5 or 6) and approximately 6 mL of the Chlamydomonas culture (Item 7) into the glass tube.
vi. Place the cork with tubing back onto the glass tube so that the tubing is within an inch of the glass tube's bottom and place the entire apparatus onto the glass tube rack (Item 8)
vii. Connect the tubing to the pump using one of the following methods
    1. If using gang valves (Item 9), connect the open end of the tubing to one of the gang valves and use another section of tubing to connect the gang valves to the aquarium pump (Item 4). See Figure B4 for reference.
    2. If not using gang valves, directly connect the open end of the tubing to the aquarium pump output as shown in Figure B5.
viii. Plug in the aquarium pump and adjust the aquarium pump output until the reactor is bubbling gently.

Figure B4. Pump to bioreactor connection

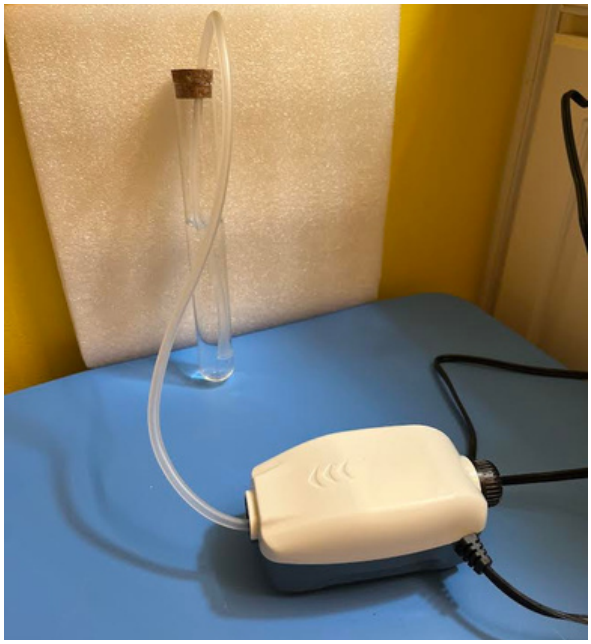


Figure B5. Bioreactor setup with gang valves

Notes: If you are experimenting with varying light levels or would like to provide light 24/7, you can use a desk lamp or similar light source with the apparatus as shown in Figure B6. Similarly, you can adjust the pump output to investigate the effects of changing the air supply.

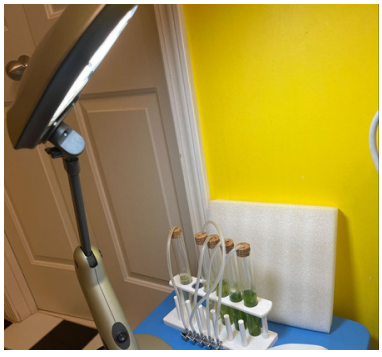


Figure B6. Bioreactors with desk lamp

2. Sample photobioreactors + prepare hemocytometer

a. Combine 3 parts isopropyl alcohol and 7 parts water in a spray bottle.
b. Clean the glass section of the hemocytometer and 1 hemocytometer coverslip using the sheets provided with the hemocytometer.
c. Place the coverslip onto the counting section of the hemocytometer.
d. Unplug the aquarium pump.
e. Carefully remove the top from one photobioreactor and use the pipette to pick up approximately 10 uL or 1 drop.
f. Pipette approximately 10 uL into one side of the hemocytometer as shown in Figure B7. The drop should disperse evenly under the coverslip.
g. Set the hemocytometer aside and replace the top of the photobioreactor.

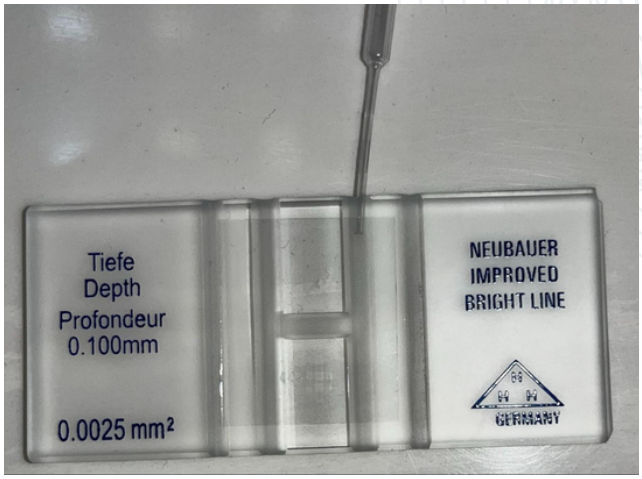


Figure B7. Dropping a sample into the hemocytometer

Notes: You can observe two samples on the same hemocytometer by using different samples on either side of the hemocytometer. To ensure that you do not "flood" the hemocytometer, ensure that there is only enough liquid to cover the observation area of the hemocytometer and that none leaks into the wells on either side.

3. Analyze cells

a. Place the 10x eyepieces into the binocular ports of the microscope. (See Figure B8)
b. Rotate the objective lenses until the 10x objective lens is directed at the observation area of the microscope.
c. Use the coarse focus knob to lower the observation area as much as possible.
d. Securely place the hemocytometer in the observation area of the microscope as shown in Figure B9. Turn the microscope on and adjust the light to the dimmest setting.
e. Adjust the binocular eyepieces until you see one image while looking through them.

Figure B8. Microscope 10x eyepieces

Figure B9. Hemocytometer on microscope

f. While looking through the binocular eyepieces, slowly raise the observation area using the coarse focus knob until an image forms. Once an image forms, use the fine focus knob to clarify the image. The image should look similar to Figure B10.
g. Rotate the objective lens to 40x, ensuring that the objective and the slide DO NOT TOUCH. Use the fine focus knob to refocus the image, which should look similar to Figure B11.
h. Collect data
    i. If collecting manually, follow the instructions included in the hemocytometer kit to calculate cell concentration.
    ii. If collecting automatically through Algaeorithm, follow the Algaeorithm tutorial to capture/upload images and extract cell data.

Figure B10. Microscope image (10x)

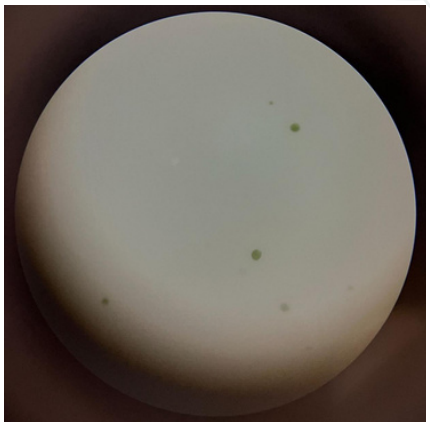


Figure B11. Microscope image (40x)

Notes: If fewer than 10 cells are visible when using the 40x objective lens, you can use the 10x objective lens to collect data once a clear image has formed.