1 August 2024

Dear Dr. Zucchetta,

We present this revised manuscript entitled "T**urbidivision: a machine vision application for estimating turbidity from underwater images**" to *PeerJ*. We thank you and the reviewers for your feedback that has strengthen the manuscript from our original submission, and believe it is now suitable for publication.

We have added detail to give readers a sense of turbidity professional standards for accuracy, availability of historical data for image analysis, and added detail to points throughout the methods. Our responses to all reviewer comments are included below.

The authors declare no conflict of interest and that no part of this manuscript has been previously published in any form. Both authors have approved the manuscript and agree with its submission to *PeerJ*.

Best Regards,

Matthew J. Wilson

Editor comments (Matteo Zucchetta)
As you can see, three reviewers have commented on your manuscript. It was well received, and the comments—both specific and general—mainly suggest changes to clarify the content or highlight the potential impact of the work.

Reviewer 1 (Anonymous)
Review Report for PeerJ
(Turbidivision: a machine vision application for estimating turbidity from underwater images)

1. Within the scope of the study, various classification operations and regression operations were performed with deep learning using underwater images.
2. In the introduction section, the importance of the subject was mentioned very limitedly. In addition, the purpose of the study and especially the originality point should be stated more clearly.

We have reviewed the introduction to identify specific points that could be bolstered in response to the reviewer's suggestion. We have outlined the significance of turbidity broadly across human and ecological contexts as well as the emergence of machine vision and how it has been applied in other water quality contexts, but not for turbidity of natural waters. Our goal statement follows these to identify the knowledge and utility gaps this study fills (i.e., originality). In response to this suggestion and others we have added a clause to the goal statement about the novelty of the model's accessibility, the availability of historical datasets this model could be applied to, and the potential use of machine vision in this context (i.e. the gap our research addresses). If there are additional or specific suggestions we have not been able to identify, we would be happy to incorporate them.

3. The type and amount of dataset used seems sufficient within the scope of the study. However, it would be more appropriate to prefer cross-validation for a more accurate analysis of the classification operations.

As we understand this comment, we did use cross-validation in our classification. We used independent data for training and validation and 50 models (i.e. 50 folds of validation vs training sets). If the reviewer is suggesting we reuse training data for additional validation within a single model then we would be overfitting the model.

4. In the model training section, it was stated that YOLOv8 was used. Although there are many different deep learning models that can be used in this context in the literature, why was this model preferred in particular? When YOLO versions are examined, it is seen that more up-to-date ones are available in the literature, why was this version preferred in particular? What are the reasons why other state-of-the-art deep learning models are not preferred in the literature within the scope of the study?

YOLOv8 was the latest stable YOLO version at the time we began model training. We have added two sentences to the methods in response to this comment and suggestions from Reviewer 3 to explain which YOLO model was used with our rationale.

5. The use of a web application within the scope of the study is a positive situation in terms of the applicability of the study. At this point, the study proves itself in terms of usability, but there are big question marks in terms of originality.

We are unsure of a suggestion change.

6. The results obtained and confusion matrix outputs seem sufficient for the first stage of the study. However, attention should be paid to model diversity.

We are unclear of the reviewer's suggestion for "model diversity". If the question is about source image diversity then we believe the absence of influence of metadata on model training addresses this concern. Or if it is the number of models or diversity of images in validation vs training sets, then our 50 models and selection of the best fit should address the concern.

In conclusion, the study is important in terms of its subject and applicability, but attention should be paid to the sections listed above in order to clarify issues such as originality and contribution to the literature.

We have made changes where appropriate to address the reviewer's concerns, as outlined above.

Reviewer 2 (Pierre Lechevallier)

Basic reporting

In this article, the authors present a model they developed to estimate turbidity from underwater images. Their findings are robust, with an extensive set of data, good model performance, and well-explained methods. This paper's content is relevant to the field of water monitoring. However, there is substantial room for improvement in the paper's form. Many sentences are ambiguous, redundant, and do not follow scientific writing standards. Therefore, I do not recommend publishing the article as it is, but I encourage the authors to carefully revisit the form, as I believe that their findings are very promising. In this review, I used the structure provided by the PeerJ editors, and used a red/orange/green colors to evaluate each point. Each evaluation is followed by a detailed explanation.

RED: The article must be written in English and must use clear, unambiguous, technically correct text. The article must conform to professional standards of courtesy and expression.

Explanation: see comments in the document. I primarily concentrated on the abstract and the introduction to provide detailed comments, but the criticism is applicable to the whole document.

Changes made where appropriate and we have responded to all in-line comments.

ORANGE: The article should include sufficient introduction and background to demonstrate how the work fits into the broader field of knowledge. Relevant prior literature should be appropriately referenced.
Explanation: The authors defined in the first paragraph of the introduction the relevance of turbidity in water quality assessment. However, despite using in the article statements such as line 47 "High turbidity can have a negative impact on aquatic life" or line 53 "overly turbid water is unsuitable for consumption by humans", they did not provide any information on what is considered a high/normal/low turbidity value in the context of natural waters and drinking waters. This is a problem because the reader is not able to assess whether the measurement range (0-55NTU) is relevant in this context.

We have added a statement to identify water quality standards from the World Health Organization. However, providing broader context for natural waters is more difficult as relatives are highly dependent on context and do not relate to specific values. We have replaced "high" with "increasing" in the manuscript where referencing turbidity effects on aquatic life because rate and intensity of change relative to baseline is more important than absolute values in a general sense. For example, baseline and relative increases in the Amazon or Mississippi would be drastically different from reefs in the Azores.

GREEN: The structure of the article should conform to an acceptable format of 'standard sections' (see our Instructions for Authors for our suggested format). Significant departures in structure should be made only if they significantly improve clarity or conform to a discipline-specific custom.
Comment: the article is brief, well-structured and information are presented where the reader expect them.

No changes made.

ORANGE: The submission should be 'self-contained,' should represent an appropriate 'unit of publication', and should include all results relevant to the hypothesis.
Explanation: The authors used the argument that historical dataset of underwater images could be analyzed with their method (line 75: "historical data from underwater images often lacks […]" or line 96 "gaining insights into historical data"). I agree that this is a very strong argument for the relevance of their model. Yet they did not provide any evidence that such data are available. I suggest citing existing dataset, and even (if possible) analyzing some of historical data with their algorithm, which would with little effort greatly improve the strength of the article.

We have added a citation/example of an existing dataset of underwater images that does not include associated in-situ turbidity data in the third paragraph of the introduction. We would also argue historical datasets need not be publicly available for the model to be valuable. Both authors are aware of multiple image datasets their owners have not released into the public domain and there are likely many other datasets in this state.

Experimental design
The submission should clearly define the research question, which must be relevant and meaningful. GREEN: The knowledge gap being investigated should be identified, and statements should be made as to how the study contributes to filling that gap.
Comment: the introduction was well structured and highlights the importance of turbidity for water monitoring, as well as the limitations of currently available techniques for turbidity measurement, leading to the research gaps.

No changes made.

ORANGE: The investigation must have been conducted rigorously and to a high technical standard. The research must have been conducted in conformity with the prevailing ethical standards in the field.
Explanation:
There is, at difference instance, a lack of justification to explain why the authors choose certain methods over others:
• Line 122: "For 38% of all photos, we included the 4.2cm Secchi". Why did you do that? What is the impact of the Secchi disk on the model performance? Did you try to train the model with images without Secchi disk? (This is relevant because for future use of the model most users will not have access to a Secchi disk).

