https://doi.org/10.48048/tis.2024.8528

Supporting Information

Critically Reckoning Spectrophotometric Detection of Asymptomatic Cyanotoxins and Faecal Contamination in Periurban Agrarian Ecosystems via Convolutional Neural Networks

Soumyajit Koley

Department of Civil Engineering, Hooghly Engineering & Technology College, West Bengal, India

(Corresponding author's e-mail: samkoley5@gmail.com)

S1. Latest case studies

Flow cytometry has recently been used for low-voltage scanning electron microscopy (SEMs). The primary objective of this study was to investigate and elucidate the properties of objects found during platelet concentrate storage using scanning electron microscopy (SEM). Extremely sensitive technology beyond standard optical microscopy is necessary for the detection of particles with sizes ranging from 25 nm to 700 nm. Coatings of conductive metals on materials as thin as small extracellular vesicles are required by conventional scanning electron microscopes (SEMs). However, low-voltage scanning electron microscopy techniques have been developed to eliminate this requirement and make it possible to investigate objects smaller than individual cells. The primary benefit of scanning electron microscopy (SEM) is its ability to capture high-resolution images. These images provide machine-learning algorithm-specific and important data for the classification. The ability to detect and achieve higher levels of accuracy may be greatly enhanced by incorporating other data obtained from the microorganisms under study, such as dry mass and fluorescence measurements. Despite their superior resolution, scanning electron microscopy (SEM) images have many drawbacks that make them unsuitable for microscopy of water quality on a broad scale. Scanning electron microscopes (SEMs) are easily identifiable owing to their massive stature, high cost, and the necessity of a space devoid of electric, magnetic, or vibrational interference. In addition, an efficient cooling water circulation requires a consistent voltage. As an additional precaution, only trained professionals with proper certification should operate this machinery. These limitations greatly affect the use of SEM for water quality monitoring because it prohibits its use for in situ applications owing to the danger of equipment damage. It is hoped that the SEM technology will become more useful as it evolves. Therefore, it is essential to assess new technical developments. The introduction of SHeM is a revolutionary method for nondestructive microscopy using helium. A picture of the surface topography and local chemical composition can be generated using this technique by interacting with a neutral helium atom in its lowest energy state with the surface of the specimen. Crucially, this approach did not compromise the integrity of the sample. A tightly focused beam of atoms was created by the SHeMs by directing an ever-thinner helium gas stream via nozzles, which in turn created an image. When atoms smash onto a sample surface, the amount of energy they return tells us a lot about the topography and chemical makeup of the area. The maximum achievable resolution of the SHeM pictures is determined by the intrinsic wavelength of the helium atoms, which is approximately 0.05 nm. The chemical composition and texture of the target surface can be detected at this level, which aids in image

identification. This theoretical resolution is yet to be attained for practical designs. The long image collection time, sample vacuuming requirement, and limited spatial resolution offered by the current technology renderer, SareHeM, are unsuitable for water quality monitoring, despite their high theoretical resolution and helpful local chemical information. Accelerating picture capturing and boosting spatial resolution are the goals of ongoing research to improve the desired signal sources and increase the sensitivity of the detectors. Future advancements in SheM should be thoroughly evaluated in terms of water quality, which requires constant monitoring and analysis. Recent practical uses of artificial neural networks include biometric security and self-driving automotive navigation, and they are making steady strides in the field of image recognition. Convolutional Neural Networks (CNNs) are among the most widely used neural network topologies for cell recognition applications. Biological datasets that are too large to be manually examined can be analysed more quickly and objectively using artificial neural network recognition. Adding image-processing techniques, such as segmentation, focus classification, and deblurring algorithms, to picture identification neural networks can increase their efficiency. After the processing was complete, neural networks were employed. Because each node in the network may be trained to perform a particular job accurately and quickly, using post-processing and recognition algorithms together increases the total efficiency of neural networks. Prior to being fed into the identification algorithms, the raw picture data may be segmented using a neural network to isolate the biologically relevant area and remove any optical defects, such as dust or droplets. Reducing the number of irrelevant pixels that must be considered can improve the efficiency of future recognition systems. In its place, computing resources are zeroed in investigating targeted areas of biology. The automatic segmentation of small pictures of Bacillus anthracis with high accuracy was accomplished by Abiri et al. [1] using UNet. Because of this method, a crucial area of biology is located. This method achieves a segmentation accuracy of 97 % and avoids the need for a human expert to evaluate scanned materials. To reduce the chances of the algorithm being affected by optical aberrations or debris, the segmented image can be processed using a neural network that has been trained specifically for bacterial identification. Segmentation algorithms are crucial for high-throughput imaging approaches such as the ultralarge-scale methodology described by Bucki et al. [2]. The reduced overall data storage requirements are the outcome of these algorithms' efficient data storage through the deletion of undesirable background areas. Picture segmentation techniques are crucial for post-processing to save and manage large-scale cyanobacterial images.

Table S1 Overview of the analytical methods currently employed for monitoring bioreactors.

	Analytical Information	**Advantages**	**Disadvantages**
Gas sensors, such as the Respiration Activity Monitoring System (RAMOS)	Partial pressure of gases (O_2 and CO_2)	Can determine the oxygen transfer rate (OTR), which is directly related to the particular growth rate.	There is no precise information available on the total signal that describes metabolic rate.
Electrochemical culture medium sensors	Enzyme modification results in a signal that is directly proportional to the concentration of a certain molecule, such as ethanol.	Rate of synthesis of a certain product can be enhanced by modifying enzymes using targeted analytics.	Enzyme activity is influenced by changes in temperature and pH.

	Analytical Information	Advantages	Disadvantages
Absorption based optical sensors	Extinction is directly proportional to concentration when it is low, such as in the case of proteins or pigments.	Cell concentration may be quantified, together with qualitative data on photoactive chemicals.	At larger cell concentrations, it is necessary to do sample preparation or dilution.
Scattering based optical sensors	Light that is dispersed by cells (the decrease in signal produced by scattering may also be quantified)	Cell concentration may be measured online.	Disruptions caused by the process of adsorption or the clumping together of cells.
Fluorescence based optical sensors	Fluorescence emitted by certain substances present in the medium or organism	Specific product information	At greater cell concentrations, it may be necessary to do sample preparation or dilution. In such cases, correction factors may need to be added.

There are very few cases of out-of-focus cells or motion blur in microscopic images obtained using flow cytometry because of the technique's tendency to capture cells in motion and its small depth of field. should remove these photos before giving them to recognition algorithms, because they affect how well they work. Deblurring techniques such as DeepFocus are useful for making motion-blurred photographs clearer again. In addition, DeepFocus and similar algorithms can detect blurry images and crop them off. Similar to how DeepFocus was developed to enhance diagnostic efficiency by removing fuzzy tissue cell shots, cyanobacterial images may also be located and fuzzy-free using a similar technique. **Table S1** lists several microscopic methods that use flow cytometry to obtain high-resolution images of live cells during motion. Data collection rates and sample processing speeds were enhanced by increasing cell throughput. With respect to ultrahigh-throughput QPI systems, motion blur may degrade the image quality, especially at higher sheath fluid flow rates. The use of generative adversarial networks (GANs) can enhance the clarity and sharpness of photos captured using fast-scanned or motion-blurred imaging equipment because of variations in the imaging settings and optical setups. Non-blind deblurring may have limitations because it relies on established methodologies tailored to certain optical systems. Blind deblurring is an extremely flexible method that uses a trained Convolutional Neural Network (CNN) or Generative Adversarial Network (GAN) to restore sharpness to a blurred image quickly. Using an end-to-end deblurring technique, DeblurGAN directly converts blurred photographs into clear photographs using a trained CNN with a different network architecture. According to this method, DeblurGAN is faster than the competing techniques. Using a Convolutional Neural Network (CNN) to forecast a sharp image from a hazy training picture is a training approach that does not require any previous data from the optical system. Additionally, to evaluate the performance of the CNN, a Generative Adversarial Network (GAN) was trained to compare projected and real sharp pictures. Improving network performance and image deblurring rates is the goal of adversarial training for generative adversarial networks (GAN) and convolutional neural networks (CNN). Higher enumeration and taxonomy identification

accuracy are direct outcomes of using deblurring techniques, which are especially effective in high-throughput imaging systems.

