AquaSent-TMMAE: A Self-Supervised Learning Method for Water Quality Monitoring from Spatiotemporal Data

Cara Lee
cara.l@fusionresearch.co
Fusion Research
San Francisco, California, USA

Faisal Nabulsi
faisal.n@fusionresearch.co
Fusion Research
San Francisco, California, USA

Michael Xu
michael.x@fusionresearch.co
Fusion Research
San Francisco, California, USA

Christopher Kan
chris.k@fusionresearch.co
Fusion Research
San Francisco, California, USA

Andrew Kan
andrew.k@fusionresearch.co
Fusion Research
San Francisco, California, USA

Rachel Yun
rachel.y@fusionresearch.co
Fusion Research
San Francisco, California, USA

Talal Nabulsi
talal.n@fusionresearch.co
Fusion Research
San Francisco, California, USA

Bryan Jiang
bryan.j@fusionresearch.co
Fusion Research
San Francisco, California, USA

Isam Kharouf
isam.k@fusionresearch.co
Fusion Research
San Francisco, California, USA

Rana Suleiman
rana.s@fusionresearch.co
Fusion Research
San Francisco, California, USA

Zaid Nabulsi
zaid.n@fusionresearch.co
Fusion Research
San Francisco, California, USA

ABSTRACT

Due to the high cost associated with in situ water quality monitoring, it is challenging to consistently identify and monitor contaminated bodies of water, especially for large or remote areas. High spatiotemporal-resolution Sentinel-2 satellite images present the opportunity to monitor water quality of different bodies of water across long time periods through remote sensing technology. While there exists a plethora of satellite imagery, obtaining labeled data is extremely costly and in many cases infeasible. In this work, we develop and evaluate AquaSent-TMMAE, a new self-supervised learning framework based on the MAE architecture tailored to spatiotemporal remote sensing data and optimized for water quality predictions. Our experiments demonstrate that when finetuned on a specific task using transfer learning, AquaSent-TMMAE is able to accurately detect unsafe levels of various water contaminants, achieving areas under the receiver operating characteristic curve of above 0.90 on multiple downstream tasks. Our work presents an important step toward developing a robust foundation model for satellite imagery.

DeepSpatial '24, August 2024, Barcelona, Spain

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Computing methodologies** → **Unsupervised learning**; • **Information systems** → **Spatial-temporal systems**.

KEYWORDS

spatiotemporal data, masked autoencoder, self-supervised learning, deep learning, remote sensing, water quality

ACM Reference Format:
Cara Lee, Faisal Nabulsi, Michael Xu, Christopher Kan, Andrew Kan, Rachel Yun, Talal Nabulsi, Bryan Jiang, Isam Kharouf, Rana Suleiman, and Zaid Nabulsi. 2024. AquaSent-TMMAE: A Self-Supervised Learning Method for Water Quality Monitoring from Spatiotemporal Data. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (DeepSpatial '24).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The health of aquatic ecosystems is a global concern, with water contamination threatening both environmental sustainability and human wellbeing [20]. Climate change is significantly disrupting water quality. Warmer temperatures and severe storms increase pollution, upsetting the delicate balance of our water system [28]. Furthermore, the impact of human activities on water quality is increasing [27] as substantial amounts of chemical fertilizers and pesticides flow into rivers [1]. Human activities such as mining cause acid mine drainage (AMD), a harmful environmental effect that is the main cause of polluted water in many countries around the world for the last several decades [2]. Contaminated bodies of water pose significant health risks to humans by serving as vectors for waterborne diseases and toxic substances that can lead to serious illnesses. Additionally, these pollutants disrupt aquatic

ecosystems by causing habitat degradation, reducing biodiversity, and impairing the reproductive and growth cycles of aquatic organisms.

Robust water monitoring programs are paramount to addressing the harmful effects of water contamination, both now and in the future. Traditional monitoring approaches, like systematic sampling, laboratory analysis, and spatial interpolation [16], are accurate but less suited for daily use by environmental authorities. They tend to be time-consuming, costly, and limited in their area of coverage. As a result, traditional water monitoring programs are ineffective at addressing the underlying issues [31].

