Imperial College London

Department of Earth Science and Engineering

MSc in Environmental Data Science and Machine Learning

Independent Research Project

Final Report

Using Deep Learning to Classify Individual Ponds of Artisanal and Small-scale Gold Mining in Ghana.

by

Olga Kostur

Email: ok121@ic.ac.uk
GitHub username: edsml-ok121
Repository: https://github.com/ese-msc-2021/irp-ok121

Supervisors:
Dr.Yves Plancharel
DR. Pablo Brito Parada

2nd September 2022

Abstract

This project developed a classification algorithm with U-Net that categorises 3 classes of gold mining ponds – actively mined; in transition (recently abandoned) and inactive (abandoned more than 6 months ago). Using Sentinel-2 images, nine classification models were built with different combinations of spectral bands and numbers of output classes. 10-channel models resulted in the highest F-1 scores (0.94-0.77). It was proven that the concentration of minor ferric iron, a proxy of acid mine drainage (AMD) risk, is statistically different between the 3 pond classes on a 99% confidence level. The developed models can be used to quantify the internal dynamics of gold mining wetlands, such as intensification or cessation of activity and the corresponding change in water acidification risk. Three such applications were demonstrated in the South of Ghana, including site comparison, seasonal cycle computation and creating a timeline to quantify how gold mining intensity influences risks of AMD. The results proved that the intensity of gold mining influences AMD risk. The developed tool and proposed applications of it, can be used to track mining activity, design restoration policies, evaluate their success, highlight spots of heavy metal pollution risk, or understand how long it takes for the environment to recover naturally.

Acknowledgements

Firstly, I would like to thank Seda Camalan for providing her unique dataset and giving me valuable insights about gold mining phenomenon. I would like to thank my supervisor Yves Planchered for his great support and guidance during this project. In addition, I would like to thank Pablo Brito Parada for helping me to frame ideas and identify areas of improvements. Finally, I would like to thank my husband – Konstantin, my friends, and my family for being my best support during this difficult time of war in Ukraine and helping me in staying focused on dissertation regardless of the circumstances.

1. Introduction

In Ghana, artisanal small-scale mining (ASM) is an important economic sector and in 2020 1.3 billion US dollars were contributed to GDP via it (Sasu, 2021). Approximately 35% more gold is extracted illegally annually, financially supporting 4.5 million people (McQuilken & Hilson, 2016). Illegal ASM is dangerous not only for miners themselves (high risk of mercury poisoning, kidney, and neurological diseases) but also for the environment - deforestation, water pollution and degradation of arable land (Nicholls et al., 2020). Water is the key element of gold mining, hence in tropics such as South Ghana, where any dig is filled up with water quickly, it is so popular. The unregulated excavation of soil for ASM converts tropical forests to novel hydrological landscapes consisting of large networks of ponds interacting with existing waterbodies and expanding at an average rate of ~2600 ha yr−1 (Barenblitt et al., 2021). When ponds are abandoned, algae start to colonise standing water, turning artificial wetlands into dangerous breeding grounds for malarial mosquitos (Bansah et al., 2018).

Pixel-based classification based on spectral signatures of proxy classes (e.g., bare land from ESA 10m 2020 World Cover Map) is a classical method for ASM detection (Boakye et al., 2019). Though, gold mining landscapes consist of multiple land classes – water, vegetation patches and soil, so models trained on macro classes such as bare land are unrepresentative of complex ASM environments (Obodai et al., 2019). Misclassifications of algae in abandoned ponds with vegetation in deforestation and re-greening studies is common due to the similar spectral signature of chlorophyll, present in both classes (Barenblitt et al., 2021). Thus, the pixel-based models lack spatial context and are unable to detect total mining area.

With advances in remote sensing and higher spatial resolution of satellites, such as Sentinel-2 (10m in visible and near-infrared bands) using deep learning for image segmentation became more popular (Kotaridis, 2021). Convolutional Neural Networks (CNN) address the problem of misclassifying individual pixels within mining sites (algae – vegetation, bare land-urban etc.) by bringing learning interrelationships between pixels. CNN uses highly non-linear functions to learn feature hierarchies and relationships from gridded data, so well-trained models can segment whole objects (Goodfellow et al., 2016). However, in the field of gold mining few studies have adopted such techniques, mainly due to the limited availability of high quality and quantity pre-labelled training data (Signoroni et al., 2019).