Table S2(a) Isolation of *Listeria monocytogenes* from healthy domestic animals.

Country	Animal	Target	Sample (n)	Positive (%)
Slovenia	Cows Calves	*L. monocytogenes*	Feces (544)	19.2[1]
Spain	Cattle	*L. monocytogenes*	Feces (956)	5.3
	Sheep	*L. monocytogenes*	Feces (488)	6.8
	Goat	*L. monocytogenes*	Feces (335)	1.3
Switzerland	Cattle	Ab to LO and IA	Serum (1858)	21
Taiwan	Chicken	*L. monocytogenes*	Carcass rinse (269)	13.4
USA	Cattle	*L. monocytogenes*	Feces (828)	41[2]
USA	Cattle	*L. monocytogenes*	Feces (528)	29.2[3]
USA	Broiler	*L. monocytogenes*	Feces (535)	18.9
Japan	Pigs, Dogs, Rats	*L. monocytogenes*	Feces (1755)	1.78
Japan	Cattle	*L. monocytogenes*	Feces (1728)	6.3
Jordan	Cattle	*L. monocytogenes*	Feces (680)	5.6
Qatar	Camel	*L. monocytogenes*	Feces (80)	4.4

[1]Stool samples were collected from cows and calves on 20 dairy farms owned by families at 2-week intervals over the course of the year.
[2]Antibodies against listeriolysin O and internalin A
[3]A total of twenty-five fecal samples were collected daily during two 2-week periods and one 5-day period.
[4]A case-control study was conducted to evaluate the spread and ecology of Listeria monocytogenes in farms. The study included 24 case farms that had at least one recent listeriosis outbreak and 28 control farms that had no outbreaks.

Image and pattern recognition make extensive use of Convolutional Neural Networks (CNNs), which are a type of artificial neural networks. CNNs have several advantages over other types of neural networks, which is why they are becoming increasingly popular. Many biological image identification applications have used neural networks extensively. These applications range from microbe recognition and illness diagnosis on the skin and leaves, to cancer cell detection and tissue categorisation. When it comes to identifying images, CNN frameworks are resilient to distortions caused by variations in lighting, optical artefacts, and picture perspectives. The structural design of CNNs makes them more efficient with resources because it decreases the amount of memory required by the algorithm. Compared to other neural networks with similar capabilities, CNNs often have a far faster training procedure because of their reduced number of network parameters. In this study, it was imperative to review the framework and methods for identifying CNNs. Similar to real synapses, convolutional neural networks (CNNs) are artificial neural networks that only fire when the sum of all inputs reaches a certain threshold. When they spot a large number of recognisable elements inside an image, they react to this as a recognition stimulus. An input layer, output layer, and many hidden recognition layers are the standard components of a conventional CNN. Notable examples of stacked layers include the convolutional, nonlinear, and pooling/subsampling layers. CNNs

designed for specific tasks often have distinct architectures tailored to their specific roles. To obtain valuable abstract information, the convolution layer analysed each individual picture block. To represent the input as a matrix, the network convolution layer uses either visual or physical information. By evaluating the matrix with a smaller filter matrix, known as a kernel, significant information can be extracted from the picture, resulting in a feature map.

The structural parameter of a CNN that influences the amount of detail recovered from an image is the kernel stride. Distinct feature maps that reflect distinct sets of sample data can be generated by using different kernels. The subsampling layer uses feature map pooling methods, including minimum pooling, maximum pooling, and average pooling, to locate objects that are similar across several pictures. The neural network sorts statistically related pictures into predefined groups by using common abstract patterns in the feature maps. A multiclass or binary structure for the classes can be used in a neural network trained to monitor water quality. There is a unique cyanobacterium species in each multiclass configuration class for both toxic and non-toxic cyanobacteria. Recognition algorithm performance is heavily affected by categorisation structure, which is achieved by combining convolutional neural networks (CNNs) with quantitative phase imaging (QPI). One exciting area of study is the integration of quantitative phase imaging (QPI) with convolutional neural network (CNN) techniques for the accurate cell detection and classification of a specific sample. Machine-learning systems that identify specimens using previously learned examples face substantial challenges when cyanobacteria are detected in a sample. Cyanobacteria do not have sexually reproducing traits that provide taxonomic information, are extremely variable in size, or exhibit polymorphisms. Consequently, to help identify each subspecies, extensive training sets are required to account for all variations. The detection and subtyping of white blood cells into B and T cells were accomplished using biomedical machine learning methods. To train a neural network to differentiate between different cell types, researchers first mapped the 3D refractive indices of cells. A trained convolutional neural network (CNN) can recognise and differentiate cyanobacteria from other microbes and non-biological trash in a water sample, even though white blood cells are larger than tiny cyanobacteria, using the data provided by a high-resolution QPI device. A Convolutional Neural Network (CNN) called 'HoloConvNet' was created by Brito et al. [3] using MATLAB. The primary motivation for developing this CNN is to enhance species identification by identifying crucial biological features using Quantitative Phase Imaging (QPI) data. Using a dataset that included four other bacterial strains with comparable but distinct morphological features, the HoloConvNet method successfully identified Bacillus anthracis. At 1 ms per cell, the algorithm achieved recognition accuracies of up to 96.3 %. As B. anthracis and smaller cyanobacteria cells are similar in size, this analytical approach is useful for determining water quality. Based on these data, it seems that this approach can detect very small cells, and a new bacterial strain for neural networks called *Listeria monocytogenes* was detected and classified by researchers using HoloConvNet. Without modifying the system, HoloConvNet achieved an accuracy of 85 % in recognising *L. monocytogenes*. For every species encountered, the program identified the crucial biological characteristics. This finding demonstrates the adaptability of neural networks in detecting novel species with little to no human input. To achieve a far greater level of accuracy in identifying this particular strain, researchers have suggested modifying the architecture and learning parameters of the HoloConvNet. The automatic detection of *L. monocytogenes* was a breeze using HoloConvNet. This was achieved by quickly training the CNN with raw QPI data after recognising the key biological features in the given images. To classify images into different groups, HoloConvNet, similar to other CNNs, gradually modifies the input through many convolution layers to produce a hierarchical representation of the picture.

As shown In **Figure 3(b)**, difficulty arises when there are several different species of cyanobacteria in which a Convolutional Neural Network (CNN) is used to identify and distinguish between faces. With a

binary-class setup that included three species, HoloConvNet could detect anthrax with a high recognition accuracy of 96.3 %. However, in a multiclass context with five species, the recognition accuracy decreased to approximately 61 %. In addition, the identification accuracy test results demonstrate that CNNs work best when trained to distinguish between two distinct types of data, anthrax and non-anthrax, in the case of HoloConvNet. This demonstrated that methods for identifying cyanobacteria can improve their accuracy. The convolutional neural network (CNN) was trained to classify specimens in QPI photos into two broad groups: those with a high potential for harm and those with a lower risk, for example, cyanobacteria and non-cyanobacteria.

Table S2(b) Isolation of *Listeria monocytogenes* from healthy wild animals.