We have added a statement that this allowed us to use all images together and moved the statement about Secchi proportions later in the methods for better logical consistency. As stated in line 194-195 of the results, Secchi disk presence or absence had no effect on model performance. Including all images did not negatively impact performance on images without a Secchi disk.

• Line 148: "We used Ultralytics YOLOv8 classification". Why did you selected this method? Did you try other approaches? Why do you believe that this method is suited for this task?

Classification models are the standard for machine vision models in this context and YOLOv8 was the most recent release at the time of our analysis. Our "Future Directions" section of the discussion is entirely devoted to a discussion of alternatives to our approach. We have added a statement to clarify this was the most recent release of YOLO.

• Line 136: "the images were broken into 11 groups". Why eleven? Why not 5? Did you optimite this number, or is it for practical reasons?

We experimented with various group sizes and found that 11 gave the best results without spreading the data so thin that there was not enough data for each class (i.e., more groups create better "fidelity", but spreads the data more). We have added a sentence to clarify this.

ORANGE: Methods should be described with sufficient information to be reproducible by another investigator.
Explanation: In general, the description about the data collection lacks information, which hinders reproducibility:
• Line 106: "whether the water was flowing". Under which criteria was the water flowing? Did you use a flowmeter? Did you use this information in the analysis?

We did not use a flow meter, however our categories of lotic and lentic images are ipso facto expressions of flow as a binary. As stated in line 194 of the results, there was no effect of metadata on model accuracy. At line 105 we have changed "data" to "metadata" to avoid any confusion.

• What about the illumination? Did you use sunlight? Did you use the camera flash? Did you collect information about it?

Whatever the natural lighting in the situation was (sunlight, room lights) with no flash. Information was not collected as the model should be able to "learn" the differences in lighting on its own and adjust accordingly (and successfully did). We have added a sentence to the first paragraph of the methods clarifying ambient light conditions.

• Line 115: "Images were collected by taking individual photographs [...] and by saving a selection of frames from a video". Why to acquisitions methods ? How many images are taken from each mode?

We did not include this in the metadata as the model reduces image quality (to increase processing speed) beyond a point where it would be relevant. If this statement is confusing we are happy to remove it, but thought it relevant for full transparency.

Validity of the findings
ORANGE: The data on which the conclusions are based must be provided or made available in an acceptable discipline-specific repository. The data should be robust, statistically sound, and controlled.
Explanation: I am concerned about the validity of the findings because some information about the dataset are missing. The authors used images takes from different sources, which is improving the range of application of their model. My concern is: is it possible that the model is not actually predicting turbidity, but the origin of the image?
Here is a possible scenario: all samples from rivers have a turbidity between 0 and 5 NTU, all samples from the lab experiment have a turbidity between 10 and 15 NTU, etc. In that scenario, the classification model could be predicting the origin of the image, which is itself correlated

with turbidity. The problem would be that if one provide a new image to the model, it will not be able to predict turbidity accurately.

As stated in our results, none of the associated metadata groups influenced model accuracy. We are unclear where in the manuscript the reviewer has identified this concern. If this were the case as suggested by the reviewer, then habitat/location/camera/etc. would have shown variable accuracy across bins.

Answer: thanks to the clarification made in the paper, my concerns are addressed.

Solution: 1) provide information, for each sub-parts of the dataset, about the range of turbidity. What about, in Figure X_REF, using a color-code for each dot representing the origin of the sample? 2) provide information about the train-validate-test splitting of the data, regarding whether images from each source is present in each one of them. 3) Even better: keep a whole section of the dataset, for instance river images, from the training and validating set. If the model still performs well on this totally new images, it means it learned how to estimate turbidity from any images. This is important because future users of your app need to verify that the model will still perform if they take a picture in a totally different context.

As we understand the suggestion, this goes against best practice and would overfit the model to specific conditions. The train-validate-test split of data is randomized to specifically prevent bias in outputs. In addition, if the model we have presented was not effective with images across contexts then the metadata would have identified differences in the accuracy with our test images.

Answer: since you clarified this point about metadata, you can disregard all my comments on this topic.

ORANGE: The conclusions should be appropriately stated, should be connected to the original question investigated, and should be limited to those supported by the results. In particular, claims of a causative relationship should be supported by a well-controlled experimental intervention. Correlation is not causation.

Comment: see comment above.

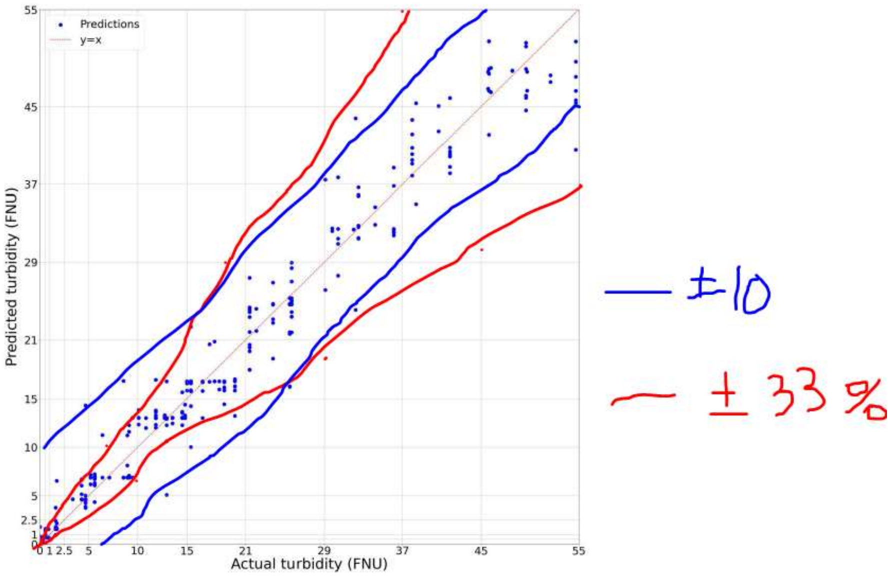
Additional comments

I have two additional remarks/questions about the result analysis.

First, why did you present the model accuracy in percentage (above 2.5 FNU)? In my opinion, the main message of this article is that with a properly trained model, any camera can give a +/- 5 FNU estimation of the turbidity, no matter the range. By giving accuracy in %, you are confusing the reader who could believe that the accuracy increases with increasing turbidity (heteroscedasticity), which is apparently not the case (I am using Figure 2, but a residual plot would enable the reader to judge heteroscedasticity more clearly).

The reviewer's comment has this relationship in reverse. As turbidity increases, the accuracy decreases as an absolute value but is consistent as a proportion. (i.e. higher values are more variable). I can understand where the reviewer is coming from with the concern for heteroscedasticity, however it is misplaced in this context. Because of the nature of our multiple linear regression not all regression equations are parallel, and as the min/max diverge from one another at increasing values then the range is proportional, without affecting the residuals.

Answer:



To me it looks like 33% accuracy is actually understating the model accuracy. Or I would express it in this way: 33% accuracy below 20NTU and +/-10 above 20 NTU. I do not have the raw data to make precise calculation, but this is what I can imply from the figure.

Second, why did you analyze the results differently below and above 2.5FNU? You can consider whether this is necessary, or if it is not overcomplicating things. In my opinion, this is not very relevant, as a natural water below 2.5 FNU can already be considered as very clean, not matter if the turbidity is 1.9 or 1.8 FNU. Again, in my understanding, your approach is not aiming at being as accurate as possible, but instead to be accurate enough to provide a cheap, accessible and quick way to estimate turbidity.