Gallwey (2020) was the first to use deep-learning to address problems of pixel-based methods and efficiently detect total land area degraded by gold mining in Ghana. Gallwey (2020) used a CNN method called U-Net, designed by Ronneberger et al. (2015) for the segmentation of medical images. It is an encoder-decoder network particularly powerful for problems with limited training data. A version of U-Net proposed by Gallwey (2020) classified multispectral Sentinel-2 images into mining, built/developed and vegetation land classes, achieving a mean class accuracy of 0.94. Camalan (2022) studied mining ponds as a separate component of ASM and created a custom dataset with 4 pond categories – background, active, transition and inactive. Actively mined ponds are brown in colour, due to high levels of sediment mixed with water. Abandoned inactive ponds are dark green, corresponding to dense climax algae growing. Lighter green ponds correspond to recent termination of mining and early stages of algae colonisation. Figure 1 shows how three pond types appear in RGB images.

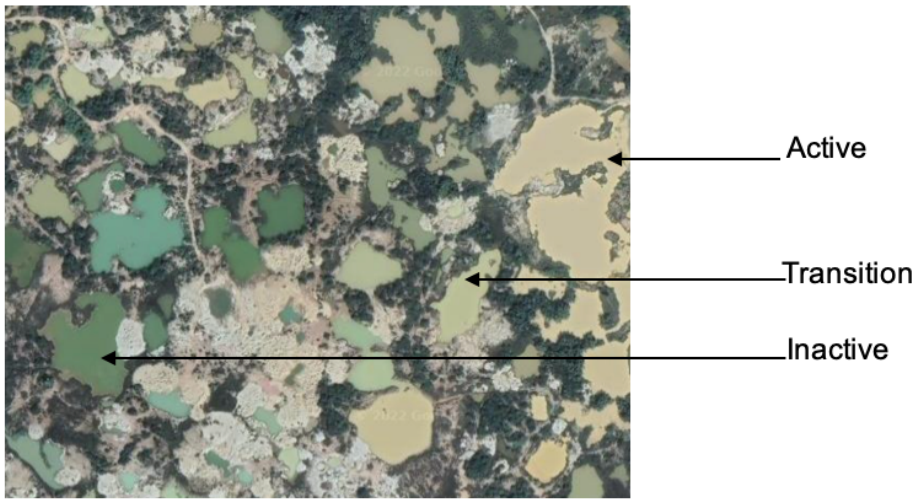


Figure 1: Mining Pond types in Central Ashanti, Ghana. Active ponds appear yellow or brown, due to mixing of sediment. Transition ponds are at early algae colonisation stages, hence light green in colour. Inactive ponds were abandoned 6 or more months ago and have developed substantial algae biomass, corresponding to dark green colour.

Camalan (2022) combined CNN and Recurrent Neural Network (RNN) in Re-CNN to detect internal temporal changes within mining sites. While the 'no change' and 'water existence' categories had very high F-1 scores – 0.99 and 0.96 respectively, classes 'increase' and 'decrease' had lower performance - 0.56 and 0.57, implying the model's weakness in determining types of ponds and their shapes. Proposed Re-CNN is a multifunctional network that in combination with an exclusive dataset resulted in the most fine-scale analysis of the internal dynamics of gold mining wetlands. However, while the combination of CNN and RNN has been used for feature extraction problems before (Jogin et al., 2018), U-Net remained a preferred image segmentation technique in remote sensing studies (McGlinchy et al., 2020, Jiao et al., 2020).