Country	Animal	Target	Sample (n)	Positive (%)
China	Rodents	*L. spp.*	Feces (341)	9.6
		L. monocytogenes		3.5
		L. innocua		4.9
Finland	Birds	*L. monocytogenes*	Feces (212)	46
Norway	Reindeer	*L. monocytogenes*	Feces (470)	6.2
		L. monocytogenes		53
France	Rooks	*L. innocua*	Feces (112)	34
		L. seeligeri		8.8
		L. monocytogenes		7.8
Germany	Pigeons	*L. innocua*	Feces (350)	6.3
		L. seeligeri		2.6
Japan	Crows	*L. innocua*	Feces (301)	1.82
Kenya	Nile tilapia	*L. monocytogenes*	Muscle (167)	1.8
Poland	Red deer	*L. monocytogenes*	Feces (120)	1.95
	Red fox	*L. monocytogenes*	Rectal swab (286)	3.7
Poland	Beech marten, Racoon	*L. monocytogenes*	Feces (65)	6.4
Switzerland	Wild boars	*L. monocytogenes*	Feces (70)	4.7
USA (Central New York)	Reptiles Mammals	*L. monocytogenes*	Tonsils (153)	17.8[5]
	Birds	*L. monocytogenes*	Feces (17)	19

[5]Tonsils, rumen (stomach contents), liver, intestinal lymph nodes, caecum contents, and faeces were collected.

Complete understanding of the subjects was achieved using this method. Here, ‘DeepNN’ means a deep neural network, ‘ConventionalNN’ is a conventional (single layer) network, and ‘Brightfield’ is an image taken using traditional brightfield microscopy. This information was derived from a study by Bretas et al. [4] who investigated the potential risks to human health and the environment caused by toxic algal blooms. Efforts have been made to assess the presence of harmful cells to reduce the required computing

resources. Compared to conventional microscopy, the additional information obtained with QPI is more valuable (see **Figure 3(c)**). In every test setup, the QPI data outperformed the standard brightfield microscopy data in terms of the recognition accuracy. A binary-class test involving three species showed an accuracy improvement of 15 % when using QPI. Both the binary and multiclass tests involving the five species exhibited the same trends. According to these results, the imaging method is more crucial than the neural network when comparing the deep and regular neural networks. More study is needed to make a direct comparison between AI learned using QPI (Quantum Phase Imaging) and AI trained with regular pictures in terms of accuracy. These results were confirmed with the help of this study. Similar to how other neural network types are trained, Convolutional Neural Networks (CNNs) are trained using the same concepts and methodologies. With the use of machine learning, artificial neural networks may be trained to mimic the recognition of biological neural networks. Layers of neurones connected by synapses with different weights form these networks. When the sum of all the inputs reaches a specific level, the activation function regulates how strongly the neurones fire. The cost function of the method can be determined by comparing the expected and actual outputs of the network, which are generally expressed as the mean squared error. The output of the algorithm is found to be inaccurate. The network can grow repeatedly across several epochs by backpropagation, which involves modifying the weights of synapses. The neural network canfies pictures that were previously unseen because of the knowledge gain made possible by incremental adjustments. Because many cycles are required to create adaptable neural connections, training neural networks is computationally difficult and time consuming. Supervised machine learning is a common method for teaching neural networks that can recognise images and entails feeding algorithm examples of recognised items to help them learn how to classify them. Subsequently, a different dataset was used to test the recognition abilities of the algorithm. Recognising images using an artificial neural network can save manual labour, but it takes a lot of work up front to train the algorithms on specific samples and identify the specimens to be used for training. One way to increase the size of a training set is to digitally alter pre-classified pictures; this is called training set augmentation. Random pixel value changes were used to rotate, flip, and noise the identified training images. With this technique, a single data point, such as a picture of a harmful cyanobacterial cell, can be enhanced to generate more than hundred distinct training examples. These examples provide several viewpoints regarding the use of neural networks to improve machine learning. Training convolutional neural networks (CNNs) still requires considerable time and resources, even if their network design is better than that of other neural networks. It usually takes researchers approximately two weeks to train image recognition neural networks to identify germs on computers with dedicated GPUs. A minimum of one thousand to ten thousand photos was required to attain good accuracy in every trainable category. 50 - 80 % of these images were used for training, and 15 - 20 % were used for testing. By utilising the data augmentation methodologies previously discussed, the total number of unique images required to teach a certain class may be drastically reduced. Nevertheless, it is extremely challenging to train a Convolutional Neural Network (CNN) that can accurately identify and evaluate the viability of a wide variety of cyanobacteria species.

Table S3 *Listeria spp.* were found in samples collected from farms and their surroundings.

Country	Target	Sample (n)	Positive (%)
Egypt	*L. monocytogenes*	Water (36)	11.6
		Silage (36)	35.8
		Manure (36)	29.4
		Soil (36)	9.3
		Milking equipment (432)	7.9
Iran	*L. monocytogenes*	Water (180)	17.7
Iran	*L. monocytogenes*	Iranian currency (108)	1.93
Jordan	*L. monocytogenes*	Bulk tank milk (305)	8.5
New Zealand	*L. monocytogenes*	Bulk tank milk (400)	5.0[6]
South Africa	*L. monocytogenes*	Roof-harvested rainwater (264)	32
Taiwan	*L. monocytogenes*	Abattoir environment (246)	14
USA(Colorado)	*L. monocytogenes*	Soil (555)	17.3
	L. spp.[7]	Stream water (196)	38
	L. monocytogenes	Soil, water, sediment, surface soil and wildlife	1.23
USA (Idaho)	*L.* spp.	fecal samples (572)	2.5
	L. monocytogenes	Dairy wastewater ponds (30)	7.7

[6]A poll was performed between November 2011 and August 2012. During this time, 25-mL milk samples were collected five times from each of the 80 randomly chosen dairy farms. This survey was designed to detect the presence of L. monocytogenes in these samples.

[7] Excluding *L. monocytogenes*.

S2. Measurement principles

Algae and their biological components can be used as alternatives to fossil fuels and synthetic goods to generate beneficial chemicals and active compounds in an eco-friendly and effective manner. Through the use of microalgae biotechnology, compounds such as polyunsaturated ω-3 fatty acids, natural colourants such as phycocyanin and chlorophyll, and powerful antioxidants such as astaxanthin, lutein, and zeaxanthin have been developed. These chemicals have pharmacological relevance and are nutritionally useful. Legislative changes and increasing consumer demand for natural and healthy products are fuelling the growth of these industries. In addition, microalgal farming has the potential for greater output and may be performed on unsuitable terrain, which has two advantages over traditional agricultural production methods. Microalgae have economic potential; however, the extraction of valuable molecules from microalgae is costly and time-consuming. Consequently, choosing the appropriate strain and controlling the growth conditions are crucial for maximising algal growth rate, cell density in culture, and product content. Reducing product yield and increasing downstream purification expenses to produce a safe product are both caused by the disruption of the culture. Therefore, it is essential to carefully observe the culture to prevent bacterial loss. These biotechnological approaches require meticulous process control and monitoring. Most analytical techniques are offline operations that require the collection, preparation, or dilution of samples. **Figures 1c - 1e** provide an extensive overview of the analytical methods employed to track biological activities. It is a common practice to measure the oxygen transfer rate (OTR) while investigating the development of cell cultures from plants or microbes. The level of light in phototropic

algae culture affects the amount of oxygen produced during photosynthesis. Therefore, the growth rate can be approximated using the OTR.

