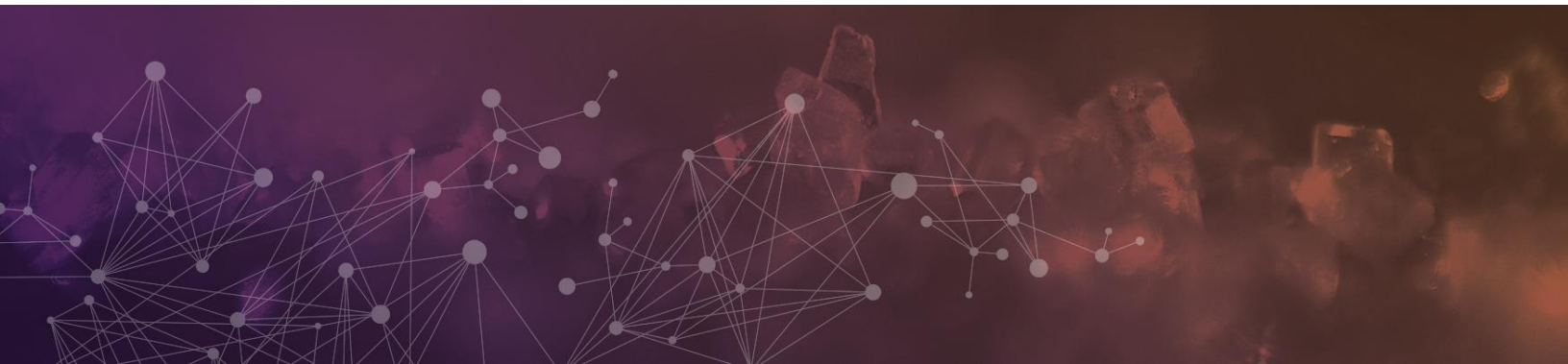
Entry to the Stockholm Junior Water Prize 2018

Employing Computer Vision and Cellulosic Biocomposites for Rapid, Automated and Cost-effective Water Analysis and Purification

Krtin Kanna Nithiyanandam

Sutton Grammar School



I. ABSTRACT

This interdisciplinary study employs a novel computer vision algorithm with a cellulosic biocomposite for the rapid classification and elimination of water contaminants. In-lab water diagnoses are currently expensive, time-consuming and inaccessible to many rural communities. However, the ubiquity of smartphone technology provides a powerful platform for water analytics. A biocomposite was engineered to filter various contaminants whilst retaining water samples for analysis with a smartphone camera system. A random forest (RF) computer vision approach was designed to categorise image features to generate a probability map. K-means clustering grouped similar pixels and a secondary RF validated different cell regions for species identification. After training, the algorithm identified heavy-metals and individually trained bacterial species within seconds with 90.16% accuracy. Biocomposite filtration removed 2,671 parts per billion (ppb) of Pb^{2+}, 2,234 ppb of Ni^{2+}, 96% of bacteria and 100% of solid particulates; initial exposure to biocomposite casein resulted in 100% coliform bacterial inactivation within the filtrate in 10 minutes. This study develops a novel, rapid and accessible system for field-analysis, mapping global water quality and purifying water in developed and developing countries.

II. TABLE OF CONTENTS

III. KEY WORDS

biocomposite, bioplastic, casein hydrolysate, coliform bacteria, computer vision, *Escherichia coli*, heavy-metals, machine learning, principal component analysis, random forest, *Sarcina aurantiaca*, *Vibrio fischeri*

IV. ABBREVIATIONS AND ACRONYMS

EDTA: Ethylenediaminetetraacetic acid	***E. coli**: Escherichia coli*	**CMOS**: Complementary metal-oxide-semiconductor
GRAS: Generally regarded as safe	**HoG**: Histogram of Oriented Gradients	**PCA**: Principal Component Analysis
ppb: parts per billion	**RF**: Random Forest	***S. aurantiaca**: Sarcina aurantiaca*
SDG: Sustainable Development Goal	**TCC**: Total Coliform Count	***V. fischeri**: Vibrio fischeri*

V. ACKNOWLEDGEMENTS

I am grateful to the entire science department at Sutton Grammar School for providing the resources and space needed to explore the wet-lab aspects of this project. I would also like to thank Mr. Costello, from Sutton Grammar School, for remaining supportive of my research endeavours and suggesting various methods to gain access to the resources I would need to conduct this project. I would like to acknowledge Dr. Palgrave, from University College London, for his inspiring discussion about the field of materials science. Moreover, I am ever grateful to my friends and peers who introduced me to the field of computer vision and have provided guidance and reassurance throughout the development of the algorithms. Finally, I am thankful to my parents for purchasing various reagents for me and for supporting this research endeavour.

VI. BIOGRAPHY

Krtin Nithiyanandam is currently a Year 13 student at Sutton Grammar School in the United Kingdom. He was recently awarded the Senior Science Winner Award at the national Big Bang Fair 2018. Krtin was also the recipient of the Scientific American Innovator Award at the Google Science Fair 2015. In 2017, Krtin was named the UK Intermediate Scientist of the Year and was the national winner of the UK Junior Water Prize, and a subsequent finalist at the Stockholm Junior Water Prize. TIME Magazine included Krtin in their list of the 30 Most Influential Teens of 2017, partly for his work on sustainable water purification. In 2018, Krtin received the Stantec Developing Communities Award. Krtin hopes to pursue a course relating to public health at university.

1. INTRODUCTION

Democratising global access to clean water is the objective of the United Nations' Sustainable Development Goal (SDG) 6 [1], with trickle-down effects to Millennium Development Goal 7 [2] and other SDGs to improve education, pioneer sustainability and eradicate hunger, poverty and disease. However, according to the World Health Organisation, over 2 billion people drink contaminated water, thus resulting in 3.4 million deaths annually from the spread of waterborne illnesses such as: *Escherichia coli* (*E. coli*), Salmonella and *Vibrio cholerae* (cholera) [3]. Furthermore, the United Nations predicts chronic water shortage will affect 1 in 4 people by 2050 [1]. Moreover, the presence of toxic heavy-metals such as mercury, lead and nickel have detrimental effects on aquatic and mammalian life; human exposure to heavy-metals can result in the impairment of vital organs and organ systems, oftentimes proving to be fatal [4].