By bringing together advanced image segmentation method U-Net and Camalan's (2022) pond classification system this study aims to develop a tool for the detection and classification of gold mining ponds with the highest possible accuracy. A U-Net model will be developed without a temporal aspect, instead focus will be on the most precise and accurate identification of ponds and the distinction of their types. Camalan's (2022) main motivation behind developing a classification system was analysing intensification and cessation of mining activity over time. However, tool can have more use cases, such as monitoring the

changes in heavy metals concentrations, as the pond transitions from actively mined to abandoned. Mercury is key problematic heavy metal in gold mining, but tracking it remotely is an emerging field and is not currently possible using open-source satellite data (Ri et al., 2022). Conversely, minor ferric iron (MFI) can be estimated via the ratio of Red/Blue bands (Adamek et al., 2019). MFI is a product of weathered pyrite, exposing minerals with sulfur and if mixed with water and oxygen leads to acidification of water. The concentration of MFI is a proxy for the risk of acid mine drainage (AMD) (Akcil et al., 2006). Knowing how AMD risk changes as algae colonises abandoned ponds is useful for targeting the most polluted sites or assessing the effectiveness of restoration projects.

The objectives of this paper are i) to design a U-Net algorithm, capable of detecting mining ponds and distinguishing whether those are active, in transition or inactive, ii) to evaluate how the combination of input Sentinel-2 bands and the number of output classes influence the models' performance iii) propose applications of developed tool to quantify how the intensity of gold mining impacts the risk of AMD

2. Methodology

2.1 Dataset

The training data for pond classes was provided by Seda Camalan (2022). It consists of Sentinel-2 top of the atmosphere images for 16 regions from Madre de Dios, Peru, 1 region in Indonesia, 1 in Myanmar and 5 from Venezuela at 2 timesteps - August 2019 & 2021. Each image came in 3 versions of band combinations:

- 3 channels – Red, Green, Blue
- 6 channels – Red, Green, Blue, NIR, SWIR1, SWIR2,
- 10 channel - Red, Green, Blue, NIR, SWIR1, SWIR2, Ultra-Blue and 3 Red Edge bands.

To each image there was a corresponding 1-channel label image with 4 classes:

- Background
- Actively mined pond
- Transition pond (recent mining, but not current)
- Inactive pond (mining terminated 6 or more months ago)

The dataset was imbalanced with background accounting for 77% of data points, active 11%, transition 5% and inactive only 3%. Labels were created manually using the LabelBox (cite label box maybe) tool by experts in alluvial gold mining characterisation. Data from all four countries were randomly shuffled and used to train the model, to allow generalisations and usability on Ghana images (most comparable to Venezuela) and other international samples.

Application of U-Net classifier was presented on gold mining scenes in Ghana. Most ASM takes place in the South along the banks of the Ofin River. Median composite Sentinel-2 images for August 2020 were acquired from Google Earth Engine (GEE) for four areas: Ashanti, Hun, Dawusasso and Nkwanta. The tropical and humid climate of South Ghana

makes the region very cloudy, thus cloud masking was essential to prevent U-Net from mistaking clouds for highly reflective mines. The cloud classification data set created by Ibrahim et al. (2021) was imported into GEE and used to train Support Vector Machine algorithm. It was then applied to composite images to detect cloud pixels and mask them accordingly.

2.2 The design of model

U-Net was proved successful in the segmentation of macro-scale mining areas by Gallwey (2020), so the suitability of this algorithm for distinguishing ponds from other components within the mining site will be tested in this research. U-Net consists of encoder and decoder paths, where sequences of layers forms blocks (DownBlock and UpBLock). A number of these blocks are stacked on each other symmetrically, according to the specified depth. In the encoder path, a DownBlock consists of convolutions, ReLU activations and maxpooling operations, following the formula:

conv_layer1 -> conv_layer2 -> max_pooling

The number of filters (image depth) in the initial convolutional layer is set as hyperparameter during model initialisation, and each following convolutional layer doubles it. A maxpooling layer half's the image's size (height x width). At the last block the maxpooling layer is dropped. In the decoder path, the image is upsized to its original size following the formula:

conv_2d_transpose -> concatenate -> conv_layer1 -> conv_layer2

At transposed convolution the image is doubled in size and halved in depth. Then, it is concatenated with the corresponding image from the contractive path, as it allows gaining more precise information. The final layer of the network is a convolutional layer with 1x1 kernel, 1 filter and SoftMax activation function. SoftMax allows getting multi-class probability distribution making the output a 1 channel image with a user-defined number of classes (Wang et al, 2020).