This is how turbidimeter specifications are typically given (including the meter we used). We did it this way to give a clear comparison with the turbidimeter and allow the reader to make a direct evaluation of which (model or direct measure) is most appropriate for their needs (lines 191-193 and 245-249).

Answer: I understand this approach. In my opinion, this overcomplicates the message of the paper regarding model accuracy, but I leave that to you to decide.

The reviewer has also provided an annotated manuscript as part of their review:
Line 18: quality and purity is redundant.

We disagree. Depending on context these terms are complimentary. Water quality in natural waters is often not about purity at all, but which materials are present in the water. For water treatment plants and drinking water standards, purity is a more critical metric.

Line 20: "additional" in reference to what?

Relative to photography. Changes made to clarify.

Line 26: only for model training? Because you used "robust" in the beginning of the sentence, I expect not only that you collected data, but that you collected an qualitative and quantitative dataset. I suggest that you add such details in the abstract

We are unclear of the reviewer's suggested change here. We have not collected any qualitative data nor do we understand how a machine vision model could be applied to qualitative data in this context.

Answer: I am advocating here that you add in this sentence information about the quality and quantity of data you collected (ex: number of photographs, turbidity range, number of sites, etc.), so that the reader can judge from the abstract the "robustness" of the dataset.

Line 28-30: unclear. Is this sentence useful? After reading the full paper, I now understand this sentence. However, the fact that at first reading, this was unclear to me, reveals that other readers might also be confused by this sentence. I suggest to skip the storyline with classification in the abstract, the readers are anyway mostly interested in the regression, i.e. turbidity prediction accuracy

We disagree with the reviewer about the importance of this statement. The regression is a product of the classification and it would therefore be misleading to lead with regression, potentially implying an inflated confidence in the precision of the model. We have added a clause in the preceding sentence about our data bins and range to clarify the sentence in question.

Answer: In my opinion, making too much emphasis on the classification is just overcomplicating the main message of your work (which is: you developed a model that can, from water images, estimate turbidity with XXX accuracy). I see the classification as a preprocessing step to the regression model, and I would just not mention it in the abstract. Or at least not mention the results of the classification.

In the current version of the abstract, it is unclear that the regression and the classification are coupled, and the reader might think that they are independent. I suggest something along this line:

“We used a two step approach based on i) a classification and ii) a regression using the classification outcomes to estimate turbidity from the images. Our approach made it possible to estimate turbidity with XXX NTU precision […]. “

Line 31: This is confusing that you use two different units to quantify accuract: FNU for measurements below 2.5FNU and % above 2.5 NTU

Please see our response to the more detailed version of this comment above, under “addition comments”

Line 39: Long paragraph, consider splitting

We appreciate the observation but have been unable to identify an appropriate approach to splitting the paragraph without creating a disjointed outcome.

Answer: what about splitting before:

“High turbidity can have negative impacts on aquatic life, since it often shares an inverse correlation […]”

Line 40: turbidity is not "impacting" water clarity, it is a measurement of water clarifty

We disagree with the suggested change and have articulated the point in response to the suggestion for Line 42 below.

Answer:

Turbidity is the measurement of an optical property of water, not of suspended particles.

I am here using this reference:

Rügner, H., Schwientek, M., Beckingham, B. et al. Turbidity as a proxy for total suspended solids (TSS) and particle facilitated pollutant transport in catchments. Environ Earth Sci 69, 373–380 (2013). https://doi.org/10.1007/s12665-013-2307-

“Suspended particles in water cause scattering of light leading to turbidity (cloudiness) of the water, which can be measured by optical backscatter sensors (OBSs, turbidity meters) in the field or by so-called nephelometers commonly used in laboratory water analysis.“

Line 42: This is not very precise. Turbidity, i.e., measurement of transparency/cloudiness, is ***often*** *correlated to the concentration of suspended particles, and can therefore be used as proxy.*

We disagree with the suggested change. Turbidity comes from the scattering of light, which requires the presence of particles and can lead to visible cloudiness as a result. Water can

take on an apparent opacity without increasing suspended material (i.e., colorometry vs. turbidity).
Answer:

I am citing here the abstract from this paper: **https://doi.org/10.1002/hyp.3360090108**
"Investigating the transport of suspended solids by water sampling usually leads to an underestimation of loads and an unrealistically high sampling frequency is required to properly characterize temporal trends. An alternative method is to use in situ optical turbidimeters to estimate the suspended solids concentration; however, the relationship between turbidity and suspended solids concentration is potentially confounded by variations in particle size, particle composition and water colour."

My answer was not well formulated. The linear relationship between turbidity and suspended solids in a single site is verified most of the time, but the parameters of the linear regression change from site to site. Reference:
https://link.springer.com/article/10.1007/s12665-013-2307-1/tables/1

Line 44: Comma between "soils" and "or".
No changes made – the "and" and "or" here are separate Booleans and not a three-part series.

Line 45: what is the subject of "depend": these praticles? Consider splitting
Yes, we have changed plant debris and microscopic organisms to a parenthetical example to clarify the sentence.

Line 46: a runoff is not a pollutant, it is transporting the pollutants. "Pollutants contained in industrial ..."
Changes made.

Line 47: what is high thurbitity?
Good point. We have changed "high" to "increasing" as "high" is context-dependent depending on the taxa involved and baseline turbidity levels in the system.

Line 49-52: the structure of the sentence is not correct
No changes made. This sentence is grammatically correct and we are unclear of the suggestion.

Clarifying my comment:
In addition, increasing turbidity can cause nonlethal effects on aquatic taxa through altered predator-prey interactions (Abrahams and Kattenfeld, 1997; Ferrari et al. 2010),

*[I do not understand the subject of this part of the sentence. Is it: "*In addition, increasing turbidity can cause nonlethal effects on aquatic taxa through*"?]* decreasing [decreased?] light penetration and reduced photosynthesis (Moore et al., 1997), and [again, what is the subject here?] physical stress such as gill deformities (Lowe et al. 2015).
I suggest splitting this sentence into three sentences.

Line 49: species?

No changes made. As many aquatic groups' response to water quality are characterized at multiple and inconsistent taxonomic levels (e.g., operational taxonomic units of macroinvertebrates), "taxa" is the more appropriate term in this context.

Line 50: what is the issue if the effects are nonlethal?

No changes made. We are unclear of a suggested change as the following series of citations and examples expand on the issues from nonlethal effects.

Answer: OK

Line 52-55: Again, incorrect sentence structure

No changes made. This sentence is grammatically correct and we are unclear of the suggestion.

Clarifying my comment:
Original sentence:
Beyond affecting aquatic ecosystems and taxa, overly turbid water is unsuitable for consumption by humans (Muoio et al., 2020) and livestock (Umar et al., 2014), as well as acting as an indicator of bacterial contamination (Gharibi et al., 2012).

The structure of the sentence is correct, but its logic is strange. Because the sentence starts with "Beyond affecting aquatic ecosystems and taxa", the reader expects to read about further negative effects of turbidity. However, "acting as an indicator of bacterial contamination" is not a negative effect, it is just a fact.

Line 56: filtration (be concise) and "filtration that remove suspended particles"

Changed to "filtering" to make the statement concise.

Line 57: acceptable levels…which are?

Dependent on filtration method and use and always below 5 NTU. We have added a statement to specify this.

Line 60-61: Incorrect structure

No changes made. This sentence is grammatically correct and we are unclear of the suggestion.

Answer:

On second reading I realized I was wrong and that the sentence is correct. Maybe the sentence can be improved in this way:
Typically, turbidity is measured with a turbidimeter which shines a light source - either white or infrared- into the fluid and a probe which measures the resulting scatter of light.