Conventional methods to remove toxic heavy-metals from water, including reverse osmosis, distillation and activated carbon, are expensive and resource intensive. Crockett *et al.* synthesised a sulphur-limonene polysulphide material capable of binding to heavy-metals and exhibiting a chromogenic response when bound to Hg^{2+} [5]. However, this material was unable to inactivate bacteria and filter solid particulate contaminants. Nithiyanandam [6] extended the previous approach by developing a cellulosic bioplastic capable of sequestering Pb^{2+} and facilitating a photocatalytic advanced oxidation process. Similarly, the bioplastic was also unable to filter solid particulates and relied on sunlight to completely inactivate bacteria. Pervious concrete composites, such as those by Jeswani *et al.* [7], have shown to effectively remove particulate contaminants, however such filters cannot remove toxic heavy-metals and cement is considered a non-renewable resource.

Studying water composition prior to purification can yield valuable data that could be used to identify and address sources of water contamination. However, literature describing methods to accurately, rapidly and simultaneously identify multiple bacterial species and heavy-metal types without strict adherence to laboratory assays is scarce. Complementary metal-oxide-semiconductor (CMOS) sensors are popular components and are almost universal within smartphone cameras due to their fast speeds, low cost and power efficiency [8]. Therefore, CMOS sensors open the possibility for rapid, camera-based identification of water contaminants. However, smartphone imaging of bacterial cells is burdened by magnification limitations and significant background noise, with heavy-metals traditionally being impossible to identify by camera-imaging due to their presence on the ionic level. Nevertheless, clip-on magnification units for smartphones exist, thus demonstrating potential for small, cheap and durable smartphone-compatible magnification systems.

Computer vision can be used to extract image features for analysis and distinguish between valuable image regions and background noise [9]. However, previous approaches have struggled with delineation due to difficulty with identifying boundaries between adjacent cells [9, 10]. Kim *et al.*'s proposition to use support machine vector classifiers to extract cells from the background was limited by overestimation or underestimation of cell regions, particularly clustered cells [10]. Random forest (RF) ensemble approaches rely on using weak learners (trees) to ultimately yield a strong learner (forest) [11]. RFs are a popular choice due to their quick training and availability of library implementations, such as scikit-learn and R. Furthermore, RF classifiers can address overfitting issues by drawing upon a random feature set injected at each node to decide the split for decision trees [11]. Nevertheless, analysing water samples through image features is challenging, with no reported literature concerning the use of computer vision to identify individual, water-present bacterial species and heavy-metals in real-time.

This study aims to design, construct and evaluate a novel system for the rapid removal of solid particulates, bacteria and heavy-metals from water, whilst accurately identifying heavy-metals and individual bacterial species. Building upon Nithiyanandam's approach [6], the potential for cellulose, casein and D-limonene to form a novel, porous biocomposite for effective water filtration was evaluated. It was hypothesised that a porous cellulosic matrix would filter contaminants, whilst cellulose-stabilised D-limonene would sequester heavy-metals. For water analysis, the potential for a cheap, smartphone-compatible van Leeuwenhoek lens system for simultaneous water purification and analysis using RF was studied. Furthermore, the use of a script to make any analysed water data and sample images publicly accessible was also explored. This project is envisioned to not only serve the vital function of sustainably increasing accessibility to potable water, but to also provide a resource for researchers and citizens to easily and rapidly analyse local water quality, to ultimately map global water quality, thus providing insight for future approaches to improve water sustainability.

2. MATERIALS AND METHODS

2.1 Materials and Chemicals

Microcrystalline cellulose was obtained from Blackburn Distributions. Cellulose fibres, urea, (R)-(+)-Limonene (referred to as 'limonene' henceforth), 0.2M aqueous sodium hydroxide (NaOH) solution and casein hydrolysate were purchased from Sigma-Aldrich. Glycerol and sodium chloride (NaCl) were obtained from Fisher Scientific. Glass beads were used to construct the magnification unit and a Samsung RV520 laptop running Windows 7 acted as a server for data processing. The algorithms were written in Python.

For sequestration studies, 2% aqueous lead acetate ($Pb(C_2H_3O_2)_2$ and nickel (II) chloride ($NiCl_2$) solution were purchased from Fisher Scientific and Eramet respectively. 1,5-diphenylthiocarbazone (dithizone) and ethylenediaminetetraacetic acid (EDTA) was obtained from Sigma-Aldrich for analysis of Pb^{2+} and Ni^{2+} concentration respectively. Lysogeny broth was made for bacterial assays [12]. Studies were conducted with *E. coli* (K-12 strain), and *Vibrio fischeri* (*V. fischeri*) and *Sarcina aurantiaca* (*S. aurantiaca*) served as model organisms for cholera and salmonella respectively. For training and live-testing of the computer vision algorithm, the aforementioned bacteria species were used. All bacterial strains used are classified under Risk Group 1 and are considered suitable for school use. Distilled water was used in all experiments.

2.2 Manufacturing a Sample-retaining Biocomposite for Water Filtration

2.2.1 Developing a Porous Biocomposite

Naturally-occurring cellulose fibres were used to form the underlying biocomposite matrix. An aqueous NaOH-urea solution (8% NaOH and 8% urea) was prepared for the swelling and partial dissolution of the cellulose fibres [13, 14], combined with 10g of microcrystalline cellulose. Furthermore, 12g of casein hydrolysate and 35ml of limonene was added to the aqueous cellulose-NaOH-urea solution. The resulting solution was formed into a three-dimensional structure prior to being heated at 120^{o}C for 30 minutes and subsequently cooled. All biocomposite samples used for analyses had a surface area of 265cm^2.

2.2.2 Synthesizing and Integrating a Non-Porous Bioplastic for Sample Retention

A transparent bioplastic was synthesised to be integrated into the biocomposite (Figure 1) to retain water samples for analysis. The synthesis methodology was adapted from Nithiyanandam's study [5], using a 1% aqueous cellulose-fibre solution combined with 10ml of limonene instead of a cellulose-TiO_2 paste. Casein hydrolysate was not used for bioplastic synthesis to avoid bacterial inactivation within the analysis sample.

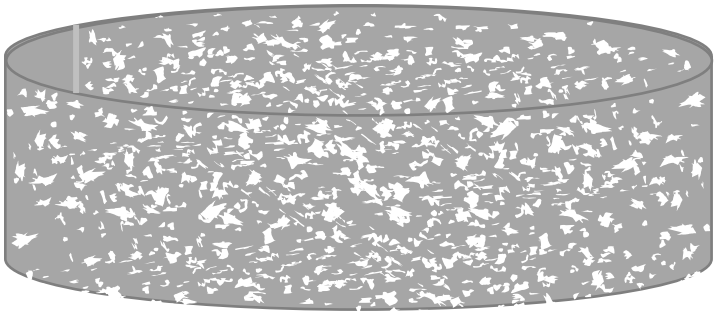


Figure 1. shows a diagrammatic representation of a porous biocomposite with a separated section for non-porous bioplastic integration to retain water samples for analysis.