In this project 9 U-Net models were built - on 3, 6 and 10 channel combinations of bands predicting 4 and 3 classes and binary classifications. Figure 2 illustrates the architecture of U-Net with 10 channel images predicting 4 classes.The key differences between the developed U-Net in this study and original proposed by Ronneberger et al., (2015) are the inclusion Batch Normalisation, switch to Adam optimiser from Stochastic Gradient Descend (SGD) and switch to Focal Loss from Cross-Entropy loss function.

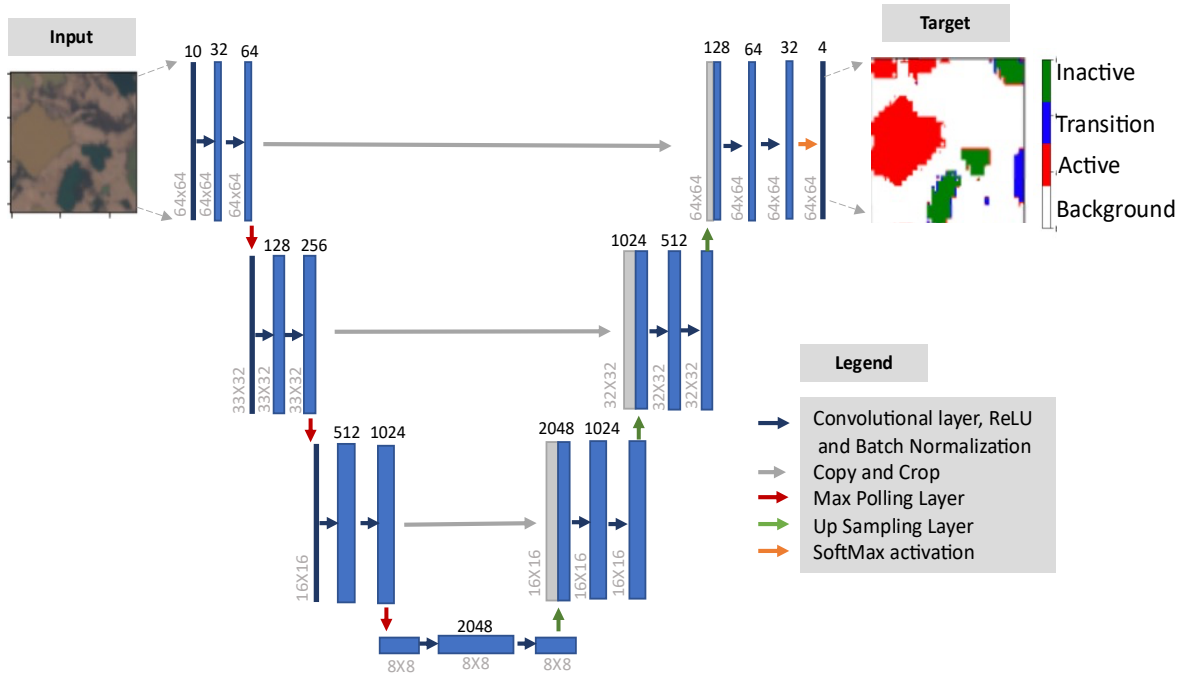


Figure 2: U-Net architecture with 3 blocks and 32 start filters. On the left of the encoder path there is an input image with 10 spectral bands (RGB image was used for visualisation purposes). On the right of the decoder path there is an output image with 4 classes.

Batch Normalisation standardises the inputs or activations of the prior layer without changing the structure of the network. It was proven that it can the required epochs by half (Liao et al., 2016), which is advantageous in this project where nine models are being trained and hyperparameter tuned. Therefore, the inclusion of this layer doubled the speed of training, allowing to test more hyperparameters, which is important for maximising the model's F-1 score.

Adaptive Moment Estimation (Adam) computes individually learning rates (the step size over the loss curve) and momentum (controls the steepness of the gradient) for each parameter. For highly imbalanced data in this project, this optimiser was most suitable, as it gives larger updates to rarely occurring instances (inactive and transitions ponds), than to frequently occurring ones (background class).