Especially in coal mining areas, Earth remote sensing (ERS) technology offers significant advantages for environmental monitoring. In particular, applying deep learning to imaging spectroscopy provides a powerful alternative or complement to traditional chemical analysis methods for effectively assessing water contamination [18]. Given a large repository of labeled data, specialist convolutional neural networks can be trained to automate the detection of water contaminants from multispectral satellite imagery [12]. However, while several large-scale satellite image datasets have been curated in recent times including Functional Map of the World (fMoW) [7] and BigEarthNet [29], annotating these datasets for specific water contaminants requires obtaining costly laboratory analysis of water samples at the same time the satellite image was captured. In the field of ERS where a myriad of opportunities for automation exist, it is *practically impossible* to curate labeled datasets for all tasks and outcomes in order to train supervised models. Therefore, it is important to develop strategies for training generalist ERS models that can be finetuned for numerous downstream tasks without requiring the curation of large-scale labeled datasets.

In this work, we developed AquaSent-TMMAE (Aquatic Sentinel Temporal Multispectral Masked AutoEncoders), a new self-supervised generalist architecture optimized for detecting various water contaminants from Sentinel-2 satellite imagery. We utilize Sentinel-2 satellites (launched by ESA in 2015 and 2017) in this work because while they offer excellent capabilities for monitoring water contamination [18, 26], their use in this field is surprisingly understudied [6, 25]. These satellites provide data in 10 narrow VNIR spectral bands with 10 and 20-meter spatial resolution. Their frequent image capture (every 2-3 days in clear temperate conditions) further enhances their potential in this space. Our contributions are the following:

- We design a novel self-supervised learning architecture that is optimized for learning robust visual features from spatiotemporal satellite images. Our architecture builds upon Masked Autoencoders [13] by incorporating both a temporal and a multispectral dimension to allow the network to learn nuanced changes in water quality.
- We curate a large dataset consisting of 3,482,319 Sentinel-2 satellite images to pretrain our network on.
- We evaluate our pretrained network (AquaSent-TMMAE) on various downstream water contamination detection tasks and demonstrate state-of-the-art performance when compared to baselines.
- To encourage further research in the field, we release our pretrained network, all training and statistical code, our large-scale pretraining dataset, and our finetuning dataset. We hope that this will aid future work in this field.

Band	Resolution	CW[2]	Mean	Std. dev.
B1: Aerosols	60m	443nm	1412.755	672.123
B2: Blue	10m	490nm	1243.624	702.457
B3: Green	10m	560nm	1281.264	871.231
B4: Red	10m	665nm	1212.754	981.457
B5: Red Edge 1	20m	705nm	1332.325	1098.264
B5: Red Edge 2	20m	740nm	1765.526	1341.375
B5: Red Edge 3	20m	783nm	1874.375	1486.274
B8: NIR	10m	842nm	1834.246	1451.247
B8A: Red Edge 4	20m	865nm	2002.264	3681.457
B9: Water Vapor	60m	940nm	606.645	481.367
B10: Cirrus	60m	1375nm	13.654	12.346
B11: SWIR 1	20m	1610nm	1834.732	1501.574
B12: SWIR 2	20m	2190nm	1290.543	1100.745

Table 1: Metadata on our curated pretraining dataset. We show the resolution, central wavelength, and mean and standard deviation of pixel values for each band.

2 MATERIALS & METHODS

2.1 Overview

Due to the limited availability of paired water quality and satellite imagery data, conventional supervised learning methods are not expected to yield optimal performance in predicting diverse water quality indicators from overhead satellite imagery. [9]. However, there does exist a plethora of unpaired (unlabeled) satellite imagery. To utilize this unlabeled data, we develop AquaSent-TMMAE in two stages. First, we conduct self-supervised pretraining on unlabeled data to develop a generalist model for satellite imagery and optimized our image masking method so the model can learn nuanced water quality features. Then, we take the pretrained encoder, add a single linear layer on top, and finetune it on various downstream tasks in our supervised finetuning (SFT) stage using smaller, labeled datasets. A schematic of our method is shown in Figure 1.

2.2 Pretraining Dataset Generation

To curate a pretraining dataset, we first obtain a list of latitude and longitude pairs that correspond to bodies of water from the US Geological Survey [1]. Next, we randomly sample these coordinate pairs, and for each point, we obtain three Sentinel-2 satellite images of the same location captured on different dates, each consisting of the full set of 13 Sentinel-2 frequency bands (B1-B12 and B8A). We choose three dates randomly, spaced at least one year apart, from the set of historical Sentinel-2 visits for that location. This process results in the aggregation of three images per location, totaling 3,482,319 images. See Table 1 for further details regarding the dataset, and Table 3 for dataset counts.