One way to successfully characterise a cultivation is by examining the concentration of the substrate, such as glucose in a photomixotrophic cultivation, or the product, such as ethanol, in the culture media. Enzymes can be attached to amperometric or potentiometric sensors to achieve this. To protect enzymatic sensors from environmental factors such as temperature and pH, specialised non-enzymatic sensors were created. Although there are sensors that perform beneficial work, bioreceptor-based sensors, such as enzymes, are designed to determine the quality of a particular molecule dissolved in the culture medium. Infrared spectroscopy, visible spectroscopy, and light scattering are just a few examples of the many ways in which optical sensors can gather information on pigments and biomass. UV-Vis spectroscopy is an inexpensive method for determining the amounts of substrates, metabolites, and products. A useful method for determining the concentration and purity of phycocyanin (CPC). On the other hand, turbidity is an issue because it decreases the precision when the concentration of cells increases. Although this approach is most effective at low doses, CPC fluorescence can be used to determine its concentration. Raman spectroscopy is a dependable method for evaluating chemicals, particularly for opaque materials. It offers a different approach that requires less sample preparation but is more expensive. However, the Raman effect is not very strong and, similar to other optical approaches, a multicomponent complex matrix results in overlapping spectra that are difficult to decipher. Both the material and vast array of metabolites generated throughout development contribute to the complexity of the sample matrix. Instead, signs of volatile organic compounds (VOCs) released by algae into the air above the bioreactor should be examined. By utilising the right analytical method, these signals can be effectively separated, minimising signal overlap, and the VOC profile can offer valuable insights into current metabolic processes. Variations in cell concentration due to metabolism can be better understood by analysing the VOC profile. Studies have shown that the volatile organic compound (VOC) profile of *L. platensis* is affected by the amount of light received. The developmental conditions of the plant affect the volatile organic compounds (VOCs) released by the liquid. The amount and type of volatile organic compounds (VOCs) in the photobioreactor headspace can be affected by a number of factors, including the physical and chemical characteristics of the liquid. The analysis of volatile metabolites often involves gas chromatography and mass spectrometry (MS). The use of this analytical procedure is time consuming, complicated, and expensive. Swapping out the mass spectrometer with an ion mobility spectrometer will make this investigation more budget-friendly. Ion mobility spectrometers are commonly used in security and military applications to identify explosives and other volatile and semi-volatile organic compounds (VOCs).

By utilising ambient air as the gas supply, a compact and energy-efficient apparatus can function at atmospheric pressure. The ion mobility of a sample can be determined by tracking the velocities of the ion swarms that originate from its molecules. The formation of ions from the neutral sample molecules is a crucial step in this process. Using photoionization, Corona Discharge ionisation, or radioactive ionisation sources, samples can also be ionised in air at normal atmospheric pressure. Ions were introduced into the drift area at microsecond intervals by using an electrical shutter. In most cases, the lengths of the drift area were between 5 cm and 15 cm. With the help of an electric field gradient and a gas flow (either air or nitrogen) that flows in the opposite direction to the ions, the collection of ions travels collectively towards a detector. A variety of drift velocities inside the ion swarm provides the groundwork for mass- and structure-based ion sorting. Rapid and accurate assessment of volatile organic compound (VOC) profiles is possible using ion mobility spectrometry (IMS), although the results are heavily dependent on the humidity levels of the gas sample. Gas chromatography is an effective solution for addressing problems caused by excessive humidity. The minimum amount that can be detected is affected by the limitation on the number

of samples that may be used. It has also been demonstrated that a membrane inlet [5] allows for the analysis of samples with high humidity using IMS. An easy and reliable way to determine the number of volatile organic compounds (VOCs) in a damp gas sample is to use membrane inlet (MI)-IMS. The integration of a custom-built molecular intelligence (MI) system into an ion mobility spectrometer was the primary focus of this study. Ultimately, it is crucial to analyse the VOC profile of the bioreactor automatically and in real-time. Additionally, it is important to measure the efficacy of the MI-IMS system by observing its response to variations in cell density and light intensity. The primary objective was to identify unique and consistent signal patterns across the stages of development. Such patterns may indicate the presence of algae-related illnesses, nutritional limitations, or technical failures, all of which could disrupt the production process. In this study, the cyanobacterium L. platensis was extensively cultured to fine-tune and verify various analytical tools used. Several other test locations yielded similar results for these cultivation methods. One of the many advantages of growing *L. platensis* is its bioactive compounds, which include antioxidants and anticancer agents, as well as renewable proteins and fatty acids. Phycocyanins, which are linked to proteins, are the primary pigments found in *L. platensis*. Phycocyanin is an antioxidant that is utilised in several cosmetic and medicinal products, and comprises approximately 15 % of the protein matrix. In addition, it is a food colouring agent. Therefore, in practice, it would be highly advantageous to have an easily accessible analytical monitoring system and to produce L. platensis commercially on a global scale.

Table S4 The median value and interquartile range (IQR) of the water quality data collected between March and December 2018 are shown. Surface water (SW), rapid sand filtrate (RSF), and ultrafiltration (UF) were sampled several times. When comparing surface water (SW) and riverside filtration (RSF), it is critical to consider the dilution effect of a combination of groundwater and surface water at a 20 to 80 % surface water ratio. The selected characteristics included total organic carbon (TOC), dissolved organic carbon (DOC), UV absorbance at 254 nm (UV_{254}) for both the unfiltered and filtered samples, specific UV absorbance (SUVA), humification index (HIX), fluorescence index (FI), freshness index (β:α), temperature, and turbidity.

		SW (*n* = 11)		**RSF (*n* = 16)**		**UF (*n* = 16)**	
Parameter	**Unit**	**Median**	**IQR**	**Median**	**IQR**	**Median**	**IQR**
TOC	mg L^{-1}	3.63	0.28	2.99	0.39	2.89	0.23
DOC	mg L^{-1}	3.94	0.31	2.94	0.29	2.18	0.26
[a]UV_{254} unfiltered	—[b]	12.8	1.4	9.7	0.7	4.5	0.4
[a]UV_{254} filtered	—[b]	13.1	1.7	9.2	2.1	5.4	1.4
SUVA	L mg^{-1} m^{-1}	3.5	0.3	3.6	0.4	2.2	0.6
[c]HIX	—	0.93	0.02	0.98	0.03	0.79	0.05
[c]FI	—	1.54	0.03	1.53	0.04	1.67	0.03
[c]$\beta : \alpha$	—	0.65	0.02	0.64	0.02	0.75	0.03
[a]Temperature	°C	6.0	0.7	8.0	0.8	6.7	0.8
[a]Turbidity	FNU	0.35	0.08	0.19	0.07	0.06	0.04

[a] Measured on-site. [b] Absorbance per meter. [c] HIX – Ex: 254, Em: (435-480)/((300-345) +(435-480)). FI – Ex: 370, Em: 470/520. β: α –Ex: 310, Em: 380/max(420-435).

S3. Error estimation

All around the world, freshwater ecosystems are feeling the effects of human-caused change. Significant threats to their survival include pollution, altered river flow, habitat degradation, overexploitation, and invasive species. Consequently, freshwater ecosystems are becoming increasingly susceptible to species losses. Dams, weirs, culverts, and other stream control measures have created thermal and hydrodynamic obstacles, which make it difficult for fish to traverse. Significant changes in species composition and a 90 % loss in the native fish population have been caused by obstacles that restrict fish circulation in Australia's Murray-Darling Basin (MDB), the world's largest freshwater system. Limiting fish migration physically and significantly influencing downstream water temperature are possible outcomes of controlling water flow and discharging water from large dam impoundments. A distinct issue in the MDB is cold water pollution (CWP), which occurs in the summer when large thermally stratified reservoirs leak cold water from their bottom. Water temperatures downstream of dams can be as much as 16 °C lower owing to cold water pollution (CWP). This temperature dip can last for months and spread up to 350 km downstream, and the internal temperature of aquatic ectotherms such as fish is dictated by the temperature of the water. For essential survival functions, including foraging, predator avoidance, and successful reproduction, temperature is essential because it affects physiological processes. The ideal range of environmental temperatures for the physiological functioning of most fish is closely related to the temperatures these creatures experience in their native habitats. Any temperature outside this ideal range has the potential to impair physiological performance. Fish may experience stunted development, stunted swimming ability, and reduced survival rates when exposed to cold water, particularly during the CWP. Juvenile silver perch (*Bidyanus bidyanus*) showed both short- and long-term negative effects on swimming ability and metabolic rate after exposure to a rapid (10-hour) drop in temperature. To simulate the conditions downstream after cold water discharge from a deep dam impoundment, this temperature change was engineered. It is not yet known whether reducing the pace of water temperature decline can lessen the effects of CWP-like temperature reductions, even though it is obvious that the magnitude of the temperature drop presents problems for *B. bidyanus*. This change in the rate of temperature decrease may be a management technique for areas affected by cold water pollution. There are more than just CWP that fish in freshwater environments face. Natural variations in water temperature, which may occur at different times of the year, are another challenge they face. Fish use a wide array of strategies to adapt to these changes. This could involve changing their physiological rate mechanisms to operate better at different temperatures through thermal acclimation or moving to more suitable environments. Through thermal phenotypic plasticity, also known as thermal compensation via acclimatisation, fish can adapt to habitats with fluctuating seasonal temperatures and shield themselves from the negative effects of climate change. Nevertheless, thermal adaptation may be affected by a number of variables such as body size, exposure time, thermal variability, and rate of temperature change. Many fish species are sensitive to temperature changes because they affect how quickly they mature and react to stress. In addition, different species, phases of life, and degrees of temperature fluctuations may have different effects on how fish react to prolonged exposure to low temperatures. Many fish species modify their physiological performance to cope with seasonal drops in water temperature. However, it is not yet understood how fish employ this technique when faced with instream obstacles that create fluctuations in the water temperature generated by humans, such as the CWP. The results of this assessment highlight key implications for the management of thermally stratified dam water-release methods.