2.3 Deployment of Glass Beads for Smartphone-Based Magnification

Commercially purchased glass beads were obtained and a mechanism to secure the glass bead to a smartphone camera was 3D printed as described by Erikson *et al.* [15].

Figure 2. shows images taken of *E. coli* (K-12) samples with a glass bead clip-on for smartphone cameras. **a)** shows an image at 150x magnification, **b)** shows an image at 400x magnification and **c)** shows an image at 600x magnification.

2.4 Computer Vision Scheme

2.4.1 Pixelwise Classification

The pixelwise classification stage aims to categorise the pixels comprising the sample images into four classes: bright-region, dark-region, halo artifacts and noise. 568 labelled, light-microscope images of *E. coli (K-12), S. aurantiaca and V. fischeri* were used as training data. Principal component analysis (PCA) was performed as a dimensionality reduction measure (Figure 5). T number of trees, *t*, comprising the forest are independently trained, $t \in \{1, ..., T\}$, using a random subsample of training data [16]. For the construction of this classifier, T=500 and the number of features at each internal node was set as the quartic root of the feature vector length. Testing points, *v*¸ are pushed from the tree root nodes until they reach the corresponding leaf nodes (terminal) (Figure 3) to evaluate unseen data. The combined prediction of each tree, ascertained by calculating the mean prediction for each class, *c*, forms the RF:

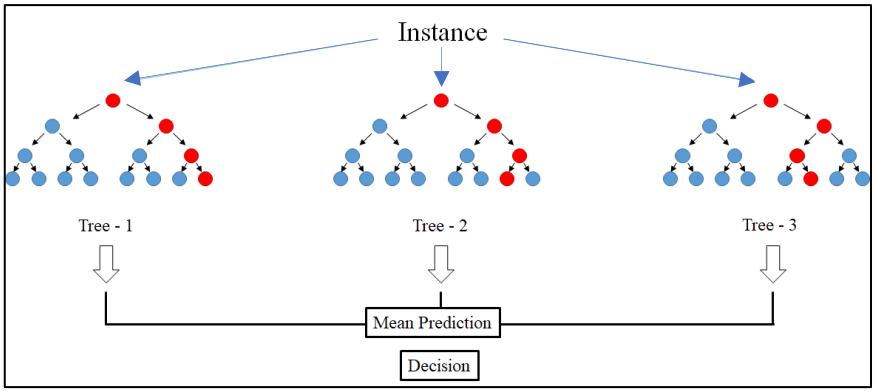


Figure 3. A simplified representation of a random forest consisting of three decision trees forming the mean class prediction.

$$P(c|v) = \frac{1}{T}\sum_{t=1}^{T} pt(c|v) \quad (1)$$

The dark-regions and bright-regions of each cell are given as a probability estimate based on the output. Artifacts are removed by restoring the sample image through dividing the ill-conditioned optics matrix, A, by the vectorised image, O, to yield the restored sample image, Rimage:

$$R_{image} = \frac{A}{O} \quad (2)$$

2.4.2 Identifying Cell Candidates

The probability estimate yielded by the pixelwise classification scheme indicates dark and bright cell regions. The next stage combines this data with identified cell outline characteristics to predict the bacterial species present. The centre peak for each cell was found using the k-means method [17] to cluster probability values (Figure 4). The output cluster which resulted in the maximisation of the probability map produced during pixelwise classification was used to identify cells. Probability maps for dark-region, bright-region and halo artifacts were noted as D, B and H respectively and the probability that a cell would be identified is:

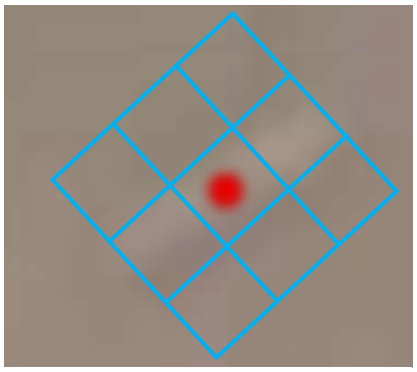


Figure 4. Division of object into sub-blocks for feature extraction to identify cell candidates.

$$P = \begin{cases} 1, & D + B + H > t, \\ 0, & otherwise \end{cases} \quad (3)$$

where t is a threshold value set to 0.8. Each cell candidate is represented as a binary value, with 1 representing a cell and 0 representing background.

2.4.3 Algorithm Validation

Finally, a validation RF is built to differentiate between individual adjacent cells and cell clusters. Cell cluster patterns are advantageous for further differentiating between species of the same genus (e.g. between *V. fischeri and V. cholerae*). Cell boundaries were treated as an individual region to determine whether the region was shared between multiple cells or belonged to an individual cell. Popular for object detection, histograms of oriented gradients (HoG) were extracted from potential cell regions [18]. The HoG acts by dividing the entire sample image into smaller regions and then combining the computed edge-oriented histograms within a larger image region. This process improves image contrast and is therefore useful for identifying individual cells and silhouettes of individual cells within clusters. The validation process decreases false-positive outputs and prevents double-detection of already identified bacterial cells.

Due to the presence of Pb^{2+} and Ni^{2+} resulting in different chromogenic responses, a simple HoG function is required to differentiate between the various chromogenic responses and background noise, therefore enabling the identification of Pb^{2+} and Ni^{2+}.

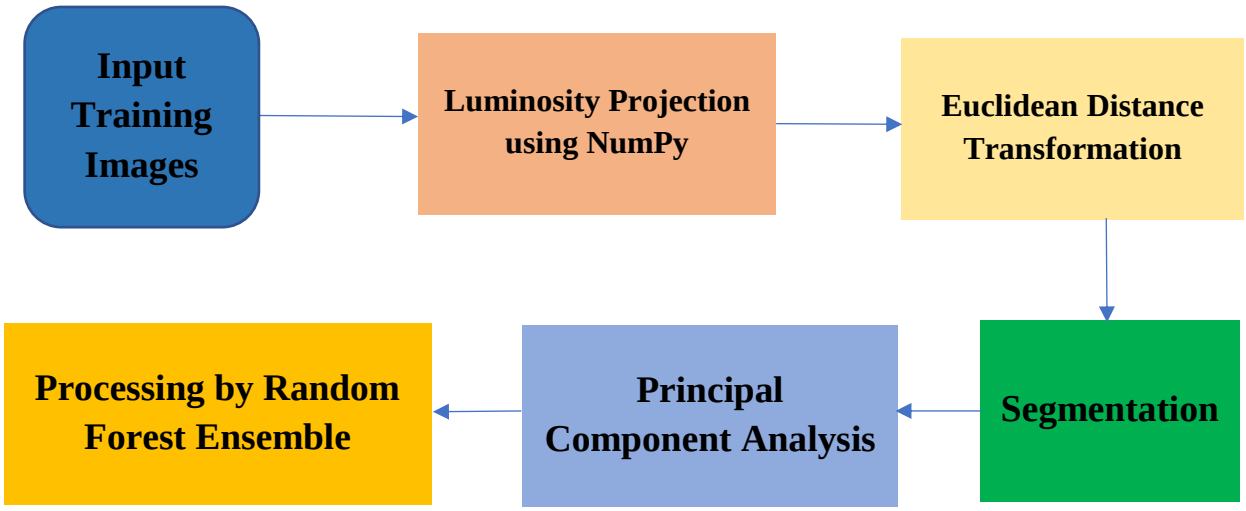


Figure 5. An outline of the algorithmic scheme by which sample images for supervised learning are processed before undergoing further processing by the random forest ensemble(s).

