Comparison of supervised machine learning algorithms for waterborne pathogen detection using mobile phone fluorescence microscopy

Hatice Ceylan Koydemir, Steve Feng, Kyle Liang, Rohan Nadkarni, Parul Benien and Aydogan Ozcan*

Abstract: *Giardia lamblia* is a waterborne parasite that affects millions of people every year worldwide, causing a diarrheal illness known as giardiasis. Timely detection of the presence of the cysts of this parasite in drinking water is important to prevent the spread of the disease, especially in resource-limited settings. Here we provide extended experimental testing and evaluation of the performance and repeatability of a field-portable and cost-effective microscopy platform for automated detection and counting of *Giardia* cysts in water samples, including tap water, non-potable water, and pond water. This compact platform is based on our previous work, and is composed of a smartphone-based fluorescence microscope, a disposable sample processing cassette, and a custom-developed smartphone application. Our mobile phone microscope has a large field of view of ~0.8 cm² and weighs only ~180 g, excluding the phone. A custom-developed smartphone application provides a user-friendly graphical interface, guiding the users to capture a fluorescence image of the sample filter membrane and analyze it automatically at our servers using an image processing algorithm and training data, consisting of >30,000 images of cysts and >100,000 images of other fluorescent particles that are captured, including, e.g. dust. The total time that it takes from sample preparation to automated cyst counting is less than an hour for each 10 ml of water sample that is tested. We compared the sensitivity and the specificity of our platform using multiple supervised classification models, including support vector machines and nearest neighbors, and demonstrated that a bootstrap aggregating (i.e. bagging) approach using raw image file format provides the best performance for automated detection of *Giardia* cysts. We evaluated the performance of this machine learning enabled pathogen detection device with water samples taken from different sources (e.g. tap water, non-potable water, pond water) and achieved a limit of detection of 12 cysts per 10 ml, an average cyst capture efficiency of ~79%, and an accuracy of ~95%.

We evaluated the performance of this machine learning enabled pathogen detection device with water samples taken from different sources (e.g. tap water, non-potable water, pond water) and achieved a limit of detection of 12 cysts per 10 ml, an average cyst capture efficiency of ~79%, and an accuracy of ~95%. Providing rapid detection and quantification of waterborne pathogens without the need for a microbiology expert, this field-portable imaging and sensing platform running on a smartphone could be very useful for water quality monitoring in resource-limited settings.

Keywords: mobile phone microscopy; machine learning; low resource settings; fluorescence microscopy; *Giardia* detection; waterborne parasites.

1 Introduction

*Giardia lamblia* is a waterborne pathogen that affects millions of people each year worldwide and causes health problems such as diarrhea, not only in developing countries but also in developed countries; it is one of the most common infectious parasites in contaminated drinking water [1–3]. Traditional treatment methods such as chlorination are not as effective to remove the cysts of the parasite because the cyst’s thick cell wall makes it moderately resistant to chlorine [1, 2]. Therefore, timely detection of the presence of the cysts in water samples is extremely important to treat and prevent the spread of the disease, especially in field settings [4–10].

*Corresponding author: Aydogan Ozcan*, Department of Electrical Engineering, University of California Los Angeles (UCLA), CA 90095, USA; Department of Bioengineering, University of California Los Angeles (UCLA), CA 90095, USA; and California Nanosystems Institute (CNSI), University of California Los Angeles (UCLA), CA 90095, USA, e-mail: ozcan@ucla.edu
Hatice Ceylan Koydemir and Steve Feng: Department of Electrical Engineering, University of California Los Angeles (UCLA), CA 90095, USA
Kyle Liang: Department of Computer Science and Engineering, University of California Los Angeles (UCLA), CA 90095, USA
Rohan Nadkarni and Parul Benien: Department of Bioengineering, University of California Los Angeles (UCLA), CA 90095, USA