2.3 Self-Supervised Pretraining

The goal of the self-supervised pretraining stage is to make use of the large scale unlabeled dataset in order to develop a generalist

[1] https://www.usgs.gov

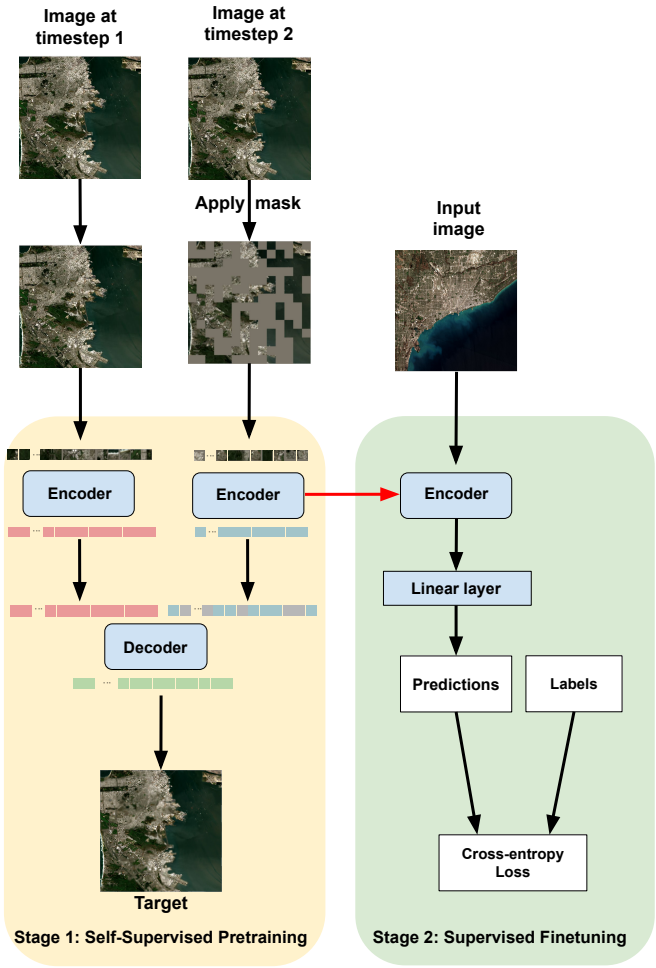

Figure 1: Schematic of AquaSent-TMMAE. Our method combines elements of SatMAE [8] and SiamMAE [10], with carefully designed masking strategies across multispectral and temporal dimensions. Once the network is pretrained, we finetune the entire network on a smaller, labeled dataset for a specific downstream detection task.

water quality satellite imagery model that can produce meaningful satellite image representations that can be later finetuned for many water-quality focused downstream tasks without the need for large-scale labeled datasets. Self-supervised learning encompasses two primary classes of methods: contrastive methods and generative methods [21, 23]. Contrastive methods depend on data augmentations (such as color jittering and blur), which are known to lose spectrum information and spatial details found in satellite imagery [11]. Thus, in this work, we leverage the benefits of generative methods and develop a temporal multispectral masked autoencoder (TMMAE) to learn water-quality focused ERS representations.

Masked autoencoder (MAE) [13] is a recently-developed, powerful self-supervised learning method that patches the input image and attempts to reconstruct it. Motivated by MAE's impressive performance on various vision tasks, many studies have applied MAE to different modalities [3–5, 14, 15, 17, 19, 22, 30]. SatMAE [8] is an extension of MAE developed specifically for either temporal or multispectral satellite imagery. However, this method does not work on both temporal and multispectral data.

In our case, we aim to apply MAE to satellite imagery which is both multispectral and temporal. The high revisit frequency of Sentinel 2 satellites allows us also to explore using temporal satellite images (i.e. multiple satellite images of the same coordinates over time). The reasoning for using temporal data is that water quality changes subtly over time, and by showing the network a satellite image of the same location over time, it can focus on representing the nuanced changes/features of the water quality as opposed to the more obvious features present in a single satellite image.

To incorporate temporal data, we utilize a Siamese encoder setup [10]. Specifically, during training, for every location in our pretraining dataset, we randomly select 2 satellite images from different dates, and pass them through the same encoder independently as described in [10]. By masking a large fraction of patches (95%) from the second image and leaving the first image unchanged, we encourage the network to focus on representing the changes between the two images. As proposed in [8], we apply consistent masking across the spectral dimension. See Figure 1 for a detailed visual of our architecture.

2.4 Supervised Finetuning Dataset Generation

Our SFT data consists of satellite images paired with binary water quality indicator labels. In this work, we explore seven water indicators: AMD, pH, conductivity, zinc, iron, manganese, and sulfate. The AMD label is derived from the other 6 (see 2.5). Our labels are sparse, meaning that for every satellite image, we only have labels for some of the indicators, and most of the labels are missing. This is because for each water sample, not all contaminants were measured.