2.5 Extent of Water Purification

2.5.1 Filtration of Solid Particulates

Water samples were passed through filter paper. Any solid present on the filter paper was washed with distilled water and gently heated. The mass after drying was determined to ascertain the extent to which the biocomposite had filtered solid particulates.

2.5.2 Removal and Inactivation of Bacteria

E. coli, *V. fischeri* and *S. aurantiaca* were transferred to separate, autoclaved lysogeny broth samples, and incubated at 25^{o}C for 48 hours. 3M Petrifilms [19] were used to record the total coliform counts (TCCs) of each sample pre-filtration, and post-filtration at time intervals of 0s, 15s, 30s, 45s, 60s, 75s, 90s, 105s and 120s. Aseptic technique was used throughout.

2.5.3 Heavy-Metal Sequestration Studies

2.00×10^{-5} mol dm^{-3} lead acetate and 7.00×10^{-6} mol dm^{-3} nickel (II) chloride stock solutions were prepared and separately passed through the biocomposite filter. Spectrophotometric analysis was performed on pre-filtration and post-filtration water samples to determine Pb^{2+} [20] and Ni^{2+} [21] concentrations.

3. RESULTS

3. 1 Chromogenic Response to Pb^{2+} and Ni^{2+} Exposure

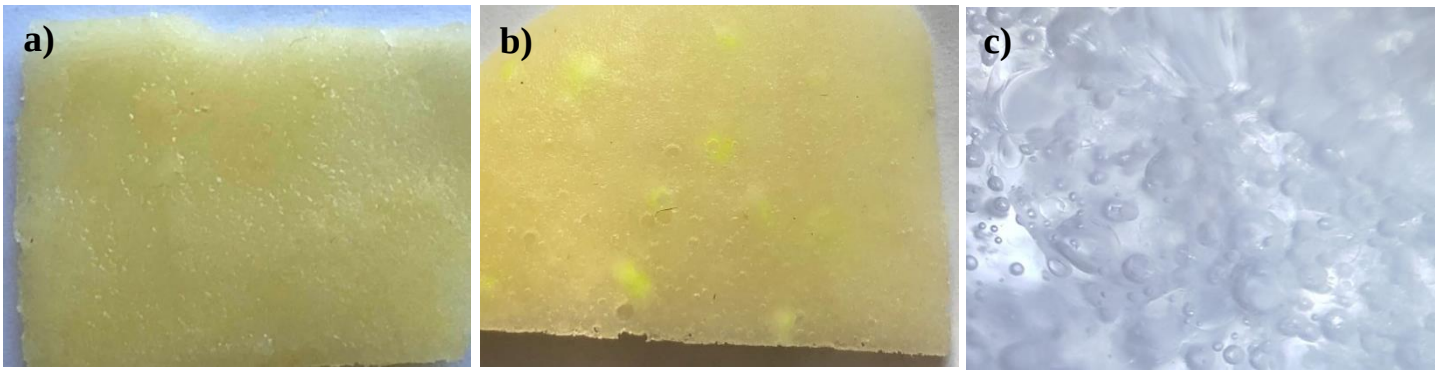


Figure 6. Sample image of the sample-retaining bioplastic after exposure to: **a)** water samples without Pb^{2+} and Ni^{2+}, **b)** water samples containing Pb^{2+} and Ni^{2+} and **c)** an image of the 2017 bioplastic sample developed by Nithiyanandam [6].

3.2 Water Contamination Classification

3.2.1 Identification of Individual Bacterial Species

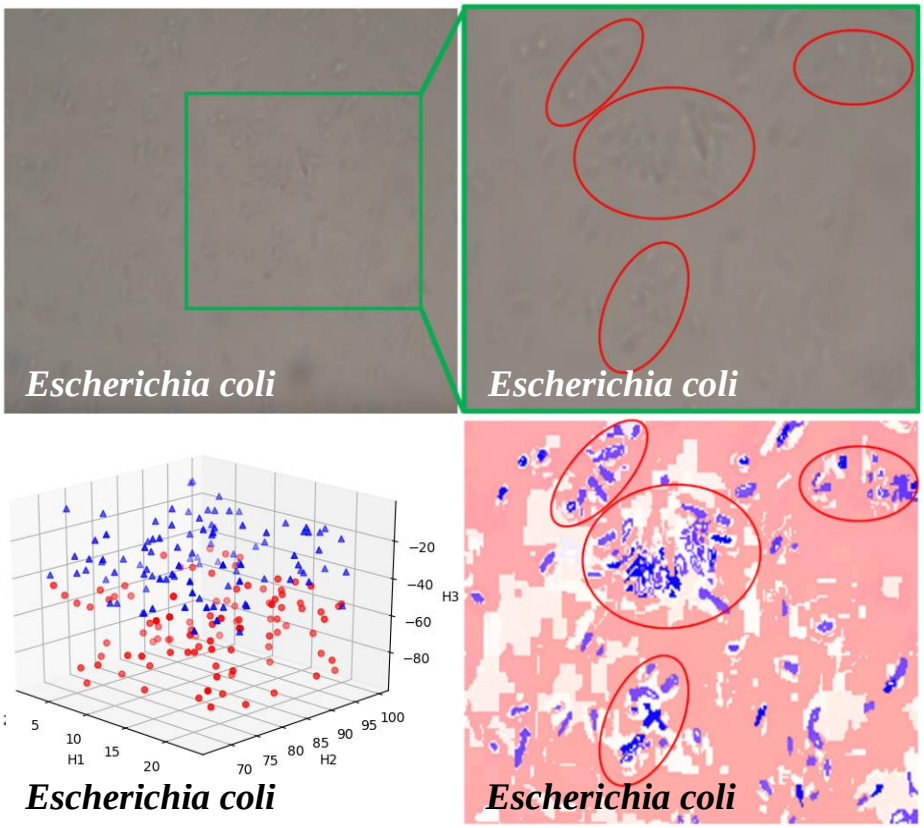


Figure 7. a) shows a sample *in-situ* image of *E. coli (K-12)* taken with the smartphone-based magnification system and **b)** shows an enlarged region of the same image for bacterial identification analysis; **c)** shows the post-PCA confidence of the computer-vision schema with blue and red dots representing cell candidates and background respectively; **d)** shows the computer-vision schema interpretation of the image with cell candidates being highlighted as blue/purple.