It was important to observe the swimming capabilities of *Maccullochella peelii* (Murray cod) for a duration of eight weeks, during which their critical (U_{crit}) and sprint (U_{sprint}) times were estimated. Fish were periodically subjected to a range of temperature changes from 24 to 14 °C. Additionally, the metabolic rate and swimming performance were measured following an 8-week exposure period to assess heat adaptation.

The approach hypothesised that fish experiencing a gradual decrease in water temperature, which impairs their swimming abilities, would be less adversely impacted than fish exposed to a rapid 10 °C temperature drop. Similarly, medically speaking, fish would do better if the water temperature dropped slowly rather than rapidly or moderately after prolonged exposure to cold temperatures, and it was possible to find the native Australian fish species *M. peelii* in the MDB. Owing to human interference, *M. peelii* populations have dropped dramatically, and in some places they have vanished altogether. Consequently, *M. peelii* is classified as 'vulnerable' according to the Environmental Protection and Biodiversity Conservation Act 1999 of the Australian Federal Government. Specifically, CWP and other obstacles that restrict fish movement were mostly responsible for the decline of this species in the MDB. From the commercial hatchery Narrabri Fish Farm in New South Wales, Australia, 88 juvenile *M. peelii* with a total length ranging from 5 to 15 cm were procured from the commercial hatchery Narrabri Fish Farm in New South Wales, Australia. The ngsters were subsequently housed in two 1000-litre recirculating aquarium systems that were supplied with carbon-filtered tap water from Brisbane. Water was changed regularly using a drip feed system to preserve water quality during the experiment. Thermochron iButtons from iButtonLink (Wisconsin, USA) were used to control and track temperature. Commercial fish meal pellets and/or bloodworms were added to the fish's daily diet until they were satisfied. The photoperiod was 12 h of darkness and light. To identify each fish before testing, they were all given visible implant elastomer tags manufactured by Northwest Marine Technology (Washington, USA). After the tagging procedure, the fish were allowed to relax for one week. To ensure that all the treatments included fish of varying sizes, random selection was performed after tagging. In an effort to reduce aggressive behaviour, three 40-litre tanks were distributed among the three treatments. Participants in each treatment group were grouped according to their size. Five to ten fish per tank were the recommended densities for stocking.

In the 'warm' treatment (n = 22), water temperature was maintained at 24 °C throughout the trial. The trial's three 'cold' therapies each maintained a temperature of 14 °C. The temperature of the 'cold' treatments was decreased from 24 °C to 14 °C at varying rates, as illustrated in **Figures 2a - 2c**. During 'gradual' cold therapy (n = 22), the water temperature was reduced by 1 °C daily for 10 days until it reached 14 °C. The water temperature in the 'intermediate' frigid treatment group (n = 22) decreased by 5 °C on the first day and then by 1 °C per day for five days until it reached 14 °C. Under the 'rapid' cold treatment, the water temperature decreased by 10 °C in a single day (n = 21). The logic behind the quicker option was to simulate the sensation of a dam that rapidly released frigid water. The chronic exposure phase commenced after each 24-hour period of frigid exposure at 14 °C. The animals were subjected to this temperature for a period of eight weeks. The growth and swimming performance of each animal was assessed at their treatment temperatures at weeks 0, 2, 5, and 8 of the protracted exposure period. In week 9, the three frigid treatments were assessed after an hour at 24 °C, which was the same temperature as the warm treatment, which caused a 5 °C decrease in the water temperature in a single day, and it continued to decrease by 1 °C per day for five days, ultimately reaching 14 °C. The water temperature decreased by ten degrees Celsius within one day of treatment commencement. Temperature was maintained at 24 °C throughout the experimental period. This study compared the effects of various interventions on samples stored at 14 °C for 1 h. Slow, moderate, rapid, control, and scorching scores were included in the treatments. From week 10 to week 14, researchers measured the metabolic rate at both treatment and control temperatures. Survival was monitored daily during the trial. Before testing swimming performance or metabolic rate, all fish were fasted for 24 h. Following completion of the swimming performance and metabolic rate tests, the growth of each fish was assessed by measuring its total length (cm) and body mass (g). The 185-litre recirculating swimming flume (Loligo, Tjele, Denmark) with thermostatic control was used for swimming performance testing. The swimming room was 87.5 cm in length, 25 cm wide, and 25 cm tall, and submerged aquarium heating rods

from Aqua Zonic in Singapore or a Hailea HC-1000A chiller from Guangdong, China, were used to regulate the water temperature. Intermittent flow respirometry was used to quantify oxygen consumption (MO_2) and determine both the routine metabolic rate (RMR) and maximum metabolic rate (MMR).

For the experiment, tubing and a water pump (Eheim, Deizisau, Germany) were used with cylindrical acrylic respirometers that held 646 or 2040 mL of fluid. A closed-circuit recirculating loop was created using this setup, which enabled the respirometers to continuously mix. A T-piece with a fluorescent oxygen sensor was inserted into a recirculating loop connected to a Fibox 3 device (Regensburg, Germany). The oxygen content of water, expressed as a percentage of air saturation, can be continuously and noninvasively monitored using this setup. After each measurement interval, the respirometer was flushed with oxygenated water by using an additional circuit. The first circuit was measured, the second circuit was activated, and resonators equipped with a recirculating pump and flush pump were used to measure the resting metabolic rate (RMR) of the fish. They were given permission to remain in this area for half an hour so that they could recuperate from any stress caused by handling. After the flush pump was deactivated, the measurement cycle commenced. Each of the three measurement cycles could run anywhere from ten to 30 min, with the exact duration dependent on the fish size and air saturation percentage. Simultaneously, resting metabolic rate (RMR) values were noted, and the one with the lowest estimates was selected. The flush pump was turned on to oxygenate water during the measurement cycles. After a tiring pursuit tactic, the fish were swiftly placed into respirometers to determine MMR. Following the fish until they were exhausted or unable to swim in short spurts was the final step in the chase process. A temperature-controlled water bath (Clayson; Narangba, Australia) was used to achieve this goal. The length, breadth, and height of the water bath were 50×30.5×19 cm^3.