Line 62-63: unclear sentence, please reformulate

No changes made. We are unsure why the review found this statement unclear or a suggested change.

Answer: You are right the sentence is clear

Line 67: be precise & Line 67-68: Citing a specific brand is probably not ethically acceptable for research. Please consider anonymizing.

No changes made. The reviewer has both suggested we be more precise about costs while removing a reference to a specific cost/model of turbidity meter. It is not possible to make changes that agree with both suggestions.

Answer: OK

Line 67: remove "quite"

Changes made.

Line 69: change "labs" to "laboratories"

Changes made.

Line 71-72: unclear, please reformulate

No changes made. We are unsure why the review found this statement unclear and there is no suggested change.

This expense can be a significant barrier to citizen science initiatives, or labs with a limited budget, decreasing the ability of laypeople to contribute to water quality monitoring efforts. The converse can also be true, that lowering costs and creating easy and consistent methods of data collection increase the quality of data and participation in data collection.

This is how I understood this statement:

High expenses can hinder citizen science initiatives and labs with limited budgets, making it harder for non-experts to contribute to water quality monitoring. Conversely, reducing costs and simplifying data collection methods can enhance both the quality of data and the level of participation.
"both the quality of data" -> I do not understand how using camera instead of turbidimeter improves the quality of data.

Line 74: This argument is not valid.

We disagree. Every additional piece of equipment for field work adds a cost in time and effort for data collection. It is unclear why the reviewer disagrees with this statement. No changes made.

Answer:

Here is the original sentence:

"Measuring turbidity with a turbidimeter can also be time-consuming, as they necessitate the transport of the meter itself, […]"

Whether you measure turbidity with a camera or a sensor, you need to go on-site and do the measurement. I do not see why using a turbidimeter takes more time that a camera. Turbidimeter can measure turbidity online, i.e. in seconds:

https://www.endress.com/en/field-instruments-overview/liquid-analysis-product-overview/turbidity-drinking-water-sensor-cus52d?t.tabId=product-overview

Line 75-76: For me, this does not belong to this paragraph, which is abou the limitation of turbidity measurement method. However this is a very good point, I would include it in the next paragraph.

Good point, we have moved this statement to the following paragraph.

Line 75: please provide reference

We have been unable to find a single reference to tie historical image datasets with in situ turbidity data and updated the statement to reflect this.

Line 77-82: this paragraph is very fuzzy. In my opinion, a logical connector is missing and this all comes out of context. You ended the last paragraph with limitations of turbidimeters, why do you transition to image vision? Also "emerging field of research" ? Satellite are used since 30 years for ocean monitoring:
Doerffer R, Fischer J. Concentrations of chlorophyll, suspended matter, and gelbstoff in case II waters derived from satellite coastal zone color scanner data with inverse modeling methods. Journal of Geophysical Research: Oceans. 1994;99(C4):7457–66.
Lavery P, Pattiaratchi C, Wyllie A, Hick P. Water quality monitoring in estuarine waters using the landsat thematic mapper. Remote Sensing of Environment. 1993 Dec 1;46(3):268–80.

We have rephrased the first sentence of the paragraph to connect this paragraph to the previous paragraph with machine vision as an opportunity to address these limitations. The use of machine vision to assess water quality is an emerging field, and we have rephrased the first sentence to clarify we are not discussing general image analysis.

Line 81: this? which work are you talking about

"this work" changed to "machine vision research for water quality"

Line 86: Our model offers

Changes made.

Line 93: also?

We are unclear of a suggested change.

Answer: I would remove the "also".

Line 101-102: Strangely formulated sentence

No changes made. The sentence is grammatically correct and we are unclear of a suggested change.

Answer:
"To develop this model, we paired two components in data collection: an underwater photograph, and a measurement of the turbidity of the water source in the image"

Here is a proposition for simplification:
To develop this model, we collected XXX underwater photographs and the reference turbidity values of the water in each photograph.

Line 102: camera

Changing "photograph" to "camera" would make the statement incongruent with one part of the statement a type of data and the other a sampling method. No changes made.
Answer: OK

Line 107: remove "and"

Changes made.

Line 111-112: Maybe, to improve reproduceability, provide more information about the selection of site.

Statement added to explain the intention behind our site selection. As site-specific, and even ecosystem-specific, features were not important for model training they are not relevant for reproducing a similar model.

Line 118-122: long sentence, consider splitting

Changes made.

Line 122: why ? why 38%?

Secchi and natural images were collected at equal proportions until testing confirmed there was no difference for model training, then all images were used together.

Line 148: why this method in particular?

Image classifications models are the standard and we provided a discussion of alternatives in the “future directions” section. (see full explanation above)

Fig 2: strange grid

Gid lines match bins used in our analysis. We have added a sentence to the figure caption to clarify.

Fig 2: change “actual turbidity” to “measured turbidity”

Changes made.

Reviewer 3 (Anonymous)

Basic reporting

This article presents the results of a machine vision model to estimate the turbidity from underwater images. The topic is interesting and has real world applications, specially since the authors have created a stand alone and a mobile application. The paper is well structured and clear.

Experimental design

The objective is clearly stated, which is to train a machine vision model to estimate turbidity values from underwater images.

The underwater images were taken with 2 types of cameras, which were taken in "auto" mode, hence each photo can have different exposure times and ISO. In some photos, a Secchi disk was in the field of view. Photos were collected in different water bodies: rivers, lakes, ponds and the ocean, additional photos were taken in controlled environments. A total of 675 images were taken. Reference measurements were taken with a LaMotte 2020i portable turbidity meter. To generate the model the Ultralytics YOLOv8 classification model was used.

Comments:

1. The most important part in this work is the model used, the YOLOv8, which is not explained. A brief explanation should be added.

1a. As I understand, YOLOv8 is mainly an object detection and image segmentation model, how was it applied to this research problem?

To address comments 1/1a we have added a sentence on the different types of YOLO models (object detection, image segmentation, and classification). We have also specified more clearly which of these was used and our justification.

1b. It is mentioned that transfer learning was used starting with a pretrained model, which data was used for that pretrained model?

YOLOv8 is pretrained on the COCO dataset. We have added this in-text and the citation to our references.

1c. It is mentioned that no changes in model accuracy where observed when the type of water body was included on the training or if a Secchi disk was in the view. Can the authors give more insights on this? All the images look very different and I wonder how the model is learning.
1d. It seems that the shutter speed and ISO do not to have any influence on the results, can the authors explain why?

We do not have a satisfactory way to effectively answer 1c/1d. By their very nature, neural networks have a "black box" aspect to model development. The best we can do is hypothesize that the aspects that allow the model to assess turbidity are not related to image background. If the reviewer has suggestions for how to address this question in-text we would be happy to incorporate them.

2. Did the authors consider taking photos from above the water surface? Which challenges can this type of images pose?

To keep the model as focused as possible while still generalizing across habitat types, we decided to stick solely with underwater images. Above and below water would create very different challenges for a model and might actually work best as two separate models (i.e., select "above" or "below" water surface). We believe this would be a great way to improve accessibility (allowing any camera to collect photos) but is outside the scope of our study.

Validity of the findings
The 675 images are provided as well as a Github repo in Zenodo (which I did not try). Additionally, stand alone and a mobile application were created, the stand alone application can be downloaded from zenodo and the mobile application from Play Store. The discussion and conclusions the results are commented and the goal of the paper is answered.

