Trinity

Nittany AI Challenge 2023
Prototype Phase

Team Members

Gayatri Bangar
Penn State Great Valley
(610) 202-0763
gdb5237@psu.edu

Ranojoy Deb
Penn State Great Valley
(484) 215-3564
rzd5485@psu.edu

Jason Durrance
Penn State Great Valley
(760) 453-0108
jbd5927@psu.edu

Mentor
Dr. Youakim Badr, Associate Professor of Data Analytics, Penn State Great Valley, yzb61@psu.edu

Industry Expertise
William Teodecki, Corporate Automation Engineer, Aqua America, wbt5083@psu.edu
Leo Daiuto, Director of the Great Valley Launchbox & Entrepreneur-in-Residence, Penn State Great Valley, lad96@psu.edu

Contents

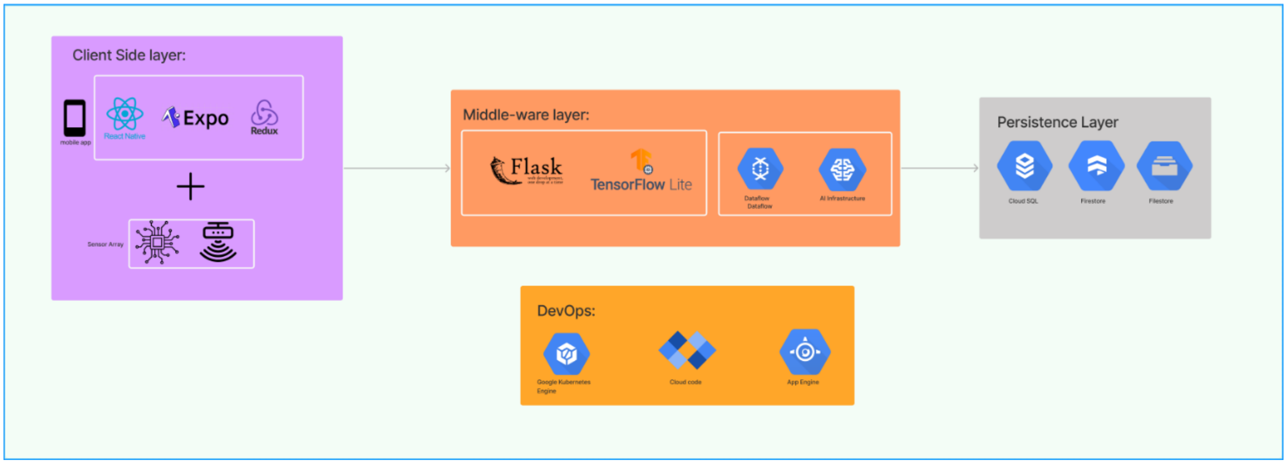
PROBLEM DESCRIPTION

Access to reliable sources of potable water is a significant problem globally as well as within large swaths of the United States. According to the Natural Resources Defense Council (NRDC), more than 170 million Americans are served by community water systems that have violated federal health-based standards in the past decade. Meaning that they have been exposed to unsafe levels of contaminants such as lead, arsenic, and bacteria. Additionally, the Environmental Protection Agency (EPA) estimates that more than 1 in 10 Americans get their drinking water from a system that is in violation of the Safe Drinking Water Act. Globally, the problem of water potability is even more acute. According to the World Health Organization (WHO), over 2 billion people lack access to safe drinking water. This is particularly prevalent in developing countries, where inadequate infrastructure and lack of resources make it difficult to provide clean water to large portions of the population.

The following facts serve to further emphasize the extent of the problem:

- Annually more than 3.5 million deaths are attributed to water-related diseases; the majority of whom are children.
- As per the CDC's estimates 1 in 44 Americans get sick from waterborne diseases, resulting in $3.33 billion in direct healthcare costs.
- Every $1 invested in water and sanitation provides a $4 economic return from just lower healthcare costs; more productivity and fewer premature deaths.