©2017, Aydogan Ozcan, et al., published by De Gruyter.
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.
Toward this goal, we previously developed a mobile phone-based waterborne pathogen detection platform [11]. This field-portable and cost-effective platform (Figure 1) uses the machine learning-based analysis of fluorescence images acquired by a mobile phone microscope and does not require a benchtop device for sample preparation or a microbiology expert for the analysis of results. This mobile phone-based platform achieved an average cyst capture efficiency of 79% and a limit of detection of 12 cysts per 10 ml of water sample [11]. In this paper, we provide a survey of supervised machine learning classifiers to further improve the same mobile platform’s cyst detection sensitivity and specificity, together with several performance improvements on the sample preparation side. For this study, we created several prototypes of our mobile phone-based pathogen detection platform to evaluate device to device variations and prototyping errors. Furthermore, a cyst training database containing >30,000 fluorescent images of individual *Giardia* cysts, each labeled by an expert, was created at different image file formats including joint photographic experts group (jpeg) and digital negative (DNG) formats. We further improved our training image database for raw images by increasing the number of dust particles (i.e. other fluorescent objects that are *not* cysts) to ~155,000 to improve the specificity of the platform. On the basis of this significantly enlarged database of images, we then evaluated the overall classification accuracies of several supervised machine learning classifiers [12–16] (including, e.g. support vector machine (SVM) [15], decision trees [14], and ensemble [16] methods) and demonstrated that a bagging classifier model gives the highest accuracy levels and the largest area under the receiver operating characteristics (ROC) [17] curve. Our results also demonstrated that raw image format (DNG) is better than other file formats in terms of detection and classification accuracy and the area under the ROC curve.

We tested this machine learning enabled pathogen detection platform using various types of water samples (e.g. tap water, pond water, non-potable water, and reagent grade water) that are screened using our pre-filtration device which removes unwanted particles/debris prior to introducing *Giardia*-specific staining into water samples. Our cyst counting results provide very good agreement with expert counting results obtained with a benchtop
fluorescence microscope as well as with our dilution tests using flow-cytometer counted samples. This waterborne pathogen detection platform with its rapid, sensitive, and automated cyst detection and counting interface can be a useful tool for monitoring of water quality in field settings and resource-limited environments.

2 Experimental section

2.1 Materials

We created three prototypes using the design described in our earlier work [11] and purchased a different smartphone for each prototype to take into account the variability among prototypes and smartphones. Mechanical parts of the prototype (Figure 1) were printed using a 3D printer (Dimension Elite, Stratasys, Eden Prairie, MN, USA) with acrylonitrile butadiene styrene material. Our mobile fluorescence microscope uses the rear camera of a smartphone (Nokia Lumia 1020, Microsoft, WA, USA) to capture raw format (i.e. DNG) and JPEG images and contains an emission filter (product no. FF01-500/LP23-3-D, Semrock Inc., Rochester, NY, USA), an achromatic lens \((f=30 \text{ mm})\) and a hydrophilic polycarbonate black filter membrane. Our mobile fluorescence microscope as well as with our dilution tests can be easily increased to absorb larger volumes of water. The porous spacing layer is used to eliminate backflow of the particles absorbed on the pads. The double-sided adhesive tape covered with black masking tape is coated with epoxy to reduce the auto-fluorescence of the tape under blue excitation. This tape is used to eliminate the movement of the black filter membrane while filtering the sample water. The casings can be washed using deionized water and reused while the other components of the cassette should be replaced with new ones for the analysis of each new sample.

We used a barrel of a 10 ml disposable syringe with BD Luer-Lok™ tip (product no. 309604, BD Company, Franklin Lakes, NJ, USA) as our sample collection tube in this study to decrease the overall cost and increase sample recovery rate by eliminating dead volumes in the capped tube. Syringes with a rubber-tipped piston ensured smooth delivery of the liquid samples. Luer caps (product no. FTLLP-1, Nordson Medical, Loveland, CO, USA) were used as syringe caps while storing or shaking the sample in a syringe barrel vigorously. A pre-filtration device was designed to capture large particles or dirt before the sample water enters into the syringe barrel. The device is composed of a filter holder (product no. SX0002500, EMD Millipore), a female Luer lug style coupler (product no. SX0002501, EMD Millipore), and a hydrophilic nylon net filter membrane with 20 \(\mu\text{m}\) pore size (product no. NY2002500, EMD Millipore). A female Luer lug style coupler (product no. FTLC-9, Nordson Medical, Loveland, CO, USA) was used to couple the filter holder to a syringe. A small piece of regular aluminum foil (product no. 01213-100, Fisher Scientific, Waltham, MA, USA) as our sample collection tube in reagent grade water (product no. 23-751628, Fisher Scientific, Waltham, MA, USA) in reagent grade water (product no. 23-751628, Fischer Scientific, Waltham, MA, USA) was used to reduce non-specific binding of particles and \textit{Giardia} cysts onto the surface of the syringe barrel and hydrate the porous spacing layer and the black filter membrane while assembling the disposable sample processing cassette.