We curate our SFT data from three states in the United States: California, Colorado, and Pennsylvania. We chose these states as their bodies of water are most likely to contain high concentrations of various water contaminants [24].

For each state, we obtain all water quality data from the US Geological Survey (USGS) Water Quality Portal [3] from 2015 through the end of 2022. We filter the water quality data, only keeping samples that measured at least one of our six contaminants of interest. For each water quality data point, we obtain the satellite image (the full set of 13 Sentinel-2 frequency bands) of the corresponding body of water from the closest date to the date of the water sampling. We binarize the water quality measurements based on the thresholds in Table 2. See Table 3 for a summary of our dataset counts.

To avoid finetuning and evaluating on the same bodies of water, we use all data from California and Colorado as our training sets, and reserve data from Pennsylvania as evaluation data.

2.5 Acid Mine Drainage Label

In this work, one of the downstream tasks we evaluate our method on is detecting the presence of AMD. AMD itself is not a contaminant that can be measured through laboratory analysis of water samples. Instead, it generally manifests as a combination of the six

[3] https://www.usgs.gov/tools/water-quality-portal

Acid mine drainage	Indicator Thresholds
pH	<5.5
Sulfate	>250 mg/L
Conductivity	>500 uS/cm
Iron	>2 mg/L
Manganese	>0.1 mg/L
Copper	>0.02 mg/L
Zinc	>0.1 mg/L

Table 2: The thresholds we used to binarize each of the indicators. For example, if a measurement had a pH less than 5.5, then it would be positive and is negative otherwise. These measurements were obtained from the US Geological Survey (USGS).

Dataset	Total	Negatives	Positives
Pretraining	3,482,319	N/A	N/A
AMD Train	1460	1271	189
AMD Evaluation	906	802	104
pH Train	1215	1107	108
pH Evaluation	787	703	84
Conductivity Train	1734	1426	308
Conductivity Evaluation	1053	902	152
Zinc Train	948	804	144
Zinc Evaluation	475	403	72
Iron Train	1511	1303	208
Iron Evaluation	702	601	101
Manganese Train	1233	1821	412
Manganese Evaluation	1231	1003	228
Sulfate Train	2571	2090	481
Sulfate Evaluation	1372	1090	282

Table 3: Number of images in our pretraining dataset and number of training/evaluation examples in our SFT dataset for each of the downstream tasks.

contaminants in this study. To derive the AMD label, we use the following procedure:

- Obtain all the satellite images that have a label for least 3 of the 6 indicators. Filter out all other images as they are too sparse.
- If at least 50% of the labeled indicators are positive, then label the image as positive for AMD.
- Otherwise, label the image as negative for AMD.

2.6 Supervised Finetuning

With a pretrained generalist network at hand, we evaluate its performance on a total of seven binary downstream water quality tasks. Given a satellite image containing a body of water, the tasks are to classify (1) presence of AMD, (2) presence of unsafe amounts of manganese, (3) presence of unsafe amounts of sulfate, (4) presence of unsafe amounts of zinc, (5) presence of unsafe amounts of iron, (6) low pH, and (7) high conductivity. We obtain train/evaluation data splits for each of the downstream tasks, consisting of paired satellite images containing a body of water and a binary label (see 2.4). We use a multi-task formulation (multiple binary tasks) as opposed to a single multiclass task because every image had only labels for some classes, and a given image can have any number of the classes present.

We use the same encoder from the pretraining stage and add a single linear layer (linear probe) on top to generate the predictions. For each of the downstream tasks, we finetune the entire network independently, initializing from the pretrained network each time.

To evaluate performance, for each downstream task, we measure area under the receiver operating characteristic (AUROC) on the held-out evaluation split of the data for that task.

Parameter	value
optimizer	AdamW
optimizer momentum	β_2, β_1 = 0.95, 0.99
weight decay	0.05
learning rate	1.5e-4
epochs	1200
augmentation	hflip, crop [0.5, 1]
batch size	1024
frame sampling gap	[4, 48]

Table 4: AquaSent-TMMAE Pretraining Hyperparameter Settings.

2.7 Additional Training Details

We follow the training settings in [13] and we develop our implementation in PyTorch, building on top of the open-source implementation of MAEs (https://github.com/facebookresearch/mae). We use the same parameters specified in [13], unless otherwise specified in Table 4. We train all experiments on 4 Nvidia V100 GPUs.