Additional comments
It is mentioned in the introduction that the approach presented herein to measure turbidity "provide an alternative to measuring turbidity when lower accuracy is acceptable, such as in citizen science and education applications". If the authors can give more information about which accuracy is needed for other users (eg. environmental agencies, municipalities) would be interesting and what would be needed to improve it.

This is a great point. In searching for accuracy standards across applications, a common theme is that most requirements focus on calibration rather than accuracy, or the method an instrument uses to calculate turbidity. We have added a few sentences to the second paragraph of the introduction about USGS testing of devices that meet EPA and ISO standards to provide a concrete benchmark for this statement.

Turbidivision: a machine vision application for estimating turbidity from underwater images

Ian M. Rudy[1], Matthew J. Wilson[2]

[1] Math and Computer Sciences, Susquehanna University, Selinsgrove, PA, United States
[2] Freshwater Research Institute, Susquehanna University, Selinsgrove, PA, United States

Corresponding Author:
Matthew J. Wilson
514 University Ave, Selinsgrove, PA, 17870, United States
Email address: wilsonmatt@susqu.edu

Abstract

The measurement of turbidity serves as a key indicator of water quality and purity, crucial for informing decisions related to industrial, ecological, and public health applications. As existing processes require both additional expenses and steps to be taken during data collection relative to photography, we seek to generate accurate estimations of turbidity from underwater images. Such a process could give new insight to historical image datasets and provide an alternative to measuring turbidity when lower accuracy is acceptable, such as in citizen science and education applications. We used machine vision to create an image classification and regression model trained on image data and their corresponding turbidity values recorded from a turbidimeter. To create a robust model, we collected data for model training from a combination of in situ field sites and lab mesocosms across suspended sediment and colorimetric profiles, with and without a Secchi disk for visual standard and binned images into 11 classes 0-55 Formazin Nephelometric Units (FNU). Our resulting classification model is highly accurate with 100% of predictions within one class of the expected class, and 84% of predictions matching the expected class. Regression results provide a continuous value that is accurate to ± 0.7 FNU of true values below 2.5 FNU and ± 33% between 2.5 and 55 FNU; values that are less accurate than conventional turbidimeters but comparable to field-based test kits frequently used in classroom and citizen science applications. To make the model widely accessible, we have implemented it as a free and open-source user-friendly web, computer, and Google Play application that enables anyone with a modern device to make use of the tool, the model, or our repository of training images for data collection or future model development.

Commented [LP1]: I suggest to rephrase this sentence.

Turbidiy sensors already exist and are able, without "additional steps", to measure turbidity fast and accuratly in waters. Ex: https://www.endress.com/en/field-instruments-overview/liquid-analysis-product-overview/turbidity-drinking-water-sensor-cus52d?t.tabId=product-overview

Commented [LP2]: Long sentence, consider splitting

Introduction

In fields ranging from environmental science to public health, assessing water quality is vital. These assessments of water quality often begin with measuring turbidity, which can impact water clarity (Davies-Colley and Smith, 2001) and potability (LeChevallier et al., 1981). Turbidity is defined as a measure of particles suspended in water, contributing to a murky or cloudy appearance caused by the scattering of light. In nature, many of these particles are agitated sediment, such as clays and soils or suspended organic matter (e.g., plant debris or microscopic organisms) and largely depend on surrounding land use (Moreno Madriñán et al., 2012). Pollutants contained in industrial and agricultural runoff can also be linked to turbidity (Rügner et al., 2013; WHO, 2017). Increasing turbidity can have negative impacts on aquatic life, since it often shares an inverse correlation with dissolved oxygen levels, creating an inhospitable environment for many taxa (Talke et al., 2009). In addition, increasing turbidity can cause nonlethal effects on aquatic taxa through altered predator-prey interactions (Abrahams and Kattenfeld, 1997; Ferrari et al. 2010), decreasing light penetration and reduced photosynthesis (Moore et al., 1997), and physical stress such as gill deformities (Lowe et al. 2015). Beyond affecting aquatic ecosystems and taxa, overly turbid water is unsuitable for consumption by humans (Muoio et al., 2020) and livestock (Umar et al., 2014), as well as acting as an indicator

Commented [LP3]: As developed in my response to your response to my initial comments, turbidity is an optical measurement of water clarity, not a measure of suspended particles. If you leave this statement as it is, which I do not recommend, please provide a reference.

Commented [LP4]: I suggest to make a new paragraph here, splitting this big intro paragraph into two.

of bacterial contamination (Gharibi et al., 2012). High turbidity also have an impact on the water treatment processes. , this necessitates filtering suspended particles to bring turbidity within acceptable levels of 5 Nephelometric Turbidity Units (NTU) or lower, depending on use and filtration method (WHO, 2017). Furthermore, monitoring turbidity can be helpful in tracking sediment runoff and pollution in bodies of water, providing valuable insights for environmental management and conservation efforts (Owens et al., 2005).

Typically, turbidity is measured with a turbidimeter which shines a source of light – either white or infrared – into the fluid and a probe which measures the resulting scatter of light. Meters that measure in Formazin Nephelometric Units (FNU) shine an infrared light into the solution and measure the scatter at a 90-degree angle of incidence. Other common units for turbidity include NTU, which are measured with white light at a 90-degree angle of incidence, and Formazin Attenuation Units (FAU), which are measured with infrared light at a 180-degree angle of incidence (Anderson, 2005). Agency standards are typically based on method of measurement rather than accuracy thresholds. For example, U.S. Environmental Protection Agency standards require 90-degree hatchure and visible radiation, with equipment tested by the U.S. Geological Survey (USGS) having 5% error or less. These results are similar to the International Standards Organization (ISO) 7027 which require a back-scatter angle of 90 degrees. Testing by USGS found these instruments to also have less than 5% error at turbidity above 40 NTU and greater than 10% below 20 NTU (Wilde and Gibs, 2008).

Turbidimeters are generally expensive, costing hundreds or even thousands of dollars. For example, the LaMotte 2020i turbidimeter used during this project has an MSRP of $1,449 USD (LaMotte, 2024). This expense can be a significant barrier to citizen science initiatives, or laboratories with a limited budget, decreasing the ability of laypeople to contribute to water quality monitoring efforts. The converse can also be true, that lowering costs and creating easy and consistent methods of data collection increase the quality of data and participation in data collection (Zheng et al., 2018; Lee et al., 2020). Measuring turbidity with a turbidimeter can also be time-consuming, as they necessitate the transport of the meter itself, or the collection of individual water samples for later measurement.

With the expansion and increasing accessibility of machine vision models their use for water quality has become an emerging field of research and offers potential to reduce time, costs, and increase accessibility of data collection. Current approaches include machine vision in conjunction with existing analytical tools (Yan et al., 2024), model development and image analysis from controlled environments (Nazemi Ashani et al., 2024), and remote sensing (Leeuw et al., 2018). The focus of machine vision research for water quality has largely been in economically important fields such as aquaculture (Li and Du, 2022) and wastewater treatment (Mullins et al., 2018). However, there has been little work with model development from in situ images for specific water quality parameters, such as turbidity. Furthermore, while historical underwater image datasets exist (e.g., Peng et al. 2023) we have been unable to identify any historical datasets of underwater images that include associated turbidity measurements, making it impossible to retroactively assess water quality for historical datasets.

Commented [LP5]: This sentences read strangely. I recommend rephrasing to something along these lines:

Increasing turbidity in natural waters have negative impact on aquatic life for different reasons. First, it is often correlated to a decrease in DO concentration, making the water inhospitable. Second, because it decrease light penetration, photosynthesis is impacted. Third, due to [effect of turbidity] …, [consequence].