We used a barrel of a 10 ml disposable syringe with BD Luer-Lok™ tip (product no. 309604, BD Company, Franklin Lakes, NJ, USA) as our sample collection tube in this study to decrease the overall cost and increase sample recovery rate by eliminating dead volumes in the capped tube. Syringes with a rubber-tipped piston ensured smooth delivery of the liquid samples. Luer caps (product no. FTLLP-1, Nordson Medical, Loveland, CO, USA) were used as syringe caps while storing or shaking the sample in a syringe barrel vigorously. A pre-filtration device was designed to capture large particles or dirt before the sample water enters into the syringe barrel. The device is composed of a filter holder (product no. SX0002500, EMD Millipore), a female Luer lug style coupler (product no. FTLC-9, Nordson Medical, Loveland, CO, USA) was used to couple the filter holder to a syringe. A small piece of regular aluminum foil (product no. 01213-100, Fisher Scientific, Waltham, MA, USA) as our sample collection tube in reagent grade water (product no. 23-751628, Fischer Scientific, Waltham, MA, USA) was used to reduce non-specific binding of particles and \textit{Giardia} cysts onto the surface of the syringe barrel and hydrate the porous spacing layer and the black filter membrane while assembling the disposable sample processing cassette.

We used a barrel of a 10 ml disposable syringe with BD Luer-Lok™ tip (product no. 309604, BD Company, Franklin Lakes, NJ, USA) as our sample collection tube in this study to decrease the overall cost and increase sample recovery rate by eliminating dead volumes in the capped tube. Syringes with a rubber-tipped piston ensured smooth delivery of the liquid samples. Luer caps (product no. FTLLP-1, Nordson Medical, Loveland, CO, USA) were used as syringe caps while storing or shaking the sample in a syringe barrel vigorously. A pre-filtration device was designed to capture large particles or dirt before the sample water enters into the syringe barrel. The device is composed of a filter holder (product no. SX0002500, EMD Millipore), a female Luer lug style coupler (product no. FTLC-9, Nordson Medical, Loveland, CO, USA) was used to couple the filter holder to a syringe. A small piece of regular aluminum foil (product no. 01213-100, Fisher Scientific, Waltham, MA, USA) as our sample collection tube in reagent grade water (product no. 23-751628, Fischer Scientific, Waltham, MA, USA) was used to reduce non-specific binding of particles and \textit{Giardia} cysts onto the surface of the syringe barrel and hydrate the porous spacing layer and the black filter membrane while assembling the disposable sample processing cassette.
to 1× according to the manufacturer’s instructions and used to label *Giardia* cysts fluorescently using the direct labeling method. No-Fade™ Mounting Medium (product no. M101, Waterborne Inc., New Orleans, LA, USA) containing an anti-fading agent (i.e. 1,4-diazabicyclo[2.2.2] octane) was used to reduce photo-bleaching of fluorescein on the antibodies. Blockout counterstain (product no. C101, Waterborne Inc., New Orleans, LA, USA) was used before applying the mounting medium on the filter membrane to enhance the contrast of fluorescein labeled cysts. All the suspensions and reagents purchased from Waterborne Inc. were stored at 4°C until use.

### 2.2 Sample preparation

Sample preparation steps are described in our previous article [11] with the exception of using a syringe barrel instead of a 15 ml falcon tube as a sample container and a pre-filtration unit (Figure 2 [11]). In summary, we first modified the surface of a syringe barrel using 0.01% Tween 20 solution. For this purpose, 10 ml of Tween 20 solution was collected into the barrel and stored at room temperature for 20 min to reduce nonspecific binding of *Giardia* cysts. After emptying the syringe barrel, we collected 10 ml of water under test (e.g. tap water, pond water, non-potable water, or reagent grade water) into the barrel while a pre-filtration unit was coupled to the syringe barrel. After the collection of the water sample into the barrel, the pre-filtration unit was removed and we added 200 μl of 1× diluted *Giardia*-specific stain into the sample and closed it using a Luer cap. After shaking the mixture gently for about 10 s, a small piece of aluminum foil was wrapped around the barrel and stored at room temperature for 30 min for labeling. The water sample was then dispensed onto a black membrane in the disposable sample processing cassette. The barrel was refilled with 5 ml of reagent water from its piston end and shaken vigorously for 15 s to wash out the remaining sample in the barrel. This step was repeated twice. Two hundred microliters of counterstain (diluted in 3:1 ratio) was dispensed onto the filter membrane, with a waiting time of about 1 min. Finally, we dispensed a few droplets of mounting medium to cover the black membrane and decrease the fading of the fluorescent stain. The cassette containing the fluorescently labeled *Giardia* cysts was then coupled to the attachment unit and an image of the filter membrane was captured using our waterborne pathogen detection platform through our custom-developed Windows-based application. We