3 RESULTS

Our results on the downstream water quality tasks are presented in Figure 2. AquaSent-TMMAE achieves high AUROCs with 3 tasks above 0.90 and at least 0.83 on all tasks except the presence of unsafe levels of zinc, where it achieves an AUROC of 0.74. Ablation study results comparing AquaSent-TMMAE to baselines and other design choices is shown in Table 6. Briefly, AquaSent-TMMAE outperforms all baselines, with the closest method being the use of independent spectral masking.

The sensitivity and specificity of AquaSent-TMMAE at a specific threshold are shown in Table 5.

To qualitatively understand the discriminative capabilities of the encoder after finetuning on the task of AMD classification, we take the embeddings outputted from the finetuned encoder (1408-dimension) and downsample to a two-dimensional space using principal component analysis. We visualize the resultant latent space embeddings in Figure 3. The encoder can clearly discriminate between satellite images of bodies of water contaminated by AMD and satellite images of bodies of water that are not contaminated by AMD.

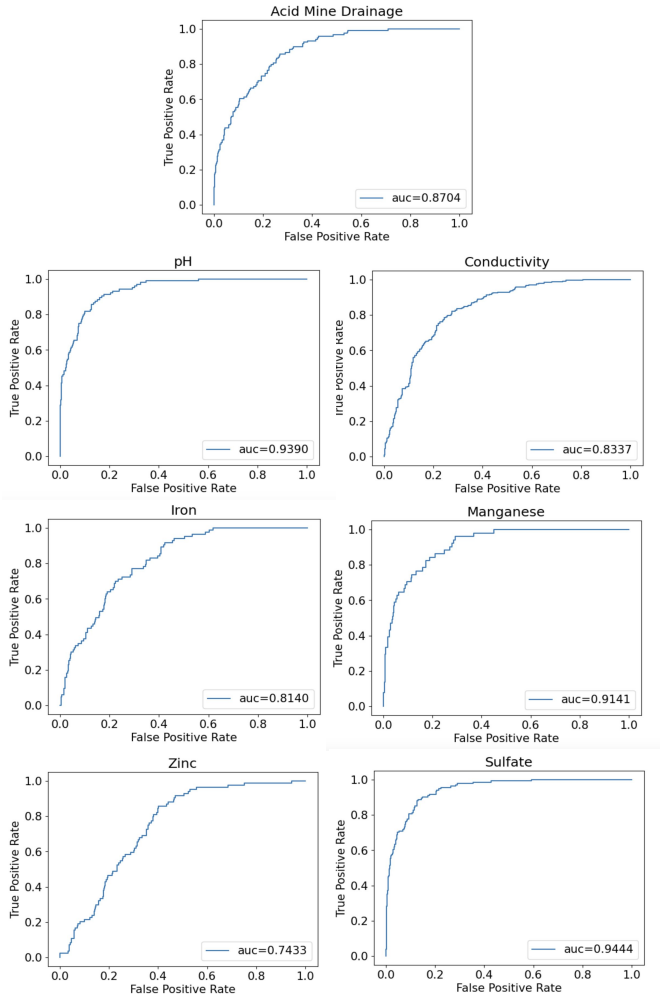

Figure 2: The area under the receiver operating characteristic curves for all seven binary downstream tasks on the held-outevaluation splits.

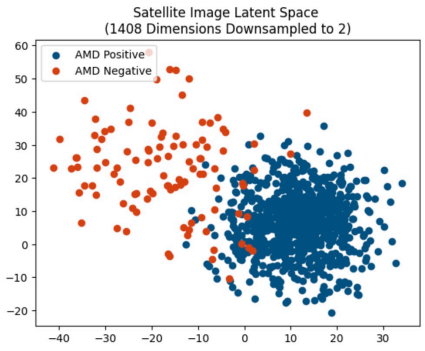

Figure 3: A visual representation of the embeddings (output of the finetuned encoder) after being reduced to two dimensions.

Metric	Value (Threshold: 0.47)
Sensitivity	90.59
Specificity	77.94

Table 5: Sensitivity and Specificity of AquaSent-TMMAE on the task of detecing AMD.

Method	Average AUROC
AquaSent-TMMAE	0.8641
Traditional Supervised Learning	0.5960
No Temporal Masking	0.7930
Independent Spectral Masking	0.8481
No Finetuning	0.8372
SatMAE [8]	0.8213

Table 6: Performance comparison of the various baselines and design decisions. across seven downstream tasks. We conducted an ablation study of various design options before arriving at AquaSent-TMMAE. Average AUROC is evaluated across the seven downstream tasks.