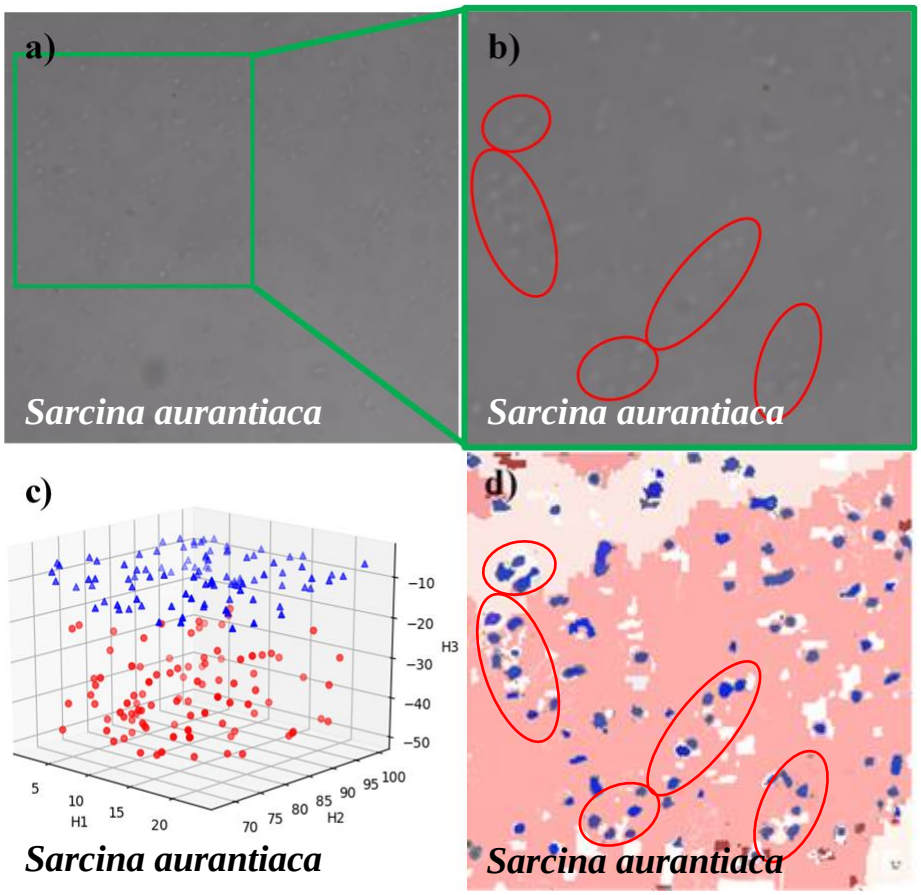


Figure 8. a) shows a sample *in-situ* image of *S. aurantiaca* taken with the smartphone-based magnification system and **b)** shows an enlarged region of the same image for bacterial identification analysis; **c)** shows the post-PCA confidence of the computer-vision schema with blue and red dots representing cell candidates and background respectively; **d)** shows the computer-vision schema interpretation of the image with cell candidates being highlighted as blue/purple.

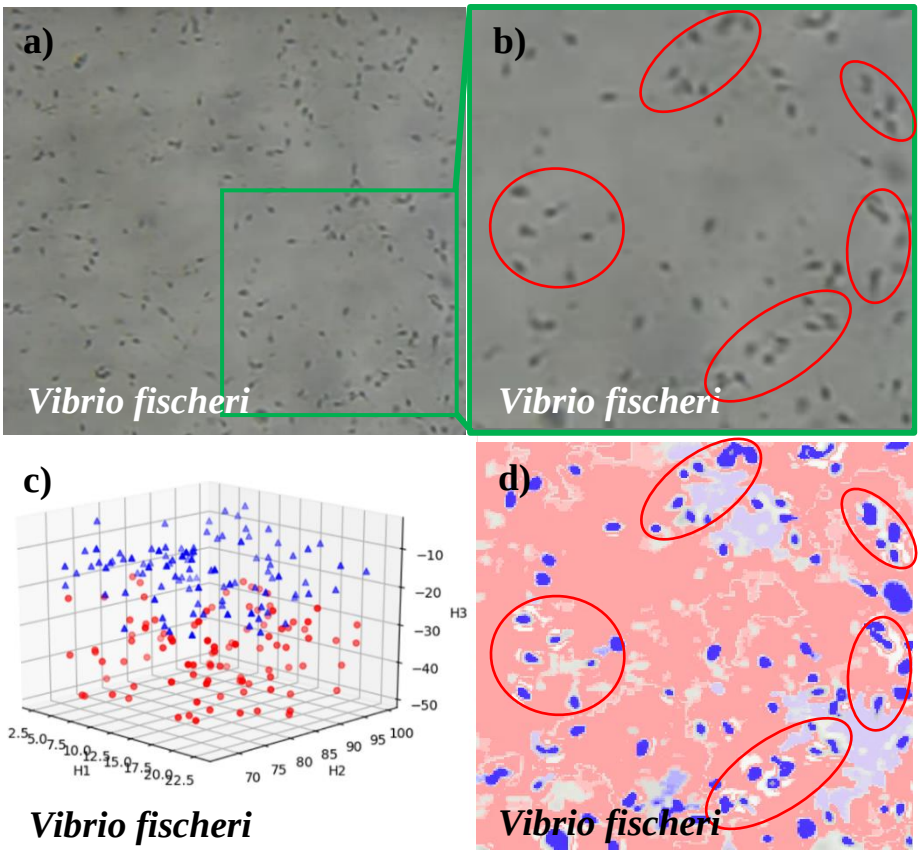


Figure 9. a) shows a sample *in-situ* image of *V. fischeri* taken with the smartphone-based magnification system and **b)** shows an enlarged region of the same image for bacterial identification analysis; **c)** shows the post-PCA confidence of the computer-vision schema with blue and red dots representing cell candidates and background respectively; **d)** shows the computer-vision schema interpretation of the image with cell candidates being highlighted as blue/purple.

3.3 Water Purification

3.3.1 Removal of Solid Contaminants

After filtration through the biocomposite, there was no measurable dry mass of solid particulates present.