---

**Figure 2:** Sample preparation, image processing, and machine learning workflow.  
(A) Steps for populating a training image database and choosing a classifier model [11]. Following a generally similar approach compared to [11], the sample preparation steps outlined in (A) have been optimized and refined to handle realistic water samples, e.g. tap water and pond water. These optimized sample preparation steps are highlighted with bold letters. The number of steps to prepare a water sample for testing is reduced by using a syringe barrel as a sample collection tube and the total time required for filtration is shortened to 21 min. (B) Steps for automated detection and classification of *Giardia* cysts using the trained model in (A).
prepared *Giardia*-spiked water samples using 10 ml of reagent grade water for the population of training data, while we used water samples from other sources (such as potable water) to independently validate the performance of the mobile device.

### 2.3 Population of training data for machine learning

We created three prototypes for three different mobile phones of the same manufacturer and model number to take into consideration potential errors that might be introduced by variations among phones and prototypes. We then tested combinations of these prototypes and mobile phones, and named each combination with a number and letter; for example, combination “1C” uses the first phone and the third prototype attachment to create a mobile phone-based cyst detection platform. To test these combinations (e.g. 1A, 1B, 1C, 2B, 3C, etc.) and assess the variation in their cyst detection performance, we prepared several water samples containing *Giardia* cysts, in addition to control samples without any cysts, and analyzed them using the procedures described in Section 2.2 to populate >30,000 cyst images to be used as training data for machine learning. Each filter membrane that contained captured *Giardia* cysts was imaged using a different combination of the prototypes and mobile phones as detailed earlier. The time sequence of capturing an image using these different combinations of mobile devices was also shuffled to reduce the systematic effect of photobleaching on our training image database. In our database, we also varied the format of the acquired image files to evaluate which file format performs better for automated recognition of *Giardia* cysts. Using our custom-developed smart application, named as *GiardiaAnalyzer*, a JPG image with 5360×7136 pixels (11.5 MB) was captured. In addition to this, using the regular application of the smartphone, named Lumia Camera, a JPEG image with 1936×2592 pixels (1.42 MB) and a raw image (DNG) with 5360×7152 pixels (47.4 MB) were also captured for each experiment. Therefore, we had three images (i.e. *GiardiaAnalyzer* JPG, Lumia JPG, and Lumia DNG) for each tested sample under a given configuration, e.g. 1C, 2B, etc.

### 2.4 Automated detection of *Giardia* cysts

After the image capture, each filter image was transferred to a local or remote PC and digital image analysis was performed using our custom-developed algorithm, which extracts >70 spatial features corresponding to each fluorescent object (i.e. cyst candidate) – see Figure 2. We used the same algorithm for each image type with some minor modifications (e.g. threshold values). The raw format images were initially converted to .TIFF images using dcraw. Following the extraction of the R, G, and B channels, the region of interest, i.e. the surface area of the membrane, was automatically cropped. After background subtraction, the fluorescent particle candidates were detected automatically and >70 features corresponding to each particle were then extracted from the image. Some of these features include eccentricity, orientation, area, perimeter, convex hull, and minor and major axis lengths as well as maximum, minimum, and mean intensity values of each detected particle at RGB, YUV, YCbCr, and HSV color spaces.