PROPOSED DESIGN SOLUTION



Predictive Framework Creation: Two deep learning models will be built using proprietary datasets. The first model will use Long Short-Term Memory Recurrent Neural Networks (LSTM RNNs) to protect certain water quality parameters that comply with WHO standards. The second model will use Vision Transformers (ViT) to classify the presence of waterborne pathogenic microorganisms, harmful water particulates, and turbidity in microscopy images. Trinity offers an innovative solution for water potability testing by combining a user-friendly mobile app, advanced deep learning models, portable IoT hardware, and secure cloud support. This comprehensive system empowers users to test water quality at their convenience, without the need for expert intervention.

The four core components of Trinity's design include:

i. User Interface: A responsive and intuitive mobile application that guides users through the testing process, offers real-time results, and visualizes data to provide actionable insights.
ii. Deep Learning Models: Advanced predictive algorithms analyze water samples' physical, chemical, and biological properties, making accurate assessments of potability and identifying potential hazards.
iii. Cloud Back-End Support: A secure and scalable cloud infrastructure that allows for seamless data storage, model training, and app updates, ensuring a reliable and efficient user experience.
iv. IoT Hardware: A compact and portable edge device equipped with an array of sensors and a digital microscope, designed to capture essential water quality data points easily and accurately.

Trinity's solution sets itself apart from other water testing methods by offering a cost-effective, easy-to-use, and accurate alternative to traditional laboratory tests. The combination of advanced deep learning models, user-friendly app design, and portable IoT hardware allows users to test water sources in real-time, making informed decisions about their water consumption.

Security and privacy are of the utmost importance for Trinity. The platform employs robust encryption and authentication methods, such as 128-bit encryption and JSON Web Tokens, to protect users' data. Trinity adheres to strict data privacy standards, maintaining de-identified information and complying with relevant regulations, such as the GDPR.

By integrating state-of-the-art technology and user-centered design, Trinity's water potability testing device enables individuals and communities to take control of their water quality assessment and make well-informed choices about their water sources.

Use Case

Given below is a use case that highlights how we envision the product to be used and the impact it can have on the life of our users:

Samantha is a mother of two living in a rural area. Her family's access to clean drinking water is always a concern, and she has been searching for a reliable solution to test their water quality. One day, she hears about the Trinity water potability testing device and decides to give it a try. She orders one online, and upon arrival, she is impressed by its compact design and waterproof casing.

To begin, Samantha downloads the smartphone app and registers for an account. Once logged in, she finds herself on the home screen, which provides informational windows on how to use the Trinity device and its various features. Eager to test her well water, she proceeds to the test screen.

As the Trinity device calibrates its sensors, Samantha sees a loading progress circle on the test screen. When ready, she selects the appropriate sensor readings for her water sample: dissolved oxygen, temperature, turbidity, pH, and conductivity.

As the test progresses, Samantha observes the Water Quality Index and notices the "Within limits" or "Beyond Limits" status for each sensor reading. Upon completion, the app clearly indicates whether the water is potable or not.

Samantha then explores the dashboard screen, which offers detailed graphics on the Water Quality Index and the number of potable water sources in her area. She can also view prediction graphics for each sensor and access a "Community Map" button. Clicking on the Community Map button, Samantha sees a map of her immediate location with markers for each tested water source. By selecting one of the markers, she can view the test results, site name, and the water's potable or non-potable status.

Thanks to the Trinity water potability testing device, Samantha and her family can now make informed decisions about their drinking water. The device not only provides accurate and reliable information but also connects them with other users, contributing to a growing database of water quality data. Together, they can work to improve the quality of life in their community.

Technology

Application of Artificial Intelligence

Water Potability Analysis
Description
The first segment in Trinity that leverages AI is the analysis of water potability by identifying key parameters, including chemical, physical, and biological properties. The primary goal of this analysis is to determine the water quality and identify potential hazards that might affect the users' health. The IoT hardware captures essential data points from water samples, which are then analyzed by the Trinity mobile app.