Focal Loss (FL) has proven to outperform other loss functions in image segmentation tasks with highly unbalanced data (Huang et al., 2020). It assigns more weights to samples that are difficult to classify and less to easily classified samples. A custom FL function was written, that could be used for binary and multiclass problems, as no in-built FL was available in PyTorch. Implementation of FL was key for achieving the adequate performance of the model. Replacing Cross Entropy Loss with FL was an important design decision, as with Cross Entropy model only detected background class and active ponds. Switching to FL as it allowed the model to detect all four classes and increased the F-1 score by 87%.

2.3 Training process

Predictions were done on all 4 classes, 3 classes (inactive and transition class combined in one) and binary (all pond types combined in a single class). By adding gradual complexity to models, it was understood whether models fail to detect ponds or distinguish between their types. Due to limited storage in Google Drive, first for each of 9 classifications, a model was

trained using standard hyperparameter combinations: start filters - 64; blocks - 4, (Ronneberger et al., 2015), gamma -2, alpha – 0.25 (Lin et al., 2017), learning rate – 0.01 and saved in the drive.

Then in a multilevel for loop, more hyperparameters were tested, but the model was only saved if it outperformed the initial one in terms of macro-averaged F1-score. Table 1 presents all hyperparameters that were tuned and explains their role in U-Net.

Table 1: U-Net hyperparameters tuned for the models in this research

Hyperparameter	Role
Start Filters	Number of filters in the first convolutional layer, setting the depth of image. It is doubled with each block
Blocks	The depth of the network itself.
Learning Rate	Step size at each iteration while moving toward a minimum of a loss function.
Gamma (Focal Loss hyperparameter)	When gamma = 0, the loss function is the same as Cross Entropy. As gamma increases, it downweighs the easily classified examples. The higher the gamma, the stronger is this downweighing.
Alpha (Focal Loss hyperparameter)	Alpha is adjusting the weights of classes. The higher the alpha the larger weights are given to rarely occurring classes and smaller to dominating class.

The values of hyperparameters were not chosen arbitrarily. Blocks had a maximum limit, as image size is halved with every DownBlock, so with 5 layers sample would reach a minimum dimension - 2x2 pixels. At such size, most details in the image are lost, particularly for Venezuela and Myanmar samples, where ponds are much smaller, than in Peru. Hence depths of 4, 3, and 2 blocks were tested to preserve variability at the model's bottleneck. High values of the start filter significantly increase training time without sufficient improvement of model performance (Ronneberger et al., 2015), thus 8, 16, and 32 were chosen. Alpha and gamma values of 0.25, 0.5, 0.75 and 2.0, 5.0, 8.0 were tested respectively. According to the original paper by Lin et al., (2017) a combination of alpha – 0.25 and gamma – 2.0 gives the lowest focal loss, but as the dataset was highly imbalanced, higher regularisation was tested.

2.4 Performance Metrics

U-Net models were evaluated using F1-score – harmonic average of Precision and Recall, Commission and Omission errors. Both overall F1 score and per class scores were computed, where 'macro' was used as the averaging method due to its suitability for imbalanced datasets, where all classes are equally important (Kumar, 2020). F1-score is defined as follows:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad \text{(Eq. 1)}$$

Precision shows the proportion of class predictions that belong to this class and Recall quantifies the proportion of class over all class examples in the dataset. Omission and commission errors were calculated for each class. Omission errors are instances that were left out from the correct class, while commission errors are samples assigned to a class where they do not belong. These errors highlight the model's mistakes and are popular metrics in remote sensing classification tasks (Bastarrika et al., 2011, Heiden et al., 2007).

2.5 Ecosystem and Workflow

Large GPU provided by Google Collaboratory (GC) makes training of deep-learning models fast and efficient (Bisong, 2019), all code was developed within it. Pro version of GC was used to access more powerful GPU (Nvidia T4) and up to 24Gb of RAM. GC connectivity to Google Drive allowed to store data, code, and models on cloud, making it safe and accessible from various machines. Figure 2 documents the workflow of this project.