### 2.5 Labeling of *Giardia* cysts for supervised learning

After capturing the images of the test filter membranes using our smartphone-based fluorescence microscope combinations detailed earlier, the filter membrane was removed from the disposable sample cassette and placed on a microscope slide to be analyzed by an expert on a benchtop microscope using a GFP filter set and a 40× objective lens. This image labeling procedure was repeated for each filter membrane that we imaged using our mobile platform. We also developed a graphical user interface for labeling each fluorescent particle (i.e. a cyst candidate) detected on an image captured using a smartphone-based fluorescence microscope as either a cyst or other/dust to populate training image data. This interface shows two images to the user. The benchtop microscope image on the left contains green and blue circles indicating which objects are single cyst and clustered cysts, respectively. The smartphone image on the right does not have any labels and is the target image that the program attempts to match the benchtop microscope image to. For this matching task (between the images of a benchtop microscope and our mobile microscope) we used a library called vl_sift, which is based on scale-invariant feature transform (SIFT). SIFT digitally matches two images to each other by finding common objects between the images and computing homography matrices, returning what it finds to be the best match. We then used this homography matrix to map coordinate points from the benchtop microscope image to the smartphone image. After the program finishes mapping, the program presents the user with a comparison between the two images, allowing the user to fill...
in data with how many cysts were in a blue circle. After finishing the labeling, the user saves the new array that contains labels for each row in the extracted feature table. This new array was used to classify unknown particles using machine learning classifiers (Figure 2A), which will be detailed in Section 3.

2.6 **GiardiaAnalyzer: our custom-developed smartphone application**

We developed a Windows-based smartphone application, named as **GiardiaAnalyzer**, to automatically detect and count **Giardia** cysts using our waterborne pathogen detection platform (Figures 2B and 3). The application allows the user to capture an image, select an existing image in the directory of the phone, upload an image to our servers for digital image processing and search for **Giardia** cysts, and view the history of measurements. The user is also guided on how to use the smartphone microscope with on-screen instructions. After selecting an image to be uploaded to our servers, the GPS location and date/time settings of the smartphone are attached to the image automatically, and the user can assign a custom job name to each image to see the results (e.g. cyst count, GPS location, and date/time) in the history. The application enables processing of more than one image on the servers and updates the user about the status of each image (e.g. uploading, processing, etc.).

For each water sample that is imaged, digital image processing and automated classification of fluorescent particle candidates captured on the filter membrane are done at our servers. Our training data contain more than 30,000 fluorescent images of **Giardia** cysts and, after selecting a test image for uploading to the servers, it takes ≤90 s to get the detection/counting results back on the screen of the smartphone.

3 **Results and discussion**

Our smartphone-based microscope has a large field of view (FOV) that covers the entire area of the filter membrane, i.e. ~0.8 cm², and has no moving parts. This handheld, field-portable and cost-effective platform weighs only ~205 g (excluding the phone, but including the sample cassette, which weighs ~25 g) and is battery-powered, i.e. does not require stable electricity to operate. Sample preparation is easy and the entire platform can be operated by a minimally trained user. There is no need for a pump or any benchtop device for sample preparation and optical detection. Also, there is no need for a microbiology expert to label and count the cysts that are captured since both of these steps are automated in our mobile platform using machine learning.

We used supervised learning algorithms to classify the detected particles on our fluorescence microscope images and be able to specifically count **Giardia** cysts in the sample. In supervised machine learning, a known set of input data and known responses to this data are used to form and train a model, which is then used to predict the labels of the newly generated data, which in our case are fluorescence images captured using our mobile microscope. To increase the detection and classification accuracy of our platform, we first populated training data and evaluated the detection performance of our image processing algorithms with the images captured using **GiardiaAnalyzer** (i.e. JPG-**GiardiaAnalyzer**) as well as the regular camera application of the phone (i.e. JPG and DNG file formats). In these experiments, we combined the image data corresponding to five mobile device configurations (i.e. 1A, 1B, 1C, 2B, and 3C – see Section 2.3) for each image file format. In our database formed with these images, the total number of fluorescent particles labeled as **Giardia** cysts, also confirmed with a benchtop fluorescence microscope, was >30,000. For example, the training dataset populated using raw DNG images is composed of 30,674 cyst images and 104,699 images of other fluorescent objects, including, e.g. dust particles and background noise. Similar numbers of images are used to form training datasets in other file formats.