Approach
Trinity employs deep learning models to analyze water quality parameters, making use of transfer learning and feature extraction techniques. Water sample data is collected and undergoes pre-processing, which includes:

i. Data normalization and scaling
ii. Removal of outliers and noise
iii. Augmentation techniques to generate more training data

The deep learning model for water quality analysis is trained on pre-processed data using various architectures, such as Vision Transformers (ViTs) and Recurrent Neural Networks (RNNs). Models like InceptionV3, DenseNet, and EfficientNet are explored to achieve consistent accuracy in water potability prediction.

Trinity Mobile App
Description
The Trinity mobile app serves as an easy-to-use interface for users to test water quality without requiring expert knowledge. It simplifies the testing process, providing real-time results and visualizations of water quality data.

Approach
The app is developed using a cross-platform framework like Flutter, which allows for a single codebase to run natively on both Android and iOS. The deep learning models are integrated into the app using TensorFlow Lite, enabling seamless execution of AI algorithms on mobile devices.

Trinity Cloud-based Platform
Trinity leverages cloud computing platforms like Google Cloud Platform (GCP) to host its deep learning models and store user data securely. The models are deployed on Google Compute Engine (GCE), while user data is stored in services like Google Cloud Firestore and MongoDB. Google Cloud Load Balancing is employed to optimize the app's performance and ensure it can handle incoming traffic.

Trinity IoT Hardware
Description
The IoT hardware for Trinity is a portable edge device featuring an array of sensors and a digital microscope to capture essential water quality data points. The compact design and ease of use make it an ideal solution for testing water quality on the go.

Approach
Trinity's IoT hardware is designed to be compatible with a wide range of smartphones, enabling users to take advantage of their mobile devices' capabilities. The hardware captures data points like pH levels,

turbidity, and presence of bacteria or contaminants. Data collected by the hardware is sent to the Trinity mobile app for further analysis by the deep learning models. In the future, Trinity's hardware could be expanded to include additional sensors for even more comprehensive water quality testing.

Trinity Data Security and Privacy

Description

Data security and privacy are critical aspects of the Trinity system. Ensuring that users' data remains confidential and protected from unauthorized access is of the utmost importance.

Approach

Trinity implements stringent security measures, such as 128-bit encryption and JSON Web Tokens, to safeguard user data. Privacy is maintained by using de-identified data and enforcing strict security clearance levels. Trinity complies with all relevant data protection regulations, including the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

Trinity User Interface and Experience

Description

A user-friendly interface is essential for the Trinity app, as it aims to provide an intuitive experience for users with varying levels of expertise. The app includes interactive visualizations of water quality data, clear instructions for using the IoT hardware, and personalized recommendations based on the analysis results.

Approach

Trinity's user interface is designed with simplicity and ease of use in mind. It employs a clean, minimalistic design with clear navigation and logical organization of features. The app is developed using a responsive design approach to ensure optimal performance and user experience across various devices and screen sizes.

Trinity Predictive Analytics

Description

One of the key goals of Trinity is to predict potential water quality issues before they become critical. By using predictive analytics, Trinity aims to empower users to take proactive steps in addressing water quality concerns and minimizing associated risks.

Approach

Trinity's deep learning models are trained to identify patterns and trends in historical water quality data, enabling the system to predict future water quality changes. As users continue to use the Trinity app and IoT hardware, the models become more accurate in their predictions. These insights are presented to users in the form of alerts, recommendations, and actionable steps to address potential issues.

Trinity Community Integration

Description

Trinity seeks to foster a community-driven approach to water quality monitoring and management. By allowing users to share their water quality data and experiences, Trinity aims to create a network of individuals working together to improve water quality and public health.