Commented [LP6]: It is really counterintuitive to place a statement about the benefits of turbidity at the end of a paragraph listing negative effects of turbidity. I suggest to move this statement somewhere else / to remove it.

Commented [LP7]: I am not sure whether these couple of sentences are relevant in this manuscript. Please consider removing them if they do not serve a specific purpose.

Commented [LP8]: Being *expensive* is a relative statement. I would argue that in comparison with most water monitoring equipment, turbidimeters are rather cheap.
I suggest:
Turbidimeters typically cost thousands of dollars.

Commented [LP9]: I do not really understand what you mean by this.

Our primary goal was to develop a machine vision model capable of estimating turbidity from underwater images that could be made publicly available and easily accessible. Our model offers a cost-effective alternative to traditional turbidimeters, making water quality monitoring more accessible for those who may not have the necessary funds to purchase expensive equipment. Given the affordability and widespread availability of waterproof digital cameras, including smartphones, this approach has the potential to democratize non-critical water quality assessment (e.g., Zheng et al., 2022). Using images can simplify the process for field sampling by requiring less equipment and eliminate sample processing. It also allows existing underwater image data without turbidity readings to be retroactively analyzed. We also evaluate the feasibility and accuracy of machine vision for predicting turbidity levels in water bodies. By assessing how well the model performs, this initiative could pave the way for innovative applications in research and citizen science, as well as gaining insights into historical data, fostering greater engagement and participation in environmental monitoring and conservation efforts.

Materials & Methods

Photo/Data Collection

To develop this model, we paired two components in data collection: an underwater photograph, and a measurement of the turbidity of the water source in the image. The turbidity was measured in FNU, and the readings were taken with a LaMotte 2020i Turbidity Meter. Two cameras were used for photography, an Olympus TG-6, and a Sony HDR-AS30V, with photos taken in the "Auto" photo mode. For each photo we recorded associated metadata for location, whether the water was flowing or not (y/n), whether a Secchi disk was present in the photo (y/n), camera information such as ISO, shutter speed, focal length, F-stop, white balance (when available), and the substrate present in the image (when possible). All photos were taken under ambient light conditions.

We collected field photos to represent a diversity of habitats with the intent of improving model robustness from water sources including rivers, lakes, ponds, and the ocean at turbidity levels between 0 and 55 FNU and calibrated the turbidity meter at each sampling location. The images were collected from bodies of water in Pennsylvania and Maryland, USA. Of the field photos taken, 298 photos were from lotic, 30 from lentic, 87 from brackish, and 25 from marine sources. We fixed the camera to the end of a metal rod and submerged it underwater either with or without a 4.2 cm Secchi disk in frame at either 12 or 23 cm from the camera. Images were collected by taking individual photographs while the camera was submerged and by saving a selection of frames from a video taken while the camera was submerged. We took care to avoid disturbing sediment to ensure the FNU reading that was recorded matched the FNU of the water in the image. We took controlled experimental photos in two different systems following the same procedure as field photos, an opaque 20 L bucket to compare with lentic field photos and a 100 L acrylic fish tank with an additional sheet of acrylic placed vertically in the tank center. The

Commented [LP10]: Long sentence, you can split it here

100 L tank included pumps to create a recirculating system to compare with lotic photos from the field.

The standard procedure for data collected from the bucket followed a graduated increase in turbidity with photos taken after each sediment addition. We added sediment in 4-gram increments and stirred to homogenize. When the water settled, we would take another turbidity reading, and additional photos. We repeated this process multiple times until the turbidity reached the upper limit of our defined range (55 FNU). To increase the robustness of the model against colorimetric changes, we also collected photos with colored ink (J. Herbin: Perle Noire, Rouge Hematit, Gris Nuage, and Vert Empire) added to the water in increments of 0.01mL/L, both with and without sediment. We repeated the same process in the recirculating aquarium with gravel added to the bottom of the tank for background texture. In total, we used 675 images in model development: 440 from the field, and 235 from the lab (114 from the bucket and 121 from the aquarium). For 38% of all photos we included the 4.2 cm diameter Secchi. Secchi and natural images were collected and equal proportions until training demonstrated no effect of Secchi presence on model effectiveness.

Commented [LP11]: In this part of the experiment, was the camera also submerged? Please clarify this

Data Preparation

Using the master image dataset, the images were broken into 11 groups (classes) based on the corresponding FNU values: 0-0.49, 0.5-0.99, 1-2.49, 2.5-4.99, 5-9.99, 10-14.99, 15-20.99, 21-28.99, 29-36.99, 37-44.99, 45-55. These classes were used as they represent the maximum number of classes we could create without reducing accuracy. We split data into training, testing, and validation groups without duplicate images within any two or all three of them. Approximately 75% of the images were used in training, 15% for testing, and 10% for validation. This training/testing/validation split follows the standard folder structure for classification datasets required for model training. All python scripts made for the project were run using Jupyter Lab, and the notebook files that contain them are included with the source code. The master dataset, including all images and associated data can be found in the GitHub repository or dataset archive (Rudy & Wilson, 2024a).

Model Training

We used the most recent release of You Only Look Once (YOLO), Ultralytics YOLOv8 model for training [add reference here]. The YOLO family of machine vision models includes object detection, image segmentation models, and classification models. We chose the YOLOv8-Classification model because of the well-established performance, computing times, and open access nature (Jocher et al., 2023). YOLOv8 provides five different size models to use as the basis for transfer learning when training a model. From smallest to largest they include, Nano, Small, Medium, Large, and Extra Large. There are different sets of these models for object detection, instance segmentation, pose estimation, object detection with oriented bounding boxes, and classification tasks, all pretrained on the ImageNet dataset. We used the Large classification model as the starting point for transfer learning which uses the COCO dataset (Lin et al. 2014). Using transfer learning and starting with a pretrained model allowed us to increase the speed of the training process by eliminating the need to build up weights suited for classification, opting instead to modify a set of pre-existing weights to better suit the new task of estimating turbidity. The final model was trained to use 320x320 pixel images, resized from the dataset, and was trained for 15 epochs, with a batch size of 16. The final model was chosen as the model with the highest fitness score from a group of 50 models trained with the specified parameters, each with different hyperparameters mutated using the AdamW optimizer and called using the tuning function provided by Ultralytics (Jocher et al., 2023). The hyperparameters of the final model are provided alongside the archived source code (Rudy & Wilson, 2024b). Training all 50 models during the final tuning process took approximately 2.2 hours using an Nvidia RTX 4080 model GPU.

Regression Prediction

Once the final classification model was trained, we performed inference on each image in the master dataset, which saves the confidence values between 0 and 1 of the top five most confident classes for each image in individual text files. From this we applied the confidence scores for each image and set all blank entries to a value of 0. We then compared these values to measured FNU by fitting a multiple linear regression model using the Weighted Least Squares method, with the predicted classes as the independent variables, and the measured value as the target variable to create a continuous estimate of turbidity in addition to the defined bins.

Results

The YOLOv8 training process generated a confusion matrix to visualize the performance of the model when inference was run on the validation set. An ideal model would have a perfect correlation, where each predicted class matches the actual class. For our model, predictions from the validation set were tightly clustered around the true classes, with 100% of predictions within one class of the expected class, and 84% of predictions matching the expected class (Figure 1).