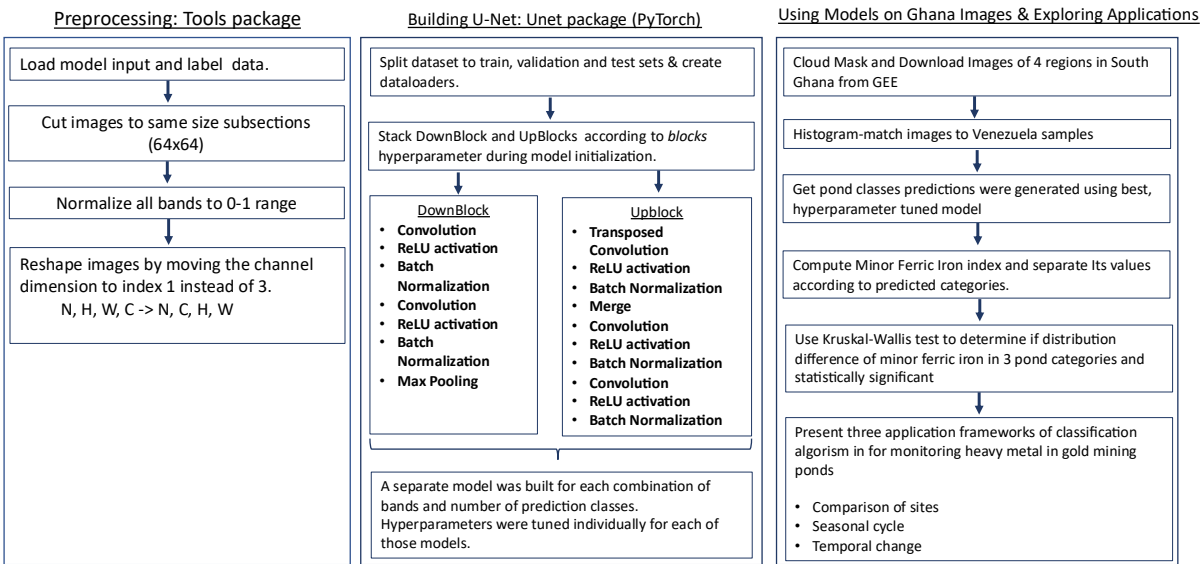


Figure 3: The pond classification workflow. Prepossessing steps are encapsulating in Tools package and model building and training in Unet package. A notebook in GitHub repository presents this workflow chronologically.

U-Net and some of the pre-processing methods were adapted from an open-source tutorial found in the repository https://github.com/johschmidt42/PyTorch-2D-3D-UNet-Tutorial (Schmidt, 2021).

Packages UNet and Tools store all functions and classes necessary for pre-processing, model building and training, visualisations, and exploration of applications in heavy metal tracking (Table 2). The notebook unifies the core methods in chronological order to showcase how models can be trained, tested, evaluated, and adapted for different research questions and data. The most important tasks include pre-processing of data for compatibility with U-Net, encapsulating layers in UpBlock and DownBlocks that could be stacked to form a U- Net and a multiclass Focal Loss function.

Table 2: Details of the custom packages used in this project.

Package Name	Description	Version	Language
Tools	Module includes packages preprocessor.py, ghana.py and visualizer.py. It focuses on data modification and cleaning for its compatibility with Unet, visualises results of models and utilises them to explore relationships of pond categories and ferric iron index concentrations.	1.0	Python 3
Unet	Implements custom PyTorch Unet Neural networks, makes custom dataset and trainer. Includes multiclass focal loss function.	1.0	Python 3
More details are provided in the README file in GitHub Repository of this project.			

2.6 Spatial and Temporal Trend Analysis

To present applications of the developed tool, Ghana images had to be histogram matched with one of the training images. Because the cumulative distribution function (CDF) of band values in Ghana is very different from those that model seen before, its performance significantly worsens in Ghana samples without this step. Histogram matching is an effective solution to calibrating CDF of an image to CDF of another image, to achieve the same range of band values as in training data (Kim et al., 2019). Experimentally it was determined that U-Net performs best when Ghana images are matched to Venezuela images.