Approach

Trinity's platform includes features for users to connect with others in their local area, share water quality data, and collaborate on projects to address water quality issues. Users can also access a global database of water quality data, providing insights into trends and patterns that can inform local actions.

Trinity's Future Developments

As Trinity continues to grow and evolve, the team will explore new technologies and approaches to enhance the system's capabilities. Potential future developments include the integration of additional sensors for more comprehensive water quality testing, improved deep learning models for greater predictive accuracy, and the expansion of the Trinity community to include partnerships with governmental and non-governmental organizations working on water quality issues.

Data Sources

Proprietary Pathogen and Hazardous Particulate Dataset Creation: To create a diverse training image dataset for Trinity, measurements will be collected over several days, ensuring variations in sample salinity and imaging conditions. Creating a labeled dataset containing phase and intensity images of pathogens and hazardous particulates will be challenging due to similarities between non-pathogen/non-hazardous particles and the target particles. To overcome this, each 50 mL filtered water sample will be spiked with a high concentration (>1000 count/mL) of a pathogen or hazardous particulate, reducing the ratio of non-target particles to <5% of detected objects. These image datasets will be further processed to remove non-target particles based on size and shape, improving label purity.

A compact neural network model (DenseNet-121) will be chosen for rapid detection of target particles, which can be further improved using larger/deeper networks at the cost of additional computation time. During training, minimizing false positives will be prioritized to ensure reliable results in practical scenarios. The trained neural network will have a 0.25% average probability of misclassifying non-target particles, which will be digitally compensated for by subtracting 0.5% of the total detected particles in a measurement from the final pathogen or hazardous particulate count. This procedure will help avoid false positive decisions and can be fine-tuned based on the water type under test. The method can also be combined with existing sample concentration methods, such as immunomagnetic separation, to enhance specificity and detection efficiency.

Turbidity Dataset Creation: Water samples (pictures) will be collected and labeled using turbidity sensor values (in NTU). A dataset will be created using the pictures and labels, and then split into train and test sets. A model will be trained using a supervised learning algorithm on the train dataset and its accuracy will be assessed using the test dataset. The model will then be deployed as a .tflite file to an Android app.

Build Datasets:

a. Build a microscopy image dataset for each waterborne pathogen for unsupervised training.
b. Build a microscopy image dataset for each harmful water particulate for unsupervised training.
c. Build a microscopy image dataset for turbidity.
d. Build a time-series sensor array (pH, Dissolved Oxygen, Conductivity, Turbidity, and Temperature) dataset with potability determination labels.