By comparison, the regression had an R^2 of 0.975, with the parameters (x) and coefficients (β) of the regression equation of form $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{11} x_{11}$ shown in Table 1, with additional regression metrics in Table 2. While the accuracy of the regression model is not significantly greater than the classification model, the numerical, rather than categorical, estimates provide a more broadly useful result for further analysis (Figure 2). Given that our model predicts turbidity values in a range from 0 FNU to 55 FNU, assessing its accuracy based on the root mean square error (RMSE) does not make sense, as an error of approximately 2 FNU at the high end of the range is considerably less impactful than an error of 2 FNU at the low end. This is why relative root mean square error (RRMSE) and relative standard deviation (RSD) were calculated. While there is no accepted rule for RRMSE or RSD, the lower it is the more accurate the model is, and our 18.46% could generally be considered good. We found that from 0 to 2.5 FNU, 95% of our model's predictions fell within ± 0.7 FNU of the true values, and that from 2.5 to 55 FNU, 95% of our model's predictions fell within ± 33% of the true values. In addition, there was no change in model accuracy when trained based on metadata, such as water body type or Secchi disk presence allowing us to use all images together in the final dataset. We also visually compared subsets of the predicted and actual values against image appearance to confirm there were no patterns between image background and accuracy (Figure 3).

Discussion

In this study, we evaluated the ability of a machine vision model to estimate turbidity and the accuracy of that model. While the accuracy of the model is lower than a physical meter, it is comparable to field test kits typically used in citizen science programs [add reference here], free, and a more accessible way to measure turbidity for those willing to accept the reduced accuracy compared to a traditional turbidimeter. In addition, this model is accurate enough to provide insights on turbidity conditions from historical data - underwater image datasets that never had turbidity recorded alongside the images - which would otherwise be impossible to go back and obtain. As an open-source program, the model can also be integrated into other systems which could be used for any variety of other potential data processing tasks related to underwater imagery. Even though the linear regression model is no more accurate than the classification, a single numerical value with a known confidence level is often more useful for analysis.

Commented [LP12]: Maybe you should include, either in the methods or in appendix, a table with accuracy of commercial turbidimeters, including references

Web App

The Turbidivision web app combines the classification and regression models into one web-based program. It can run on any modern web browser, including those on desktop/laptop computers, as well as on phones and other mobile devices. Once the classification model is downloaded, the code runs fully on the client machine; the server is only used to provide the initial download of webpage code and assets, which should remain cached in the browser for future use. After processing each image, the model outputs confidence values for each class, and the linear regression converts that to a discrete numerical estimate. For each image processed, the name of the image, confidence values for each class, and output of the linear regression are added to a CSV file. The web app has a skeuomorphic GUI designed to look like a turbidimeter, but a version using basic unadorned html elements is available for increased compatibility. The user simply needs to click on the file input button, which opens the file browser, and select the image(s) they want to process. The images will then be processed on the user's computer, and a button to save the output CSV file will appear once the processing is complete. Binary distributions of the web app for Windows (.exe), Linux (AppImage), and Android (.apk), as well as the files for the web app for serving locally, are available (Rudy & Wilson, 2024c). In addition, the Android application has been released on the Google Play store as "Turbidivision" and the web app is available for use (Rudy & Wilson, 2024d).

Future Directions

One potential improvement for future research could be changing the underlying model. We are predicting a continuous variable (turbidity), but instead of directly predicting a continuous variable, we first predicted a discrete variable, a class, and then used that to run a secondary regression model. If we were instead to create a neural network that ends in a fully connected layer which outputs a single continuous variable, from 0 to 55, we could effectively run regression directly on the images. However, such a process would require much more data and the creation of a custom convolutional neural network. The exact amount of data needed to train such a model would depend upon the architecture of the model itself, along with the desired accuracy. In addition, since regression is continuous, and not based on classification, direct model comparison is not possible. However, if we assume that quantizing turbidimeter measurements based on the reported accuracy of the meter provides an adequate analogue to number of classes, we can then calculate a minimum recommended amount of images needed for training such a regression model.

Given that the LaMotte 2020i turbidimeter is accurate to within ± 0.05 FNU between 0-2.5 FNU (LaMotte, 2024), we can say that there are approximately 25 unique ranges of distinct accuracy within this range (2.5/0.1); and since it has a reported accuracy of ± 2% from 2.5-100 FNU, we can split it into 79 unique ranges of distinct accuracy between 2.5 and 55 FNU ($2.5 \times 1.04^{count} > 55, solved\ for\ count$). The sum of these two values, 104, gives us a rudimentary analogue to the number of classes needed. If we assume such a regression model would require a similar amount of data as a classification model and a recommended minimum of 150 images per class (Shahinfar et al., 2020), we can use our analogue class count multiplied by 150 pictures per class to estimate an absolute minimum of 15600 images, ideally with the turbidity values of these images evenly distributed between 0 and 55 FNU. Such a model would not be able to make use of the development speed increases of transfer learning, as all weights would need to be trained from scratch. This training would require a larger dataset, potentially larger than mentioned above, as well as more time required for model training. If such a model were developed with a similar architecture to Yolov8, processing times for the end user should be comparable to the model created in this project.

Within the existing modelling framework, more images, but fewer than those needed for a custom convolutional neural network, would improve the training process and should be able to achieve greater accuracy. More images would allow the model to learn the features of each class better, and a greater variety of images, such as in location, water quality, and lighting conditions, should help the model gain more resilience against features not related to turbidity. A larger dataset could also allow for the creation of more classes (smaller bins), which would allow the model to make more granular predictions.

Conclusions

Measuring turbidity can be critically important for contexts ranging from human health to food webs. Our goal was to determine efficacy, accuracy, and precision of a machine vision model for measuring turbidity from underwater images. We successfully demonstrated the potential of modern machine vision techniques as viable for estimating the turbidity of natural bodies of water, with an accuracy comparable to commercial test kits up to 55 FNU and down to near 0 FNU (e.g. LaMotte Turbidity Test Kit #7519-01). The model we created offers an accessible alternative to traditional turbidimeters and can provide turbidity measurements within an acceptable margin of error in many applications.

The application developed as part of this research project is also the first photo-based turbidity measuring tool accessible to the public. To make the model widely accessible, we implemented it as a free, user-friendly web application that would enable anyone with a modern web-enabled device to make use of the tool. Additionally, the web application is designed to allow the data to be processed on the user's device, keeping their data secure and avoiding unnecessary use of internet bandwidth by preventing the need to upload their files to a server. The app is compatible with a wide range of devices and has a simple user-interface, allowing anyone to easily benefit from the results of this research. In addition, its use of images as an input allows users to retroactively gain insights on turbidity from historical underwater image datasets and understand past trends.

Acknowledgements

We thank Nabeel Siddiqui for his early feedback on model approach and for connecting computer science to freshwater research at Susquehanna University.

References

Abrahams, M.V. and M.G. Kattenfeld. 1997. The role of turbidity as a constraint on predator-prey interactions in aquatic environments. Behavioral Ecology and Sociobiology, 40:169-174. https://doi.org/10.1007/s002650050330

Anderson, C.W. 2005. Turbidity: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 9, Chapter A6.7. https://doi.org/10.3133/twri09A6.7

Davies-Colley, R.J. and D.G. Smith. 2001. Turbidity, suspended sediment, and water clarity: a review. Journal of the American Water Resources Association, 37:1085-1101. https://doi.org/10.1111/j.1752-1688.2001.tb03624.x

Ferrari, M.C., K.R. Lysak, and D.P. Chivers. 2010. Turbidity as an ecological constraint on learned predator recognition and generalization in a prey fish. Animal Behaviour, 79:515-519. https://doi.org/10.1016/j.anbehav.2009.12.006

Gharibi, H., M.H. Sowlat, A.H. Mahvi, H. Mahmoudzadeh, H. Arabalibeik, M. Keshavarz, N. Karimzadeh, and G. Hassani. 2012. Development of a dairy cattle drinking water quality index (DCWQI) based on fuzzy inference systems. Ecological Indicators, 20:228-237. https://doi.org/10.1016/j.ecolind.2012.02.015

Jocher, G., A. Chaurasia, and J. Qiu. 2023. Ultralytics YOLO. Version 8.1.8. [Computer software]. https://github.com/ultralytics/ultralytics.