3.3.2 Removal and Inactivation of Bacteria

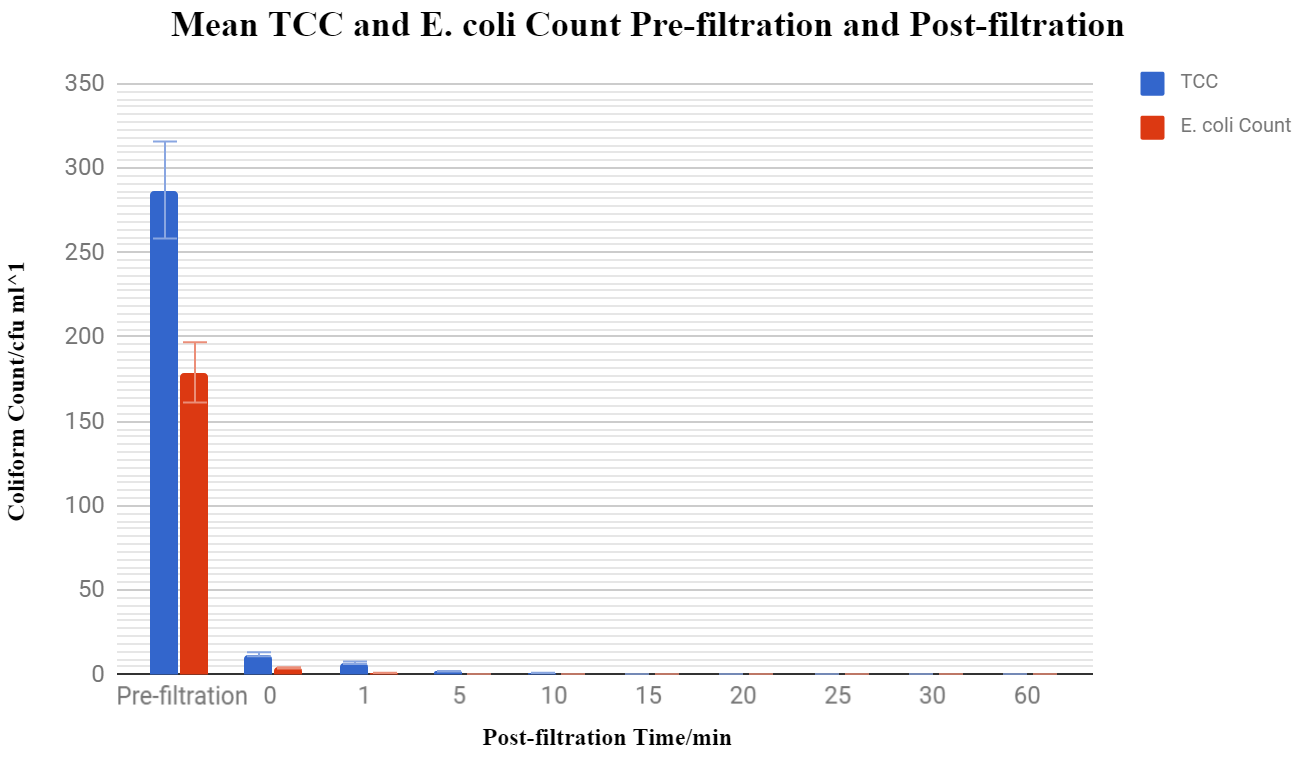


Figure 10. shows the changes in the mean TCC of water samples pre-filtration, and at set time intervals post-filtration.

3.3.3 Sequestration of Pb^{2+} and Ni^{2+}

Ions	Decrease in Concentration of Inorganic Ions/ppb														
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10	Trial 11	Trial 12	Trial 13	Trial 14	Trial 15
Pb^{2+}	2542	2693	2689	2661	2800	2678	2834	2587	2863	2593	2618	2545	2847	2621	2666
Ni^{2+}	2231	2189	2269	2052	2369	2286	2254	2224	2250	2189	2276	2459	2326	2185	2132

Ions	Decrease in Concentration of Inorganic Ions/ppb	
	Mean ($\bar{x}$)	**Standard Deviation (s)**
Pb^{2+}	**2671**	**113**
Ni^{2+}	**2234**	**91**

Table 1. a) shows the calculated decrease in Pb^{2+}/Ni^{2+} concentrations for each trial and **b)** shows the mean and standard deviation of the decrease in Pb^{2+}/Ni^{2+} concentrations.

4. DISCUSSION

4.1 Identification of Water Contaminants

The filtration of $PbCl_{2\ (aq)}$ through the biocomposite structure resulted in the rapid formation of yellow structures on/in the sample-retaining bioplastic whilst $NiCl_{2\ (aq)}$ resulted in the formation of green structures (Figure 6b & 6c). Interestingly, both colour changes were not reversed even after washing the bioplastic unit, suggesting that the structures may be immobilised complexes formed between the limonene and heavy-metal ions. The resistance of the structures to washing allows the smartphone-based computer vision algorithm to identify the exhibited chromogenic responses and ascertain the types of heavy-metals present.

The RF module was initially provided with 10-dimensional data extracted from the images of the bacterial and heavy-metal water samples to allow the children trees to identify deeper underlying trends between image features. However, PCA (Figures 7c, 8c & 9c**)** was performed to reduce the resource intensiveness of the algorithm on smartphone systems. The absence of literature concerning smartphone-based water analysis in real-time suggests that this study may be the first to explore the potential of such technologies. Table 2 shows the accuracy at which the smartphone-based computer vision algorithm can identify water-based contaminants from post-PCA image samples.

Sample Type	Accuracy/%
Escherichia coli (K-12 strain)	83.2
Sarcina aurantiaca	89.6
Vibrio fischeri	85.6
Pb^{2+}structures	97.3
Ni^{2+} structures	95.1

Table 2. shows the accuracy at which the computer vision algorithm identified different types of markers present as, or formed as a result of, contaminants within water samples.

4.2 Purification of Water

The mean pre-filtration TCC was calculated to be 287 cfu/ml (100%). Immediately after filtration, the mean TCC decreased to 13 cfu/ml (5%), and further decreased to 0 cfu/ml (0%) by 10 minutes. The TCC was observed to remain at 0 cfu/ml after 10 minutes post-filtration. Biocomposite filtration is thought to immediately filter coliform bacteria which do not have the correct orientation, or are too large, to pass through the filter. Furthermore, the natively antimicrobial casein hydrolysate within the biocomposite is believed to stimulate the complete inactivation of remaining coliform bacteria within the filtrate, thus resulting in the 100% inactivation of coliform bacteria within the water within 10 minutes.