Security/Privacy *(if needed)*

Trinity Data Security and Privacy

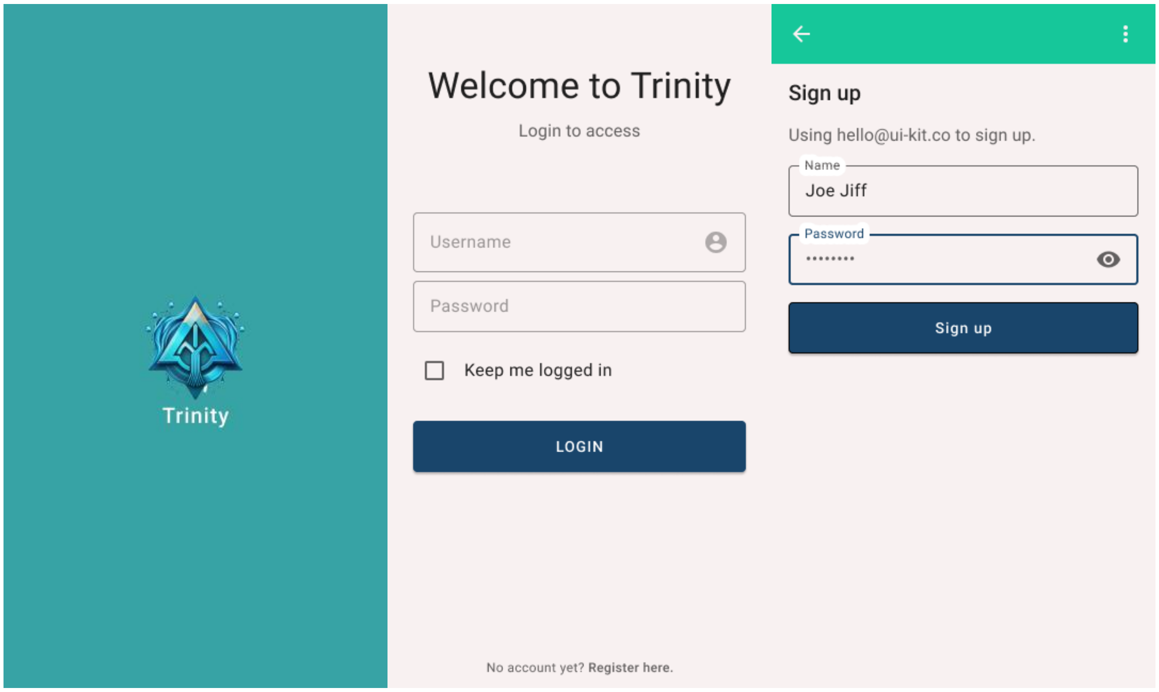
Description

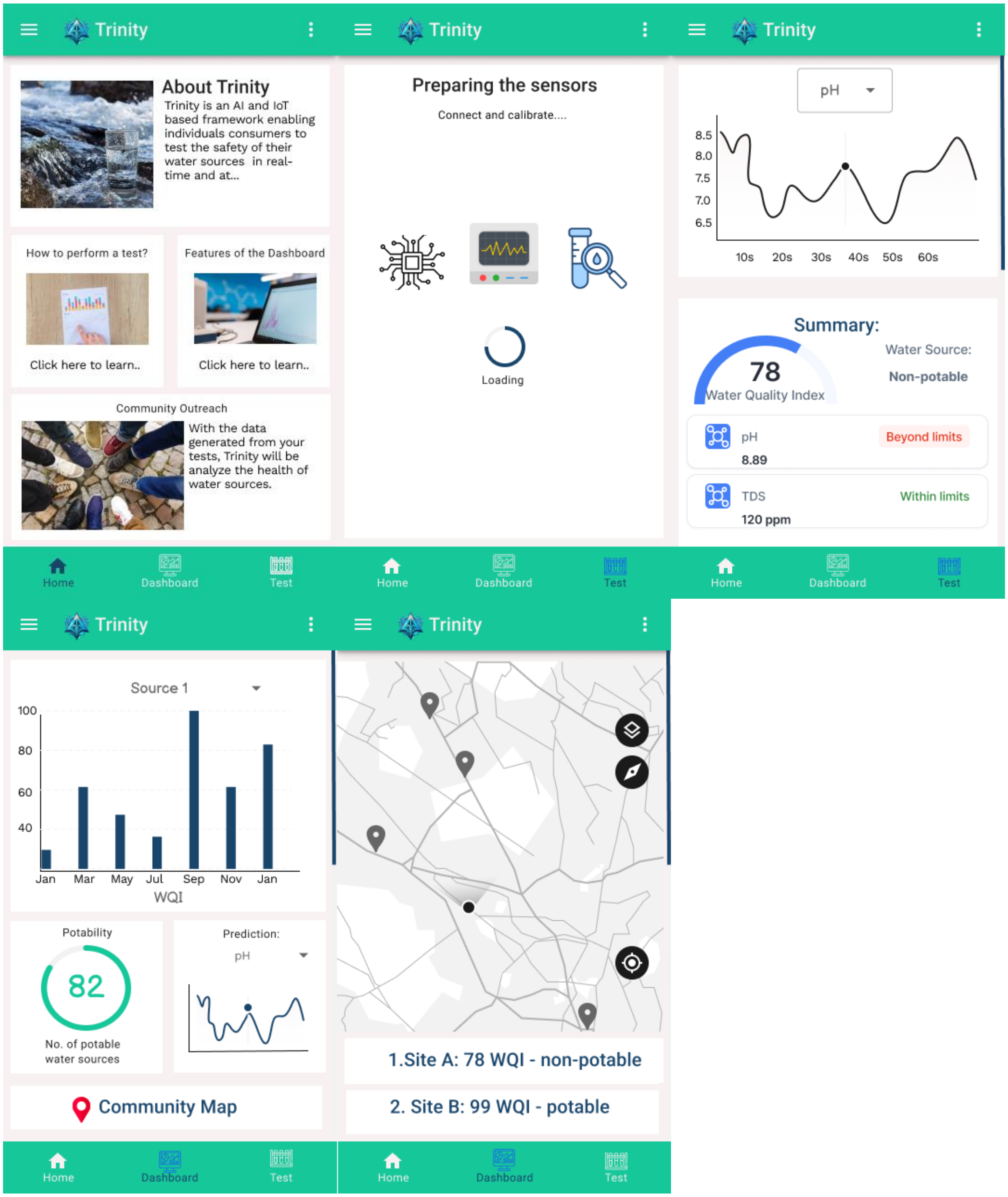
Data security and privacy are critical aspects of the Trinity system. Ensuring that users' data remains confidential and protected from unauthorized access is of the utmost importance.