After populating the training data for each image file format, we assigned “1” as the label for a single cyst or a cluster of cysts and “0” as the label for other fluorescent objects (non-cysts) in our mobile phone images to have binary classification of fluorescent objects. We chose 5% of this populated data for each image format as our independent test data and 95% as our training data to compare the performances of different prediction models. In addition to the classification accuracy, the area under ROC curve (AUROC) is also another measure that we used to demonstrate how well the selected algorithm performs; for example, the accuracy of a test is classified as “excellent” if its AUROC is >0.90. In our comparisons, we tested several supervised learning algorithms such as SVMs [18], nearest neighbors [13, 19], and ensemble methods [20]. Figure 4 summarizes the effect of the image file format on the overall classification accuracy and the specificity of different classifier models. The classification accuracy of all the supervised learning methods that we tested are >81%, while the AUROC values are >0.7 regardless of the image file format. The best performance is obtained with
the raw image format (i.e. DNG images), which is due to the fact that it preserves spatial details, retaining more information in the extracted features of each fluorescent object or cyst candidate. Lumia Camera JPG images resulted in a lower classification accuracy compared to that of JPG images captured using our GiardiaAnalyzer application since Lumia Camera JPG images are significantly more compressed, lowering the image quality. As shown in Figure 4, the best cyst predictive accuracy of ~95% and an AUROC value of 0.97 are obtained using bagged trees (number of trees, $N_{\text{tree}} = 400$) as a classifier model using DNG images. Fine and cubic k-nearest-neighbor classifiers [18, 21, 22] provided fast fitting speeds but their predictive accuracy values are relatively poor, especially for

---

**Figure 3:** (A) Following our previous design reported in [11], graphical user interface of our smartphone application, GiardiaAnalyzer, is shown. The new version of the application has several additional features: (1) it enables parallel processing of the acquired images in the server, saving processing time; (2) it provides continuous feedback to the user about the status of the image analysis; and (3) it enables assigning a job name for each image to be uploaded to server. (B) Screenshots of the application for different stages of the image analysis. (B-i,ii,iii) Photographs demonstrating the feedback to the user about the image analysis and processing of the results.
Next, we tested the performance of the best prediction model, i.e. bagging trees with $N_{\text{tree}} = 400$, for the detection of *Giardia* cysts in different types of water including, e.g. reagent grade water, non-potable water, pond water, and tap water samples using the training database populated with raw format DNG images. Using 71 spatial features and ~104,000 images of non-cyst fluorescent objects in our database, our mobile platform measured 4.00 ± 2.16, 2.67 ± 2.49, 3.00 ± 2.16, and 3.00 ± 2.16 cysts for 10 ml of *Giardia*-free non-potable water, pond water, reagent grade water, and tap water, respectively (see Figure 5A). When we used a training database containing more features per fluorescent particle (e.g. 96 spatial features instead of 71), false positive cyst counts per 10 ml of clean water sample decreased slightly for all the control samples. Some of the spatial features or image structures in these 25 extra features are in fact common for both *Giardia* cysts and other non-specific fluorescent particles. This introduces a sampling bias in the training set and results in our machine learning algorithm to have more false negatives as compared to the training database with 71 features and 104,000 dust particles as part of our database. When we used a training database containing more non-cyst fluorescent objects, e.g. ~155,000 images of fluorescent

---

**Figure 4:** The effect of the image file format on the classification accuracy and specificity of various classifier models. DNG images resulted in much better classification accuracy compared to other file formats.

---

JPG images. On the other hand, SVM and bagged ensemble classifiers are very good at their prediction accuracy values, while their training speeds are slower compared to other methods.

---

### Table: Comparison of Supervised Machine Learning Algorithms for Waterborne Pathogen Detection

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy (%)</th>
<th>Area under the ROC curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic SVM</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Bagged trees</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Medium Gaussian SVM</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Quadratic SVM</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Nearest KNN</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Coarse KNN</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Complex tree</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Fine Gaussian SVM</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Cubic KNN</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>Subspace KNN</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>Boosted trees</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>Sample tree</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Fine KNN</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>RUS Boosted trees</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>

---

---

---
dust particles, we observed a further improvement on the standard deviations and cyst counts per 10 ml of sample (i.e. 2.00 ± 0.00 and 1.00 ± 0.81) corresponding to non-potable water and pond water samples, respectively.

In addition to these control samples and false positive detection rates reported above, Figure 5 also reports predicted cyst count values for 10 ml of *Giardia*-spiked water samples against total cyst counts obtained using a benchtop fluorescence microscope. *Giardia* cyst count predictions trained with an image database containing 71 spatial features per object and ~104,000 images of non-cyst objects were in very good agreement ($R^2 = 0.9807$) with the cyst counts obtained using our gold standard method, especially for high concentrations of *Giardia* cysts in water samples (Figure 5B). When we used a training image database containing a larger number of non-cyst fluorescent objects, the predicted cyst count values were much better for lower concentrations (e.g. <120 cysts per 10 ml) – see Figure 5C. Therefore, a combination of both of these training databases to separately handle low and high cyst concentration ranges would provide better overall performance. These results summarized in Figure 5 demonstrate that our machine learning-based mobile microscopy platform can automatically detect, and specifically and sensitively classify *Giardia* cysts from different water sources.