The peak absorbance wavelength of Pb-dithizone complexes was found to be 502nm and the absorbance of the Pb-dithizone complexes is reported to follow Beer's Law (Equation 4) in the concentration range of 0.001-10 mg l^{-1} [19], where the molar extinction coefficient constant for Pb-dithizone chelate solution, ε, was graphically calculated to be 1.90×10^4 l mol^{-1} cm^{-1} and the path length, b, is 1cm. After filtration, the Pb^{2+} concentration showed a mean decrease of 2,671 parts per billion (ppb), which is approximately 178 times the actionable lead level of 15 ppb issued by the United States' Environmental Protection Agency [4].

$$A = \varepsilon bc \tag{4}$$

Ni-EDTA complexes displayed maximum absorbance at 380nm and the absorbance of Ni-EDTA complexes is also reported to follow Beer's Law in the concentration range of 0.030-2.730 mg l-1 [20], where ε for Ni-EDTA complexes was graphically calculated to be 5.86×104 l mol-1 cm-1 and b is 1cm. A mean decrease of 2,234 ppb of Ni^{2+} was recorded after filtration of the Ni^{2+} samples through the biocomposite.
The null hypothesis was that there would be no significant difference in the mean decrease of Pb^{2+} concentration compared to Ni^{2+} concentration. The t-test (Equation 5), where $\bar{x}$ is the mean decrease in ion concentration, s is the standard deviation for the decrease in ion concentration and n is the number of trials (Table 1), was applied to test the null hypothesis:

$$t = \frac{(\bar{x}_{Pb} - \bar{x}_{Ni})}{\frac{(s_{Pb})^2}{n} - \frac{(s_{Ni})^2}{n}} \tag{5}$$

The calculated value for the t-test was 25.26, which is greater than the critical value of 3.674 at the 0.1% significance level at 28 degrees of freedom. Therefore, $p \leq 0.001$ and we can reject the null hypothesis. The significant difference between the decrease in Pb^{2+} concentration compared to Ni^{2+} concentration after filtration may potentially be explained by lead (II) being more reactive than nickel (II), thus resulting in lead (II) binding to biocomposite limonene more readily than nickel (II). The maximum Pb^{2+} and Ni^{2+} removed by biocomposite filtration is believed to be limited by the relatively short time during which the water is in contact with the biocomposite. Moreover, the surface area of the biocomposite in contact with the water sample is also believed to limit the extent of Pb^{2+} and Ni^{2+} sequestration, due to a greater surface area correlating with increased exposed limonene binding sites.

4.3 Examples of Applications

The system designed in this study demonstrates the potential to be a rapid, portable and sustainable unit for the simultaneous analysis and purification of water. The system is envisioned to assist citizens and scientists, in both developed and developing regions, due to the biocomposite-enabled water filtration being a passive process which does not require input energy and the smartphone-based water analysis only requiring a functioning smartphone. Moreover, the system is inexpensive and automated, thus increasing its accessibility and operability by removing the necessity for excessive human input. In addition to providing potable water, the water analysis and purification unit may also serve field-medics by aiding disease diagnosis through the identification of waterborne illnesses, such as heavy metals and pathogenic bacterial contaminants.

The smartphone-based and biocomposite-enabled water analysis and filtration also enables citizen scientists to easily analyse samples from local water sources; ultimately a global map of water quality may be built by collating the data generated by widespread analysis of local water sources. Such a publicly-available, citizen-built global map could provide the means to develop sustainability measures by analysing changes in water quality over time on a local, national and international scale. For example, if a hazardous contaminant is independently found to be present in a local water source on numerous occasions, and that data is publicly available, data concerning that particular contaminant can be cross-referenced with local activity in the region to quickly identify and address potential sources of contamination.

The author understands that whilst widespread, smartphones are not universal and certain communities may not be able to perform the smartphone-based analysis of water samples. However, the system designed in this study still allows for sustainable water purification without a smartphone, thus ensuring that potable water is made accessible to as many individuals as possible. Whilst the aforementioned system has several point-of-

use applications, the biocomposite engineered in the study could remove water contaminants from water to be distributed to swathes of the population, without resulting in harmful byproducts, such as chlorinated hydrocarbons, which often result from the standard practice of water chlorination. Similarly, the biocomposite may serve as an inexpensive and sustainable alternative to traditional household filters, which oftentimes do not remove heavy-metals.

4.4 Safety and Sustainability

Cellulose, the primary constituent of the biocomposite, is often cited as the world's most abundant organic resource. Moreover, cellulose is an insoluble dietary fibre that is not digested or absorbed by humans and several forms of cellulose have been categorised by the U.S Food and Drug Administration (FDA) as generally recognised as safe (GRAS). An antimicrobial feature of the biocomposite is derived from the innate properties of fragmented casein. Casein is found in mammalian milk, is commercially sold as a health supplement and many casein-containing products are FDA GRAS certified. Similarly, glycerol is widely accepted as non-toxic and is FDA GRAS certified.

Unlike Nithiyanandam's bioplastic [6], the biocomposite developed in this study does not require titanium dioxide, thus reducing the safety and environmental concerns commonly associated with nanoparticles. Urea is used for the swelling of the cellulosic fibres, however it is later removed in the drying stage. Nevertheless, further toxicity studies must be conducted to thoroughly evaluate the safety of the biocomposite.

4.6 Cost Analysis

Raw Material	Cost per kg (£)	Mass required for purification of 1 litre of water (kg)	Initial* cost of purification of 1 litre of water (£)
Microcrystalline cellulose	3.55	0.0100	0.035
Cellulose fibre	5.67	0.0140	0.079
Casein hydrolysate	15.43	0.0120	0.185
Urea	47.20	0.0013	0.061
(R)-(+)-Limonene	45.50	0.0029	0.131
Glycerol	3.80	0.0672	0.255
0.2M NaOH solution	26.38	0.0021	0.055

NaCl	2.10	0.0330	0.069

Table 3. shows the raw cost of materials for bioplastic & biocomposite synthesis.

*****The initial cost of materials required to synthesise 142.5g of biocomposite (*i.e.* the mass of biocomposite required to purify 1 litre of water) is £0.87 (GBP). However, with usage, the cost of biocomposite for purification per litre of water is expected to decrease from £0.87, due to the ability of the bioplastic to withstand several usage cycles.