Approach

Trinity implements stringent security measures, such as 128-bit encryption and JSON Web Tokens, to safeguard user data. Privacy is maintained by using de-identified data and enforcing strict security clearance levels. Trinity complies with all relevant data protection regulations, including the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

User Interface





The Trinity application can be installed on mobile devices, with support for both Android and iOS platforms. Utilizing Flutter for the front-end framework ensures a seamless user experience across platforms.

Upon installing the app, users are greeted with a welcome screen that prompts them to either log in or register for a new account. The registration process is straightforward, requiring only a name and password. Once logged in, the user is presented with a bottom toolbar featuring three buttons: 'Home', 'Dashboard', and 'Test'.

The home screen offers useful information about Trinity, including how to perform a test, an overview of the dashboard's features, and information on community outreach. The test screen guides users through the process of connecting and calibrating the Trinity device, with a loading progress circle indicating the preparation of the sensors.

Once the device is ready, the test screen displays a pull-down list of the five sensor readings (dissolved oxygen, temperature, turbidity, pH, and conductivity). A Water Quality Index is provided, along with a 'Within limits' or 'Beyond limits' status for each sensor reading. The app then determines whether the water sample is 'potable' or 'not potable'.

The dashboard screen offers various performance graphics, such as the Water Quality Index and the number of potable water sources. Users can also view a pull-down list of each sensor reading, displaying a prediction graphic for the selected sensor. A 'Community Map' button is available at the bottom of the screen, leading to a map displaying tested water sources in the user's vicinity. Selecting a marker on the map reveals the test results for that specific water source, along with its site name and potability status.

With the Trinity app, users can easily test water quality, analyze the results, and share valuable information with their community. The intuitive interface ensures that the process of water testing is accessible and user-friendly, empowering individuals to take control of their water safety.