The highest calculated MMR was employed for the analysis, and aquatic oxygen concentrations were measured every minute for 15 min. The equation $\dot{M}O_2 = -1 \times \Delta O_2 \times V \times \beta O_2$ was used to calculate M O_2 (mg O_2 h^{-1}). In the respirometer, ΔO_2 represents the rate of change in oxygen saturation (as a percentage of air saturation per hour), V is the volume of the respirometer minus the mass of the fish (assuming a density of 1 g mL^{-1}), and βO_2 is the solubility of oxygen in water at a specific temperature. The Rstudio platform (version 0.98.1103), a statistical program, was used for the analysis. This study examined how several factors, including test and acclimatisation temperatures and acclimation length, affect different performance metrics using linear mixed-effects models. The lmerTest package is used in this study. The variables 'Treatment' and 'Week' were considered as fixed parameters to analyse the development of swimming performance measurements over time. Finally, swimming performance and metabolic rate at the treatment and opposite temperatures were evaluated with the variables 'Treatment' and 'Test temperature' held constant. Because the same fish and tanks were used for several measurements, the 'Fish ID' and 'Tank' random effects were employed to adjust for this. The variables 'total length' and 'mass' were used to evaluate swimming performance and metabolic rate, respectively. To ensure that the growth and metabolic rate data met the model assumptions, they were converted using a logarithmic function. When the 'step' function from the lmerTest package wasn't applicable, the 'ranova' function was utilised to examine the impact of random effects on the variance of the models. Tukey's HSD post hoc test was used to compare the least squares means and find differences between treatments or test temperatures. The Cox mixed-effects model, which is part of the survival and coxme packages, was used to model survival data. The random effect 'Tank' was embedded into this model. In order to compare survival curves after the fact, the 'survdiff' function from the survival package and the 'glht' function from the multcomp package were employed. Rapid drops in water temperature caused by cold-water discharge from thermally stratified dams can significantly affect the physiological functions of aquatic ectotherms downstream of the canal. One way to manage the rate of thermal change is to adjust the downstream water-delivery rate. As a result, creatures

have more time to adapt to the new temperature, which can reduce the overall impact of thermal change. This theory has been investigated in juvenile *Macaca peelii*. Their metabolic rate and swimming performance decreased when exposed to a temperature of 14 °C. Nevertheless, juvenile *M. peelii* individuals were able to adapt to the adverse effects of low temperature over time. However, it was surprising that the acclimatisation reactions of fish exposed to cold treatments were unaffected by the pace of the first exposure to cold water, contradicting the initial projections. Reducing the rate of temperature fall may not be a viable way to ameliorate the consequences of CWP on *M. peelii*, as studied in this paper. This discovery has implications for dam release management: how quickly temperatures vary and how much they fluctuate over the exposure period are two factors that can affect the ability to adapt to thermal conditions. Unmeasured physiological parameters may have responded differently to different rates of temperature change, although the investigation indicated that the acclimation ability was unaffected by the rate of change. Varieties of springtails, *Drosophila melanogaster*, *Linepithema humile*, *Coenagrion pulchellum*, and *Coenagrion armatum* have thermally sensitive features that are regulated by the rate of temperature exposure. When attempting to determine the critical thermal limits, for instance, the rate of temperature variation is a key variable to consider. Prior research has shown that these critical limits may be raised or lowered by varying the experimental rate of temperature change. In addition, as Ahmad et al. [6] highlighted, real temperature settings may cause rapid changes in thermal phenotypes, because they mimic the actual thermal conditions found in laboratory investigations. Although acclimatisation responses did not differ among the cold treatments, *M. peelii* demonstrated that thermal plasticity could differ across physiological parameters because of its variable acclimation responses.

In a recent study, a similar phenomenon was observed for *B. bidyanus*. Fish abruptly exposed to a 10 °C decrease in water temperature underwent a complete metabolic rate compensation. Conversely, only half of the U_{crit} adjustments were observed after ten weeks of exposure. Distinct physiological processes may cause acclimatisation to vary between traits. The heat sensitivity of various muscle fibre types and the metabolism that fuels such performance can be connected, for instance, to variations in performance across various swimming patterns in response to temperature. Additionally, metabolic indices, enzyme activity, sequence of fibre recruitment, and contractility of muscle fibres are all affected by temperature acclimatisation. In skeletal muscle, anaerobic enzymes such as lactate dehydrogenase become less active at low temperatures, whereas aerobic enzymes such as citrate synthase and cytochrome c oxidase become more active. Acclimatisation may be easier for swimming tactics that use aerobic pathways more frequently than for those that use anaerobic metabolism more frequently. Because of individual differences in thermoregulation, a plethora of physiological data are required to ascertain the effects on fish, and different people may react differently to acclimatisation depending on factors, including body size and the duration of exposure to them. Although only ng *M. peelii* was examined in this study, it is important to note that there is a large range of sizes within this species, sometimes greater than three orders of magnitude. Could the study predict the same acclimatisation responses in adults or larvae by conducting the same experiments? The temporal sensitivity of acclimatisation means that people might not have had sufficient time to adequately adapt to the tested conditions if the experimental acclimation durations were too short, leading to an underestimation of their adaptability. In addition, because of their rapid heating and cooling rates, smaller species are more thermally sensitive and adapt faster than larger species. Depending on the organism, this can significantly alter its acclimatisation reactions at various points during its lifespan. Using data from only one body size or life history stage to make management choices is not reflective of the full life cycle. Additionally, previous CWP experiments only used juvenile fish, highlighting the need for future research to incorporate both adult and larval fish to gain a more comprehensive understanding of how CWP affects these particular phases of development. Cold water flows may have fatal consequences for

downstream species, before considering the few negative effects reported in this study. *B. bidyanus* had a survival rate of less than 50 % in cold water channels compared to warm water channels, according to Achour et al. [7]. Cold shock stress, caused by sudden changes in temperature, may be responsible for a cascade of physiological responses that explain why survival rates vary. This emphasises the importance of the rate and amount of temperature reduction when delivering cold water. Dam water-release techniques have the potential to lower water temperatures by as much as 16 °C, which can have both short- and long-term negative effects on downstream fish populations. However, these findings provide evidence that thermal plasticity and compensatory response mechanisms may aid in reducing sublethal effects on performance in some fish species. The results of this study, together with those of earlier studies, provide strong evidence that fish can be physiologically and organismally affected by temperature conditions, similar to CWP.

Based on these considerations, the influence of impediments on river fragmentation can be better understood. These results indicate that phenotypic plasticity might not be sufficient to fully alleviate the consequences of CWP on *M. peelii*, even though it can help certain species cope with cold stress that results from CWP events. This is particularly the case when adaptation is time-consuming or when exposure to cold water might be lethal. Downstream freshwater communities are significantly affected by large impoundments in terms of management. Because it has a direct effect on performance and reproduction, maintaining water temperatures within ideal thermal ranges throughout the mating season is crucial for the survival and recovery of native fish species in thermally contaminated and fractured streams. Determining the temperature ranges of target species at different points in their life cycles is essential. This information will allow for the establishment of exact temperature goals for dam release. If releases are expected to significantly lower downstream temperatures, it is preferable to schedule them outside important periods, such as the mating season, if feasible. A different approach is to increase the temperature of the water released from thermally stratified dams to decrease thermal stratification. Several existing mitigation methods have been hindered by practical and logistical concerns, making the deployment of this strategy a continual issue. Physiological responses to laboratory-simulated temperature drops, similar to those observed in CWP, were measured to highlight the significance of using physiological data to influence future treatment choices. A growing number of people perceive physiological data as a practical way to address conservation issues by informing policies and management decisions. Conservation physiology has a distinct advantage in recognising problems and possible remedies related to fish migration. Laboratory-based assessments of swimming performance (U_{crit} and U_{sprint}) are crucial for understanding the capabilities of fish in their natural environments, especially when field measurements are limited by costly equipment, small sample sizes, and specialised training. Freshwater ecosystems downstream of large, thermally stratified dam impoundments may be better managed using physiological data, which is a challenging but promising area of CWP management.

Table S5 The results of Bayesian mixed models indicated a substantial positive or negative correlation between $\log_{10}$ FIB levels and geographical characteristics. This was determined using practical significance (PS) values greater than 0.50, probability of direction (PD) values greater than 0.75, and ROPE values less than 0.25.