5. CONCLUSIONS

1. The novel cellulosic biocomposite developed in this study effectively filtered solid water contaminants and coliform bacteria. This conclusion is supported by the 100% removal of solid particulates and a 96% decrease in mean TCC immediately after filtration.
2. Further bacterial inactivation studies supported the hypothesis that bacterial inactivation would continue to occur within the filtrate; the mean TCC decreased to 0 cfu/ml by 10 minutes post-filtration. Whilst the mechanism of post-filtration bacterial inactivation is unknown conclusively, the author hypothesises that the mechanism relies on natively antimicrobial casein hydrolysate.
3. The biocomposite demonstrated Pb^{2+} and Ni^{2+} sequestration capabilities. Pb^{2+} and Ni^{2+} sequestration is supported by qualitative observations of Pb^{2+} and Ni^{2+} deposit formation.
4. Furthermore, quantitative analysis employing UV-Vis spectrophotometry demonstrated a post-filtration decreased in Pb^{2+} and Ni^{2+} concentration of 2,671 ppb and 2,234 ppb respectively. The value obtained for the t-test demonstrated there was a significant difference between the decrease in Pb^{2+} concentration compared to Ni^{2+} concentration at the 0.1% significance level.
5. A cheap glass-bead magnification system for smartphones was able to produce a real-time image of water samples that could be assessed to identify potential contaminants.
6. A novel random forest schema, in combination with a HoG function, was trained using 568 labelled images in a supervised learning approach. The schema was trained to identify 3 different species of bacteria: *E. coli (K-12)*, *S. aurantiaca* and *V. fischeri.*
7. Principal component analysis was able to compress 10-imensional data presented to the classifier into fewer dimensions to decrease resource demands during image processing.
8. Image bright regions, dark regions, halo artifacts and noise were used as principal features for image analysis. A probability threshold of 0.8 was set after assigning weighted values for the

aforementioned image features; the algorithm either classified bacteria as *E. coli (K-12), S.* aurantiaca, V. *fischeri* or 'unrecognised'.

9. The accuracy at which the algorithm identified bacterial species and heavy-metal deposits from images varied by species and heavy-metal deposit type: *E. coli (K-12), S.* aurantiaca, V. *fischeri*, Pb^{2+} deposits and Ni^{2+} deposits were identified with accuracies of 83.2%, 89.6%, 85.6%, 97.3% and 95.1% respectively.
10. The cost of the raw materials for synthesising 142.5g of biocomposite is £0.87. However, it is expected that the purification cost per litre of water will decrease over time with usage.

6. REFERENCES

1. United Nations Sustainable Development. (2016). High-Level Political Forum on Sustainable Development 2017. [online] Available at: http://www.un.org/sustainabledevelopment/sustainable-development-goals/ [Accessed 13 Dec. 2016].
2. UNDP. (2018). Millennium Development Goals. [online] Available at: http://www.undp.org/content/undp/en/home/sdgoverview/mdg_goals.html [Accessed 1 Sep. 2017].
3. World Health Organization. (2016). Drinking-water. [online] Available at: http://www.who.int/mediacentre/factsheets/fs391/en/ [Accessed 14 Dec. 2016].
4. US EPA. (2016). Basic Information about Lead in Drinking Water | US EPA. [online] Available at: https://www.epa.gov/ground-water-and-drinking-water/basic-information-about-lead-drinking-water [Accessed 14 Dec. 2016].
5. Crockett, M., Evans, A., Worthington, M., Albuquerque, I., Slattery, A., Gibson, C., Campbell, J., Lewis, D., Bernardes, G. and Chalker, J. (2015). Sulfur-Limonene Polysulfide: A Material Synthesized Entirely from Industrial By-Products and Its Use in Removing Toxic Metals from Water and Soil. Angewandte Chemie, 128(5), pp.1746-1750.
6. Nithiyanandam (2017). A Novel, Photocatalytic, Lead-Sequestering Bioplastic for Sustainable Water Purification and Environmental Remediation. Unpublished manuscript.
7. Jeswani, Hansa & Oke, Ninad & Mahapatra, Nikita & Sayyed, Rumana & Choksi, Parth & Naik, Amey. (2014). Use of Pervious Concrete as Gravity Filter. 10.13140/2.1.3651.0408.
8. Sangwan, S., Kedia, Jyoti., Kedia, Deepak (2013). A COMPARATIVE ANALYSIS OF DIFFERENT CMOS LOGIC DESIGN TECHNIQUES FOR LOW POWER AND HIGH SPEED. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2(10), pp.2627-2634.

9. J. Pan, T. Kanade, and M. Chen, “Learning to detect different types of cells under phase contrast microscopy,” in Microscopic Image Analysis with Applications in Biology Workshop, 2009.
10. Kim, H., Pang, S., Je, H., Kim, D. and Yang Bang, S. (2003). Constructing support vector machine ensemble. Pattern Recognition, 36(12), pp.2757-2767.
11. Breiman, L. (2001). Machine Learning, 45(3), pp.261-277.
12. Li, R., Wang, S., Lu, A. and Zhang, L. (2015). Dissolution of cellulose from different sources in an NaOH/urea aqueous system at low temperature. Cellulose, 22(1), pp.339-349.
13. Mooragar.com. (2017). [online] Available at: https://www.mooragar.com/wp-content/uploads/2014/07/LB-Lysogeny-Broth-Media-Recipe.pdf [Accessed 21 Nov. 2017].
14. Xiong, Bi & Zhao, Pingping & Hu, Kai & Zhang, Lina & Cheng, Gongzhen. (2014). Dissolution of cellulose in aqueous NaOH/urea solution: Role of urea. Cellulose. 21. 1183-1192. 10.1007/s10570-014-0221-7.
15. Pnnl.gov. (2018). PNNL: News - Want to print your own cell phone microscope for pennies? Here’s how.. [online] Available at: https://www.pnnl.gov/news/release.aspx?id=1071 [Accessed 9 Oct. 2017].
16. User.engineering.uiowa.edu. (2017). [online] Available at: http://user.engineering.uiowa.edu/~image/reading/Slides-Milan-shb-4-RFs-classification-chapter.pdf [Accessed 14 Sep. 2017].
17. T. S. Xu, H. D. Chiang, G. Y. Liu and C. W. Tan, "Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data," in IEEE Transactions on Power Delivery, vol. 32, no. 2, pp. 609-616, April 2017. doi: 10.1109/TPWRD.2015.2479941
18. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in CVPR, June 2005, vol. 1, pp. 886–893.
19. Multimedia.3m.com. (2018). [online] Available at: http://multimedia.3m.com/mws/media/1388563O/petrifilm-coliform-count-plate-interpretation-guide.pdf [Accessed 12 Dec. 2017].
20. Humaira Khan, M. Jamaluddin Ahmed, and M. Iqbal Bhanger (2006). A simple spectrophotometric method for the determination of trace level lead in biological samples in the presence of aqueous micellar solutions. Spectroscopy, 20: 285-297.
21. Carpani , I. , E. Scavetta , and D. Tonelli . 2004 . Spectrophotometric determination of aluminium and nickel . Ann. Chim. 94 : 365 .