Development Timeline

<table>
<tr><th></th><th>Pre-kickoff</th><th>Sprint 1</th><th>Sprint 2</th><th>Sprint 3</th><th>Sprint 4</th><th>Final Sprint</th></tr>
<tr><th>Timeline</th><th>Mar 21 - Apr 04</th><th>Apr 05 - May 02</th><th>May 03 - May 30</th><th>May 31 - Jun 27</th><th>Jun 28 - Jul 25</th><th>Jul 26 - Aug 07</th></tr>
<tr><td>Model Development</td><td>Procure custom dataset of:
1. Images of pathogens, and prticulate
2. water quality parameters with potability labels</td><td>Scaling predictive models for water qaulity parameters</td><td>1.Scaling predictive models for image classification
2.Create REST API endpoints for models</td><td>Review of mode performance with experts</td><td>Setting up model builing pipeline and API to serve built models</td><td rowspan="4">Integration testing and final bug fixes before final release</td></tr>
<tr><td>Sensory Array Development</td><td>Procure hardware for MVP, and finalize design</td><td>Consulting with experts and review of initial hardware</td><td>Integrate the chemical sensors and micropose</td><td>Fabricate hardware and integrate with smartphone app</td><td rowspan="3">End-to-end integration of hardware and application</td></tr>
<tr><td>Cloud platfrom</td><td>Set up individual services in GCP suite, including a DEV and QA environments</td><td>Set up schema for databases, and populate with historical data</td><td>Set up DevOPS pipeline for CI/CD</td><td>Perform load testing on the services.</td></tr>
<tr><td>Modile Application Development</td><td>Finalize UI mockup design and breakdown list of features to be developed</td><td>Replicate UI as per desgins</td><td>Create other API required to perform CRUD applications</td><td>Integrate the UI with all API sources</td></tr>
</table>

Official kick of development phase will be April 05 2023

Each sprint will be roughly 4 weeks in duration

Team Capabilities

Gayatri Bangar (Design Lead):
She has an extensive background in academic research, having published seven papers in the areas of Internet of Things and artificial intelligence while pursuing a bachelor's degree in information technology Due to her prior experience developing a Smart Waste Monitoring System combined with building Deep Learning models, she will be in charge of developing the smart edge devices.

Jason Durrance (Technical Lead):
He has 18 years of defense industry experience in systems development, has translated client provided specifications into functional designs to drive mission critical functions. He offers his expertise in technology development, bridging the gap between technological approaches, development strategies, and the required resources. He has worked with clients from various government stakeholders, including the United States Marine Corps (USMC), United States Navy (USN), and the U.S. Customs and Border Protection (CBP).

Ranojoy Deb (Team Lead):
He was employed by Mu Sigma Inc. for the past four years as a decision scientist, eventually becoming an apprentice lead in charge of a client delivery account. He actively engaged in the development of edge devices based on the Raspberry Pi, RPA and process automation, NLP chatbots, and full-stack web development during his time there. He can oversee the complete solution development thanks to his undergraduate degree in civil engineering and his experience directing the solution design, implementation, and delivery of data science projects for Fortune 500 clients.

References

1. WHO Drinking Water

2. The World Counts - Deaths from dirty water and related diseases

3. CDC – Waterborne Disease & Outbreak Surveillance Reporting

4. Water.org – An Economic Crisis

5. Office of the Director of National Intelligence - The Future of Water: Water Insecurity Threatening Global Economic Growth, Political Stability

6. PMID2019 - Google Drive

7. Forecasting water quality parameters using artificial neural network for irrigation purposes

8. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach

9. M. S. U. Chowdury, T. B. Emran, S. Ghosh, A. Pathak, M. M. Alam, N. Absar, K. Andersson, and M. S. Hossain, "IOT based real-time River Water Quality Monitoring System." [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919309391.

10. C. Z. Myint, L. Gopal and Y. L. Aung, "Reconfigurable smart water quality monitoring system in IoT environment," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 2017, pp. 435-440, doi: 10.1109/ICIS.2017.7960032.

11. A. Gupta and E. Ruebush, "AquaSight: Automatic Water Impurity Detection Utilizing Convolutional Neural Networks."

12. Ge, Z., McCool, C., Sanderson, C., Corke, P. (2015, September). Modelling local deep convolutional neural network features to improve fine-grained image classification. In 2015 IEEE International Conference on Image Processing (ICIP) (pp. 4112-4116). IEEE.