4 BASELINES & ABLATION STUDY

In developing AquaSent-TMMAE, we performed ablation studies of different design choices and baselines. Results are shown in Table 6. To ensure a fair comparison across the methods, we kept all other aspects (i.e. model architecture, hyperparameter sweep strategy, datasets) consistent. Here are the following variations we explored in this work:

- *Traditional Supervised Learning*: Here, we do not pretrain the network and simply just train it on the labeled data for the specific task.
- *No Temporal Masking*: Instead of using Siamese encoders in the pretraining stage and having a temporal dimension, we use a vanilla MAE with consistent masking across the spectral dimension. Performance is substantially worse, demonstrating the benefit of having a temporal dimension, which forces the network to focus its representations on the changes of the satellite image between dates.
- *Independent Spectral Masking*: Instead of applying consistent masking on the spectral dimension (as proposed in [8]), this approach uses independent masking, as defined in [8]. We note that performance is marginally lower than consistent masking.
- *No Finetuning*: Instead of finetuning the entire encoder for each downstream task, this approach freezes the weights of the encoders and only trains the linear layer (linear probing).
- *SatMAE [8]* : As described in the paper, this approach uses temporal but not multispectral masking. This shows the benefit of having a multispectral dimension, since performance is worse without it.

5 CONCLUSION

The use of remote sensing for water quality monitoring presents a monumental opportunity to safeguard our water resources and promote environmental sustainability. While there has been a plethora

Figure 4: Reconstruction quality of AquaSent-TMMAE. The top row shows the original images, and the bottom row shows AquaSent-TMMAE's reconstructed images.

of prior work on developing computer vision algorithms for predictive tasks from satellite imagery, there is little prior work specifically for water quality prediction, largely because of the lack of labeled datasets. In this work, we propose a new self-supervised learning framework based on the MAE architecture tailored to multipsectral, temporal remote-sensing data and optimized for water quality predictions. Our novel masking strategy in a joint temporal and spectral space enables AquaSent-TMMAE to learn meaningful representations. Our experiments on seven downstream water quality prediction tasks demonstrate the effectiveness of AquaSent-TMMAE.

Because all downstream water quality tasks achieved an AUROC of 0.83 or greater except zinc, which achieved an AUROC of 0.74, our work has shown the promising potential of automating water quality monitoring using deep learning on satellite imagery. Further research could include running an ablation study to generate even more data using image augmentation techniques to test the model's consistency in detecting unsafe water. In addition, a system of satellite water safety detection paired with physical on-site tests has numerous advantages to current detection techniques, namely faster detection, avoiding blindsights, and the idea that overdetection is better than underdetection when regarding water safety. It would also be advantageous to expand our research beyond the realm of water quality and develop a more general foundational model for remote sensing data.

ACKNOWLEDGMENTS

We would like to thank Professor Stefano Ermon for his mentorship and guidance throughout the course of this study. We also would like to thank Aditya Grover for his help in problem formulation and advice on dataset creation. Furthermore, we would like to thank Evan Sheehan for his advice on architectural design choices and for sharing resources on generative self-supervised learning. Lastly, we would like to thank Vineet Kosaraju and Curtis Staples for their feedback on and review of the manuscript.

REFERENCES

[1] Kuk-Hyun Ahn and Venkatesh Merwade. 2014. Quantifying the relative impact of climate and human activities on streamflow. *Journal of Hydrology* 515 (2014), 257–266. https://doi.org/10.1016/j.jhydrol.2014.04.062

[2] Ata Akcil and Soner Koldas. 2006. Acid Mine Drainage (AMD): causes, treatment and case studies. *Journal of Cleaner Production* 14, 12 (2006), 1139–1145. https://doi.org/10.1016/j.jclepro.2004.09.006 Improving Environmental, Economic and Ethical Performance in the Mining Industry. Part 2. Life cycle and process analysis and technical issues.