Factor	Effect estimate	89 % Credibility Interval		PD	PS	ROPE
		Lower	Upper			
E. coli Latitude (°)	−0.24	−0.35	−0.12	1.02	0.94	<0.01
Urban site (Rural = Reference)	−0.45	−0.71	0.10	0.95	0.75	0.10
[d]Cold Brook Waterway (Onondaga Creek = reference)	−1.13	−1.83	−0.38	0.99	0.98	<0.01
Harbor Brook	−0.40	−0.65	0.06	0.92	0.93	0.23
Hopper Brook	−0.54	−1.12	0.32	0.97	0.90	0.14
Kimber Brook	−1.86	−3.39	−0.32	0.96	0.97	<0.01
West Branch	−0.23	−0.58	0.16	0.84	0.72	0.34
Enterococcus Elevation	−0.36	−0.54	−0.17	1.02	0.98	<0.01
Enterococcus Latitude (°)	0.17	0.07	0.28	0.99	0.82	0.17
Urban site (Rural = reference)	0.33	0.14	0.62	0.99	0.95	<0.01
Harbor Brook	−0.38	−0.52	−0.07	0.99	0.91	0.04
Ley Creek	−0.36	−0.68	−0.07	0.99	0.93	0.03
West Branch	−0.63	−1.07	−0.14	0.98	0.96	<0.01
Fecal coliforms Elevation	−0.29	−0.28	−0.19	1.00	1.00	<0.01
Fecal coliforms Latitude (°)	0.18	0.14	0.18	1.00	0.98	<0.01
Urban site (Rural = Reference)	0.54	0.64	0.69	1.02	1.08	<0.01
Cold Brook	−0.13	−0.49	0.15	0.78	0.54	0.43
Harbor Brook	−0.12	−0.19	−0.04	0.98	0.53	0.49
Hopper Brook	0.49	0.29	0.76	1.06	1.02	<0.01
Kimber Brook	−0.59	−1.30	0.38	0.96	0.91	0.10
Ley Creek	−0.28	−0.41	−0.20	1.05	0.88	<0.01
Sanders Creek	−0.15	−1.77	1.47	0.66	0.61	0.10
West Branch	−0.57	−0.80	−0.49	1.03	1.09	<0.01

[d]For multilevel categorical variables, if any level fulfils the PS, PD, and ROPE requirements (PS > 0.50, PD > 0.75, and ROPE < 0.25), the report includes the results for all levels. Variables not included in the table were determined to be statistically insignificant and were hence excluded.

S4. Policy Implications and future perspectives

The goal of this study was to determine how well image recognition algorithms work using quantitative phase imaging (QPI) pictures in two and three dimensions. Although the 3D method yields more precise results, the 2D method requires fewer computational resources. Research and development efforts pertaining to quantum photonic integration (QPI) are currently underway. Method 3 emphasises how this technology is attracting increasing attention. It is anticipated that other innovations will be prompted by the growing use of machine learning to manage massive amounts of data. An improved platform for real-time

monitoring of cyanobacteria populations will be available in the future in a design that is similar to the in situ (4Deep S7 Submersible Microscope) device but has greater spatial resolution and integrated flow cytometry capabilities. A methodology for real-time monitoring of cyanobacteria was proposed based on a literature analysis of existing technologies and methodologies. Quantitative phase-imaging microscopy and a neural network for machine learning were utilised in the approach made for cell identification. Early identification of harmful algal blooms (HABs), evaluation of water treatment efficacy, and provision of precise risk assessments based on microbial taxonomy and cell counts are all made possible by the ability to monitor and evaluate microbial activity in real time. The most feasible of the three potential QPI procedures (Methods 1-3) was selected to propose an enhanced cyanobacterial monitoring station. The theoretical groundwork for developing a practical water quality monitoring system was presented in this study. The goal of this strategy was to make HAB and cyanobacterial monitoring easier and more accurate. Real-time monitoring of cyanobacteria and algae is a powerful tool for the rapid detection of harmful algal blooms, evaluation of water treatment performance, and individual risk assessment of water quality according to the cell type and number. A collaborative approach integrating neural network image identification with a small imaging equipment was assessed and suggested in this study. Prior research on imaging microscopes and picture recognition was also reviewed. Because quantitative phase imaging provides more information than other options, it was deemed the most promising imaging technique among those that were studied. Image recognition neural networks, namely the convolutional neural networks outlined here, can benefit greatly from these data. In situ sample-collecting buoys and on-shore sample processing were part of the image processing approach and cyanobacteria monitoring system developed based on previous publications. This system can be constructed using readily accessible technology, which enables precise and continuous monitoring of water quality.

This presents a significant economic and environmental challenge owing to the continuously increasing levels of natural organic matter (NOM) in the surface waters of Northern Europe and boreal regions. Modern drinking water treatment facilities (DWTPs) frequently employ online spectrophotometres to measure absorbance at high frequencies to monitor changes in the chromophoric fraction of dissolved organic matter (CDOM) over time. This dataset contains valuable information that might be utilised to enhance the NOM removal process at different phases of treatment and/or pinpoint the root causes of the DWTP's underwhelming performance. Extensive pre-processing, variable selection, and signal processing are required to understand the massive datasets produced by automated monitoring systems. An analysis framework for time-series data from in-situ spectrophotometres, AbspectroscoPY, was presented in this study. By effectively handling duplicate entries, systematic temporal changes, baseline alterations, and outliers, the toolkit addresses significant data-preparation difficulties. Several spectral metrics, such as absorbance ratios, exponential fits, slope ratios, and spectral slope curves, can be generated for time-series data using automated algorithms. A drinking water treatment plant utilised AbspectroscoPY to assess three online spectrophotometres' 15-month datasets to prove its efficacy. The study described has been able to detect, quantify, and link variations in the spectrophotometric profiles of treated water to changes in lake turnover or DWTP operating conditions, even when the surface water quality has changed very little over time. With this toolkit, scientists have come a long way to create automated early warning systems that can detect and react to threats to treatment performance that might be triggered by sudden changes in water quality being fed into the system. Infected animals and humans can spread Listeria monocytogenes through their excrement even if they do not feel sick. Environmental contamination can cause diseases to be transmitted to animals and people through food and feed. *L. monocytogenes* may spread from agricultural environments to susceptible human populations through food manufacturing processes. Along with a brief overview of what is now known about asymptomatic carriers in animals and humans, this article

summarises the data on the effects of different droppings of different species on the food web, and discusses the methods by which asymptomatic individuals can contaminate animals and humans with listeriosis. Numerous food recalls and outbreaks of enteric diseases have been linked to the surface water pollution caused by human waste. Therefore, it is crucial to understand the causes of faecal pollution in agricultural water sources.