There are several optical features that contribute positively to the sensing performance of our waterborne detection platform. The first feature is the wide FOV of our design. Our waterborne pathogen detection platform with its wide FOV (~0.8 cm$^2$) enables capturing an image of the entire sample filter membrane all at once, whereas, for comparison, several tens of images should be taken using, e.g. a 10x objective lens and stitched together to have the whole image of the same filter membrane using a benchtop fluorescence microscope. In our platform, the time that is needed to capture an image of this filter membrane is only 4 s without the need for any mechanical scanning or digital image alignment/stitching. Another important feature of our design is the use of a long-pass filter rather than a single-bandpass filter. As a part of our sample preparation protocol, *Giardia* cysts are labeled with fluorescein (490 nm/525 nm) conjugated antibodies for specific detection of the cysts in water samples. There may be some other fluorescent particles (e.g. dust particles) captured on the sample filter membrane, yielding fluorescence both in the green and red regions of the spectrum. Our platform utilizes a long-pass filter (i.e. transmitting >515 nm) to enhance the detection and classification performance of our supervised machine learning classifier, because some of the features included in the training database are based on intensity values of the detected particles at green and/or red image channels. We should emphasize that we used the green channel of an image acquired using our platform to detect possible fluorescent particles that are captured on the membrane. The blue channel of the acquired image does not provide any useful information since our long-pass filter blocks

**Figure 5:** Predicted cyst levels corresponding to 10 ml of (A) *Giardia*-free samples from various water sources, (B) *Giardia*-free and *Giardia*-spiked water samples, (C) low concentrations of *Giardia*-spiked samples using different training databases for a bagged tree classifier with 400 trees.
the blue region of the spectrum, while the red channel of the image contains information about dust particles only. This way, our machine learning framework has additional degrees of freedom to differentiate true positives (cysts) from non-specific fluorescent objects/particles.

4 Conclusions

We demonstrated a hand-held and cost-effective platform for automated detection and counting of Giardia cysts in large volumes of water samples using machine learning. This platform includes a mobile phone-based fluorescence microscope to capture an image of fluorescently labeled Giardia cysts captured over a filter membrane that has a large FOV of ~0.8 cm². This fluorescence image of the sample filter membrane, captured using a custom-developed application, is processed at our servers for automated detection and enumeration of Giardia cysts using a training image dataset, containing 71 spatial features per object and labels of >30,000 images of cysts and >100,000 images of non-cyst objects. The total time of this analysis per 10 ml water sample under test, from sample preparation to automated cyst counting, is less than an hour. By comparing several supervised machine learning classification models, we demonstrated that this mobile platform achieves the best performance for prediction of cyst counts using a bagging classifier with a classification accuracy of ~95%. We also demonstrated that this platform can be used for automated detection and counting of Giardia cysts in water samples taken from different sources (e.g. tap water, non-potable water, reagent grade water, and pond water) with a detection limit of ~12 cysts per 10 ml. This portable and cost-effective platform provides a powerful tool for rapid and sensitive detection of waterborne pathogens in water samples, even in resource-limited settings with minimal training of the users.

Acknowledgments: This project was funded by the Army Research Office (ARO). The Ozcan Research Group at UCLA gratefully acknowledges the support of the Presidential Early Career Award for Scientists and Engineers (PECASE), the ARO (W911NF-13-1-0419 and W911NF-13-1-0197), the ARO Life Sciences Division, the National Science Foundation (NSF) CBET Division Biophotonics Program, the NSF Emerging Frontiers in Research and Innovation (EFRI) Award, the NSF EAGER Award, NSF INSPIRE Award, NSF Partnerships for Innovation:Building Innovation Capacity (PFI:BIC) Program, Office of Naval Research (ONR), the National Institutes of Health (NIH), the Howard Hughes Medical Institute (HHMI), Vodafone Americas Foundation, the Mary Kay Foundation, Steven and Alexandra Cohen Foundation, and KAUST. This work is based upon research performed in a laboratory renovated by the NSF under Grant No. 0963183, which is an award funded under the American Recovery and Reinvestment Act of 2009 (ARRA).

References