[3] Roman Bachmann, David Mizrahi, Andrei Atanov, and Amir Zamir. 2022. MultiMAE: Multi-modal Multi-task Masked Autoencoders. arXiv:2204.01678 [cs.CV]

[4] Sebastien Baur, Zaid Nabulsi, Wei-Hung Weng, Jake Garrison, Louis Blankemeier, Sam Fishman, Christina Chen, Sujay Kakarmath, Minyoi Maimbolwa, Nsala Sanjase, Brian Shuma, Yossi Matias, Greg S. Corrado, Shwetak Patel, Shravya Shetty, Shruthi Prabhakara, Monde Muyoyeta, and Diego Ardila. 2024. HeAR – Health Acoustic Representations. arXiv:2403.02522 [cs.LG] https://arxiv.org/abs/2403.02522

[5] Louis Blankemeier, Joseph Paul Cohen, Ashwin Kumar, Dave Van Veen, Syed Jamal Safdar Gardezi, Magdalini Paschali, Zhihong Chen, Jean-Benoit Delbrouck, Eduardo Reis, Cesar Truyts, Christian Bluethgen, Malte Engmann Kjeldskov Jensen, Sophie Ostmeier, Maya Varma, Jeya Maria Jose Valanarasu, Zhongnan Fang, Zepeng Huo, Zaid Nabulsi, Diego Ardila, Wei-Hung Weng, Edson Amaro Junior, Neera Ahuja, Jason Fries, Nigam H. Shah, Andrew Johnston, Robert D. Boutin, Andrew Wentland, Curtis P. Langlotz, Jason Hom, Sergios Gatidis, and Akshay S. Chaudhari. 2024. Merlin: A Vision Language Foundation Model for 3D Computed Tomography. arXiv:2406.06512 [cs.CV] https://arxiv.org/abs/2406.06512

[6] Varun Chitturi and Zaid Nabulsi. 2021. Predicting Poverty Level from Satellite Imagery using Deep Neural Networks. arXiv:2112.00011 [cs.CV] https://arxiv.org/abs/2112.00011

[7] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. 2018. Functional Map of the World. arXiv:1711.07846 [cs.CV]

[8] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. 2023. SatMAE: Pre-training Transformers for Temporal and Multi-Spectral Satellite Imagery. arXiv:2207.08051 [cs.CV]

[9] Linus Ericsson, Henry Gouk, Chen Change Loy, and Timothy M. Hospedales. 2022. Self-Supervised Representation Learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine* 39, 3 (May 2022), 42–62. https://doi.org/10.1109/msp.2021.3134634

[10] Agrim Gupta, Jiajun Wu, Jia Deng, and Li Fei-Fei. 2023. Siamese Masked Autoencoders. arXiv:2305.14344 [cs.CV]

[11] Xuejie Hao, Lu Liu, Rongjin Yang, Lizeyan Yin, Le Zhang, and Xiuhong Li. 2023. A Review of Data Augmentation Methods of Remote Sensing Image Target Recognition. *Remote Sensing* 15, 3 (2023). https://doi.org/10.3390/rs15030827

[12] N Hassan and C S Woo. 2021. Machine learning application in water quality using satellite data. *IOP Conference Series: Earth and Environmental Science* 842, 1 (Aug 2021), 012018. https://doi.org/10.1088/1755-1315/842/1/012018

[13] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. 2021. Masked Autoencoders Are Scalable Vision Learners. arXiv:2111.06377 [cs.CV]

[14] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. 2022. GraphMAE: Self-Supervised Masked Graph Autoencoders. arXiv:2205.10803 [cs.LG]

[15] Sahar Kazemzadeh, Jin Yu, Shahar Jamshy, Rory Pilgrim, Zaid Nabulsi, Christina Chen, Neeral Beladia, Charles Lau, Scott Mayer McKinney, Thad Hughes, Atilla P Kiraly, Sreenivasa Raju Kalidindi, Monde Muyoyeta, Jameson Malemela, Ting Shih, Greg S Corrado, Lily Peng, Katherine Chou, Po-Hsuan Cameron Chen, Yun Liu, Krish Eswaran, Daniel Tse, Shravya Shetty, and Shruthi Prabhakara. 2023. Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists. *Radiology* 306, 1 (Jan. 2023), 124–137.

[16] Thomas Kemper and Stefan Sommer. 2002. Estimate of Heavy Metal Contamination in Soils after a Mining Accident Using Reflectance Spectroscopy. *Environmental Science & Technology* 36, 12 (2002), 2742–2747.

[17] Atilla P. Kiraly, Corbin A. Cunningham, Ryan Najafi, Zaid Nabulsi, Jie Yang, Charles Lau, Joseph R. Ledsam, Wenxing Ye, Diego Ardila, Scott M. McKinney, Rory Pilgrim, Yun Liu, Hiroaki Saito, Yasuteru Shimamura, Mozziyar Etemadi, David Melnick, Sunny Jansen, Greg S. Corrado, Lily Peng, Daniel Tse, Shravya Shetty, Shruthi Prabhakara, David P. Naidich, Neeral Beladia, and Krish Eswaran. 0. Assistive AI in Lung Cancer Screening: A Retrospective Multinational Study in the United States and Japan. *Radiology: Artificial Intelligence* 0, ja (0), e230079. https://doi.org/10.1148/ryai.230079 arXiv:https://doi.org/10.1148/ryai.230079 PMID: 38477661.