To examine differences in MFI distribution in pond classes, the MFI index was computed via Red/Blue band ratio and overlayed with pond predictions derived using the best U-Net model. Then, values of MFI were conditionally extracted for each pond class and differences in their distributions were tested with the Kruskal-Wallis test. Kruskal-Wallis is a standard nonparametric test for comparing medians of 2 or more groups (McKight et al., 2010). Boxplots and line charts were used to illustrate how the change in mining intensity influences the risk of AMD. Mining intensity was determined by the proportion of active ponds versus transition and inactive ones and the risk of AMD was estimated from the skewness of MFI values distribution to higher or lower values. The code for the implementation of applications can be found in ghana.py module in and Applications section of the notebook.

2.7 Code metadata

Code was developed in Python 3.8 in the Google Collaboratory platform and can be accessed in the GitHub repository - https://github.com/ese-msc-2021/irp-ok12. The project does not rely on platforms outside of the workflow outlined in Figure 2 if the satellite data is available. Otherwise, the GEE interactive tool can be used to cloud mask Sentinel-2A images and downloads them to Google Drive. Machine Learning was implemented with PyTorch, including building UNet blocks, custom datasets, and training. As there was no in-built Focal Loss function in PyTorch, module focal.py in Unet package implements a version of it that can be used for binary as well as multiclass classification problems.

A full list of required packages to run the program are listed in requirments.txt file in GitHub repository. Details of functions are provided in docstrings, explaining functionality, arguments taken and outputs. The models built are stored in DropBox - https://www.dropbox.com/scl/fo/wzx82h2n84vtralzl2vrz/h?dl=0&rlkey=ursqm6gf9r52w7yszg7a8t77o The user is encouraged to run this classification workflow online within GC, especially if re-training models, as it will require high volume of RAM.

3 Results

F1-scores showed that, model's performance is directly proportional to number of channels used in input images, with 10-channel models having highest scores (Table 3). Binary classifications had better results, than multiclass for all band combinations detecting pond shapes almost exactly as per targets (Figure 4, A). Errors in multiclass models propagated from differentiation between transition and inactive ponds resulting in lower F1-scores and higher omission and commission errors (Table 4).
3 channel models had poorest results, indicating underfitting – common problem when there are not enough parameters in a model to explain data's complexity (Level et al., 2016).

Whilst 10-channel model had high errors in transition and inactive ponds, it was best suited to represent 4 classes accurately. Experiments to train models for more epochs were carried out, with intension to improve F-1 scores, but led to overfitting. Thus, models were possibly not deep enough, so larger image patches and more blocks should be tested for future development.

Table 3: Performance of 'best' models according to the F1-Scores on the test set, not used for model training. The datasets were split as 70%-20%-10% for the training, validation, and test sets, respectively

		Overall F1	Background	Active	Transition	Inactive
10 channel	4 classes	0.77	0.97	0.81	0.67	0.61
	3 classes	0.86	0.97	0.82	0.80	
	Binary	**0.94**	**0.97**		**0.91**	
6 channels	4 classes	0.67	0.96	0.77	0.50	0.56
	3 classes	0.80	0.96	0.69	0.74	
	Binary	**0.91**	**0.96**		**0.87**	
3 channels	4 classes	0.62	0.88	0.57	0.51	0.53
	3 classes	0.65	0.86	0.54	0.55	
	Binary	**0.73**	**0.84**		**0.61**	

Table 4: Commission and Omission Errors of best models produces. Errors are calculated per class of each model -3, 6 and 10 channels producing binary, 3 and 4 category classifications

			Background	Active	Transition	Inactive
10 channels	4 classes	Commission Error	3%	18%	30%	46%
		Omission Error	3%	21%	35%	28%
	3 classes	Commission Error	2%	19%	23%	
		Omission Error	3%	17%	16%	
	Binary	Commission Error	**3%**		**7%**	
		Omission Error	**2%**		**11%**	
6 channels	4 classes	Commission Error	2%	27%	58%	39%
		Omission Error	5%	18%	39%	48%
	3 classes	Commission Error	2%	33%	21%	
		Omission Error	5%	14%	26%	
	Binary	Commission Error	**4%**		**12%**	
		Omission Error	**4%**		**14%**	
3 channels	4 classes	Commission Error	15%	29%	57%	47%
		Omission Error	9%	51%	37%	46%
	3 classes	Commission Error	18%	32%	43%	
		Omission Error	9%	55%	61%	
	Binary	Commission Error	**26%**		**23%**	
		Omission Error	**7%**		**51%**	