Therefore, this study aimed to examine the relationship between surface-water faecal indicator bacterial (FIB) levels (including E. coli, Enterococcus, and coliforms) and environmental factors. Additionally, this study aimed to examine the possibility of detecting host-specific faecal markers. The Syracuse New York Area, which includes both urban and rural areas, was the subject of this study, which employed data gathered from 224 sites along three streams. The concentrations of Enterococcus, coliforms, and E. coli were measured in 2,816 water samples collected between 2008 and 2017. A total of 31 samples were also examined for indicators of the existence of microorganisms of ruminant or human origin. Every site has its own weather dataset in addition to water quality metrics, including turbidity and nitrate levels. To examine the connection between each microbiological objective and factors including land use, water quality, and weather, univariable Bayesian regression was employed. The likelihood of the association's direction and strength was determined for each model based on the overlap between the direction and region of practical equivalence (ROPE). Despite the lack of a correlation between FIB levels and environmental factors, they have been shown to be associated with FIB levels. More precisely, there was a positive correlation between FIB levels and sediment, temperature, and nutrient levels. As the salinity and nitrate levels grew logarithmically from base 10 to 1, the E. coli and Enterococcus levels rose by 0.20 and 0.68, respectively, with a confidence range of 0.11 to 0.31 and 0.08 to 1.24. Microbial, sediment, and nutrient pollution in the studied watersheds were all caused by the same mechanisms, according to these results. The strength of the link between land use and faecal contamination varied with the type of faecal indicator bacteria (FIBs) and the distance from the land used to determine land use metrics, but overall, the association was substantial. The pasture cover was greater at 122,366 and 1,098 m from the sample site when E. coli and human markers were present. In contrast, pasture cover was only linked to the levels of Enterococcus and coliforms at a radius of 1,098 m rather than 122 or 366 m. There was a favourable correlation between pasture cover and ruminant presence markers at 122 m; however, no such correlation was observed at 366 or 1,098 m. The need to consider neighbouring land use, especially non-point sources of pollution, should be emphasised when devising strategies to reduce faecal hazards in agricultural and recreational water. The geographical scale, such as the difference between 122 and 1,098 m, must also be considered when constructing these strategies, as the cultivation of algae in either open raceway ponds or closed bioreactors enables the sustained generation of biomass that can be applied in several sectors, such as the food, pharmaceutical, cosmetic, and chemical industries. However, to ensure the efficient and cost-effective production of these chemicals, appropriate growth conditions are necessary. Therefore, accurate and frequent measurements are needed to control and monitor algal growth and identify process disruptions early. However, such analytical procedures are not readily available. Consequently, a framework was developed to track the real-time release of VOCs from algae in the airspace surrounding the bioreactor. Ion mobility spectrometry (IMS) and membrane inlet (MI) are the two components of this technique. Instead of aggregating signals, IMS can identify whole spectra in real time, which is its unique selling point. Principal component analysis (PCA) was used to automatically assess the spectral patterns created in the ion mobility spectrum. Each algal culture has its own distinct pattern of peaks, which helps to distinguish between development stages and shows how various experimental variables have an impact. These breakthroughs pave the way for the real-time tracking of algal cultures, which in turn enables the early identification of biotechnological process disruptions.

S5. Deterministic prototypes

The economic and recreational value of water bodies diminishes when human waste contaminates them, posing a threat to human health. There have been several outbreaks of enteric illnesses, the most probable of which is faecal pollution of agricultural or recreational water. A Salmonella outbreak struck the US in 2008, affecting many states. Irrigating spicy peppers with water contaminated with Salmonella was the source of the outbreak. Tomato growers in the United States lost $25 million due to the pandemic, which produced 1,200 instances of salmonellosis. The actual culprit, it turned out, was hot peppers, not tomatoes, which were formerly thought to be the source of the disease. Similarly, the annual cost to the US economy from gastrointestinal illnesses caused by recreational water sports is between $2.2 and $3.7 billion, or $1,220 per 1,000 participants. Reducing the quantity of human waste in water sources is crucial owing to the health risks and astronomical costs associated with excrement pollution. Reducing the amount of faeces that enters a system can be achieved, but only after pinpointing the specific points of entry and the wider regions from which the pollution is emanating. Moreover, it is of utmost importance to comprehend the temporal and geographical variations in these inflows, as well as surface water quality. Numerous studies have attempted to establish a connection between non-specific pollution sources and low levels of faecal indicator bacteria (FIB), which is a measure of surface water quality. The processes that facilitate the transport of contaminants from these sources to the bodies of surface water are also the focus of these studies. For instance, in a thorough investigation of water quality in the Iowa watershed, it was found that the levels of *E. coli*, a faecal indicator bacterium (FIB), were significantly correlated with the quantity of agricultural land at the location. On the other hand, research into streams in Pennsylvania indicated that FIB levels were significantly and strongly correlated with land use in close proximity to urban areas but only weakly correlated with land use in the watershed that was used for agriculture. Several studies have demonstrated that FIB may be stored in stream sediments inside the channel, and that storms and other events that disturb these sediments might return the bacteria to the water column; however, factors in waterways, including flow rate and particle size, can alter the resuspension process. Additionally, different streams and canals have different specific or diffuse sources of faecal pollution. Consequently, it is challenging to apply these and related study results to create workable strategies for reducing faecal contamination in particular water bodies.

Potential pollutant sources may vary greatly across different bodies of water or even within the same body of water, from one area to another. In addition, the biological diversity observed in freshwater settings influences faecal contamination through complex pathways. It is not easy to choose the most effective method for reducing faecal pollution in several streams that drain into Onondaga Lake from Syracuse, New York (NY). **Figure 4(c)** shows that Onondaga Creek, Ley Creek, and Harbor Brook were the three impaired waterways identified by the New York State Department of Environmental Conservation. Microbes, nutrients, ammonia, and turbidity are abundant in certain streams. Efforts to clean these waterways have been ongoing for more than 30 years. Combined sewage overflow (CSO) volumes and faecal coliform loads in Onondaga Lake originated in Onondaga Creek in the mid-1990s. It was also determined to be the principal entrance point for the sediment into the lake. Rivers are still in a bad state despite all initiatives to fix the problem (such as lowering Combined Sewer Overflows) and constant and thorough monitoring. Consequently, some rivers have strict regulations that make it illegal or very difficult to enjoy water sports, such as swimming and fishing. Between 2000 and 2007, monitoring efforts revealed that 16 % (34 out of 215) of dry weather days at a rural sample location and 75 % (162 out of 215) at an urban sample location had faecal coliform levels exceeding New York's water quality limits. The results show that combined CSO discharge is not the only source of pathogens found in the Onondaga Creek and other Syracuse waterways. Data from other regional monitoring programs corroborated this finding, showing that fluctuations in faecal

coliform bacterial levels cannot be explained by discharges caused by precipitation alone. In addition, there was a clear trend in the data, indicating that the bacterial counts were significantly higher within the city limits of Syracuse than outside. It was determined that an outdated sewer system with damaged pipes, inadequate crosslinks, or prohibited connections was responsible for dry weather discharge, which was undeniably a significant finding from the enquiry. Multiple studies have linked septic system density, sewer system age, and proximity to discharge sites/septic systems to surface and groundwater contamination; therefore, this reasoning is in line with prior findings. There were three primary goals of this study: (i) to find out where faecal contamination occurs in Onondaga Creek, Ley Creek, and Harbor Brook (waterways near Syracuse, NY), (ii) to see how microbial indicators of faecal contamination relate to things like weather and nutrient levels, and (iii) to suggest ways to prioritise mitigation efforts.

References

[1] R Abiri, N Rizan, SK Balasundram, AB Shahbazi and H Abdul-Hamid. Application of digital technologies for ensuring agricultural productivity. *Heliyon*, 2023; **9**, e22601.

[2] M Bucki, KN Dhufaigh, C O'Brien, A Weatherley, N Walshe and T McElligott. Comparison of ovine faecal Strongyle egg counts from an accredited laboratory and a rapid, on-site parasite diagnostic system utilising a smartphone app and machine learning. *Veterinary Parasitology*, 2023; **320**, 109976.

[3] LF Brito, HR Oliveira, BR McConn, AP Schinckel, A Arrazola, JN Marchant-Forde and JS Johnson. Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. *Frontiers in Genetics*, 2020; **11**, 793.

[4] IL Bretas, JCB Dubeux, PJR Cruz, KT Oduor, LD Queiroz, DSM Valente and FHM Chizzotti. Precision livestock farming applied to grazingland monitoring and management: A review. *Agronomy Journal*, 2023; **116**, 1164-1186.

[5] H Álvarez-García, JA Sillero-Medina, ME Pérez-González and JD Ruiz-Sinoga. Spatial variations of runoff thresholds associated with changes in land use over the period 1990-2018 in the Mediterranean side of Andalusia. *Investigaciones Geográficas*, 2024; **82**, 25-42.

[6] F Ahmad, A Farooq and MUG Khan. Deep Learning Model for Pathogen Classification Using Feature Fusion and Data Augmentation. *Current Bioinformatics*, 2021; **16**, 466-483.

[7] B Achour, M Belkadi, I Filali, M Laghrouche and M Lahdir. Image analysis for individual identification and feeding behaviour monitoring of dairy cows based on Convolutional Neural Networks (CNN). *Biosystems Engineering*, 2020; **198**, 31-49.