[18] Veronika Kopačková. 2019. Mapping Acid Mine Drainage (AMD) and Acid Sulfate Soils Using Sentinel-2 Data. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. 5682–5685. https://doi.org/10.1109/IGARSS.2019.8900505

[19] Vineet Kosaraju, Yap Dian Ang, and Zaid Nabulsi. 2019. Faster Transformers for Document Summarization. *Vineet Kosaraju* (2019).

[20] Li Lin, Haoran Yang, and Xiaocang Xu. 2022. Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. *Frontiers in Environmental Science* 10 (2022). https://doi.org/10.3389/fenvs.2022.880246

[21] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. 2021. Self-supervised Learning: Generative or Contrastive. *IEEE Transactions on Knowledge and Data Engineering* (2021), 1–1. https://doi.org/10.1109/tkde.2021.3090866

[22] Zaid Nabulsi, Vineet Kosaraju, and Shuvam Chakraborty. 2019. MRNGAN: Reconstructing 3D MRI Scans Using A Recurrent Generative Model. *Vineet Kosaraju* (2019).

[23] Zaid Nabulsi, Andrew Sellergren, Shahar Jamshy, Charles Lau, Edward Santos, Atilla P. Kiraly, Wenxing Ye, Jie Yang, Rory Pilgrim, Sahar Kazemzadeh, Jin Yu, Sreenivasa Raju Kalidindi, Mozziyar Etemadi, Florencia Garcia-Vicente, David Melnick, Greg S. Corrado, Lily Peng, Krish Eswaran, Daniel Tse, Neeral Beladia, Yun Liu, Po-Hsuan Cameron Chen, and Shravya Shetty. 2021. Deep learning for distinguishing normal versus abnormal chest radiographs and generalization to two unseen diseases tuberculosis and COVID-19. *Scientific Reports* 11, 1 (01 Sep 2021), 15523. https://doi.org/10.1038/s41598-021-93967-2

[24] S V Pyankov and N G Maximovich. 2021. Monitoring Acid Mine Drainage's Effects on Surface Water in the Kizel Coal Basin with Sentinel-2 Satellite Images. *Remote Sensing* 13, 23 (2021), 4796.

[25] Aliyeh Seifi, Mahdieh Hosseinjanizadeh, Hojjatolah Ranjbar, and Mehdi Honarmand. 2019. Identification of Acid Mine Drainage Potential Using Sentinel 2a Imagery and Field Data. *Mine Water and the Environment* 38 (2019), 707 – 717. https://api.semanticscholar.org/CorpusID:202579735

[26] Evan Sheehan, Zaid Nabulsi, and Chenlin Meng. 2018. Utilizing Latent Embeddings of Wikipedia Articles to Predict Poverty. *Stanford University* (2018).

[27] N. V. Sidabutar, I. Namara, D. M. Hartono, and T. E.B. Soesilo. 2017. The effect of anthropogenic activities to the decrease of water quality. *IOP Conference Series: Earth and Environmental Science* 67, 1 (1 June 2017). https://doi.org/10.1088/1755-1315/67/1/012034 Publisher Copyright: © Published under licence by IOP Publishing Ltd.; 2017 7th International Conference on Environment and Industrial Innovation, ICEII 2017 ; Conference date: 24-04-2017 Through 26-04-2017.

[28] Signe Stroming, Molly Robertson, Bethany Mabee, Yusuke Kuwayama, and Blake Schaeffer. 2020. Quantifying the Human Health Benefits of Using Satellite Information to Detect Cyanobacterial Harmful Algal Blooms and Manage Recreational Advisories in U.S. Lakes. *GeoHealth* 4 (08 2020). https://doi.org/10.1029/2020GH000254

[29] Gencer Sumbul, Marcela Charfuelan, Begum Demir, and Volker Markl. 2019. Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. https://doi.org/10.1109/igarss.2019.8900532

[30] Zhan Tong, Yibing Song, Jue Wang, and Limin Wang. 2022. VideoMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training. arXiv:2203.12602 [cs.CV]

[31] Sze Yee Wee and Ahmad Zaharin Aris. 2023. Revisiting the "forever chemicals", PFOA and PFOS exposure in drinking water. *npj Clean Water* 6, 1 (Aug 2023). https://doi.org/10.1038/s41545-023-00274-6

Received 11 June 2024