LaMotte. 2024. 2020i Portable Turbidity Meter. Retrieved 24 April 2024: https://lamotte.com/2020i-portable-turbidity-meter#specifications.

LeChevallier, M.W., T.M. Evans, and R.J. Seidler. 1981. Effect of turbidity on chlorination efficiency and bacterial persistence in drinking water. Applied and Environmental Microbiology, 42:159-167. https://doi.org/10.1128/aem.42.1.159-167.1981

Lee, K.A., J.R. Lee, and P. Bell. 2020. A review of citizen science within the earth sciences: potential benefits and obstacles. Proceedings of the Geologists' Association, 131:605-617. https://doi.org/10.1016/j.pgeola.2020.07.010

Leeuw, T. and E. Boss. 2018. The HydroColor app: Above water measurements of remote sensing reflectance and turbidity using a smartphone camera. Sensors, 18:256-271. https://doi.org/10.3390/s18010256

Li, D. and L. Du. 2022. Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish. Artificial Intelligence Review, 55:4077-4116. https://doi.org/10.1007/s10462-021-10102-3

Lin, T.Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick. 2014. Microsoft COCO: Common objects in context. *In* Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, 13:740-755. Springer International Publishing. https://doi.org/10.48550/arXiv.1405.0312

Lowe, M.L., M.A. Morrison, and R.B. Taylor. 2015. Harmful effects of sediment-induced turbidity on juvenile fish in estuaries. Marine Ecology Progress Series, 539:241-254. https://doi.org/10.3354/meps11496

Moore, K.A., R.L. Wetzel, and R.J. Orth. 1997. Seasonal pulses of turbidity and their relations to eelgrass (*Zostera marina* L.) survival in an estuary. Journal of Experimental Marine Biology and Ecology, 215:115-134. https://doi.org/10.1016/S0022-0981(96)02774-8

Moreno Madriñán, M.J., M.Z. Al-Hamdan, D.L. Rickman, and J. Ye. 2012. Relationship between watershed land-cover/land-use change and water turbidity status of Tampa Bay major tributaries, Florida, USA. Water, Air, and Soil Pollution, 223:2093-2109. https://doi.org/10.1007/s11270-011-1007-2

Mullins, D., D. Coburn, L. Hannon, E. Jones, E. Clifford, and M. Glavin. 2018. A novel image processing-based system for turbidity measurement in domestic and industrial wastewater. Water Science and Technology, 77:1469-1482. https://doi.org/10.2166/wst.2018.030

Muoio, R., C. Caretti, L. Rossi, D. Santianni, and C. Lubello. 2020. Water safety plans and risk assessment: A novel procedure applied to treated water turbidity and gastrointestinal diseases. International Journal of Hygiene and Environmental Health, 223:281-288. https://doi.org/10.1016/j.ijheh.2019.07.008

Nazemi Ashani, Z., M.F. Zainuddin, I.S. Che Ilias, and K.Y. Ng. 2024. A combined computer vision and convolution neural network approach to classify turbid water samples in accordance with national water quality standards. Arabian Journal for Science and Engineering, 49:3503-3516. https://doi.org/10.1007/s13369-023-08064-5

Owens, P.N., R.J. Batalla, A.J. Collins, B. Gomez, D.M. Hicks, A.J. Horowitz, G.M. Kondolf, M. Marden, M.J. Page, D.H. Peacock, and E.L. Petticrew. 2005. Fine-grained sediment in river systems: environmental significance and management issues. River Research and Applications, 21:693-717. https://doi.org/10.1002/rra.878

Peng, L., C. Zhu, and L. Bian. 2023. U-shape transformer for underwater image enhancement. IEEE Transactions on Image Processing, 32:3066-3079. https://doi.org/10.1109/TIP.2023.3276332

Rudy, I. and M. Wilson. 2024a. Underwater Image/Turbidity Dataset (1.0.0) [Data set]. https://doi.org/10.5281/zenodo.10951021

Rudy, I. and M. Wilson. (2024b). Turbidivision: Estimating Turbidity with Underwater Photos - GitHub Repository (1.0.2). https://doi.org/10.5281/zenodo.10982478

Rudy, I. and M. Wilson. (2024c). Turbidivision Web App and Binaries (1.0.2). https://doi.org/10.5281/zenodo.11175905

Rudy, I. and M. Wilson. (2024d). Turbidivision live web app. https://turbidi.vision/.

Rügner, H., M. Schwientek, B. Beckingham, B. Kuch, and P. Grathwohl. 2013. Turbidity as a proxy for total suspended solids (TSS) and particle facilitated pollutant transport in catchments. Environmental Earth Sciences, 69:373-380. https://doi.org/10.1007/s12665-013-2307-1

Shahinfar, S., P. Meek, and G. Falzon. 2020. “How many images do I need?” Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. Ecological Informatics, 57:101085. https://doi.org/10.1016/j.ecoinf.2020.101085

Stormsson, C., G. Jocher, and W. Xin. 2024. Image Classification Datasets Overview. Ultralytics. Retrieved 10 April 2024: https://docs.ultralytics.com/datasets/classify/.

Talke, S.A., H.E. de Swart, and V.N. De Jonge. 2009. An idealized model and systematic process study of oxygen depletion in highly turbid estuaries. Estuaries and Coasts, 32:602-620. https://doi.org/10.1007/s12237-009-9171-y

Umar, S., M.T. Munir, T. Azeem, S. Ali, W. Umar, A. Rehman, and M.A. Shah. 2014. Effects of water quality on productivity and performance of livestock: A mini review. Veterinaria, 2:11-15.

Wilde, F.D. and J. Gibs. 2008. 6.7: Turbidity. U.S. Geological Survey TWRI Book 9, 4/98. 30 pages.

World Health Organization. 2017. Water quality and health - review of turbidity: Information for regulators and water suppliers. Retrieved 10 April 2024: https://iris.who.int/bitstream/handle/10665/254631/WHO-FWC-WSH-17.01-eng.pdf.

Yan, J., J. Lee, L. Liu, Q. Duan, J. Lei, Z. Fu, C. Zhou, W. Wu, and F. Wang. 2024. A novel method for multi-pollutant monitoring in water supply systems using chemical machine vision. Environmental Science and Pollution Research, 31:26555–26566. https://doi.org/10.1007/s11356-024-32791-3

Zheng, F., R. Tao, H.R. Maier, L. See, D. Savic, T. Zhang, Q. Chen, T.H. Assumpção, P. Yang, B. Heidari, and J. Rieckermann. 2018. Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. Reviews of Geophysics, 56:698-740. https://doi.org/10.1029/2018RG000616

Zheng, S., H. Li, T. Fang, G. Bo, D. Yuan, and J. Ma. 2022. Towards citizen science: On-site detection of nitrite and ammonium using a smartphone and social media software. Science of the Total Environment, 815:152613. https://doi.org/10.1016/j.scitotenv.2021.152613