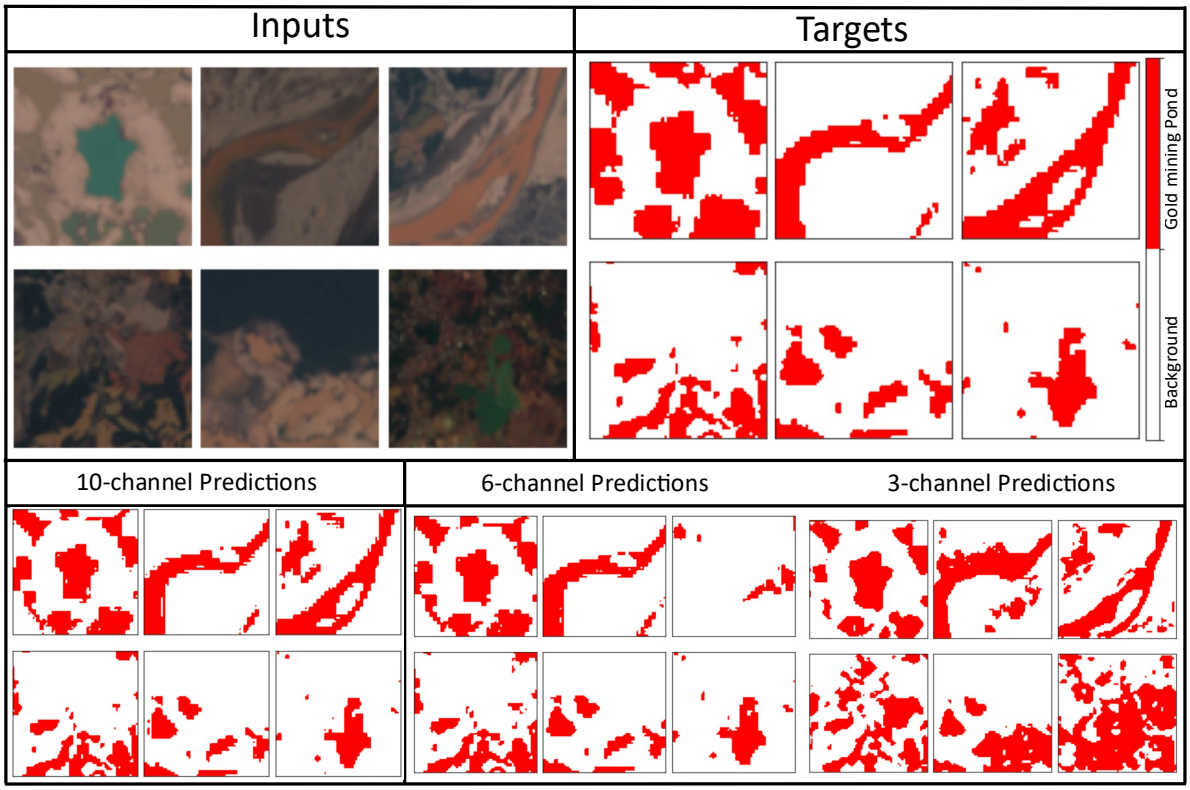
All models had instances of outperforming labelling mistakes. Figure 4, B top middle image (selected with black dashed square) was labelled as active, but was light green colour, more like transition class with some dark green (inactive) patches. The corresponding target label was not only red (active) but also selected the whole region, while there was evidently a complex micro-network of light green, elongated ponds. 10-channel model distinguished inactive ponds, but still labelled those of transition as active (as per target). The 6-channel model better captured the variability within this micro-network, while the 3-channel one labelled all sections as inactive, most like ground truth, given the light green colour.

Similar inconsistencies can be found in 4-class models. The selected RGB image has brown-coloured pond, but the target is the transition. 10 and 6-channel models predicted it as active, but 3-channel as inactive (more like target). Thus, the source of some errors in Table 4 could be the data itself, not only models, but it is not possible to quantify it.

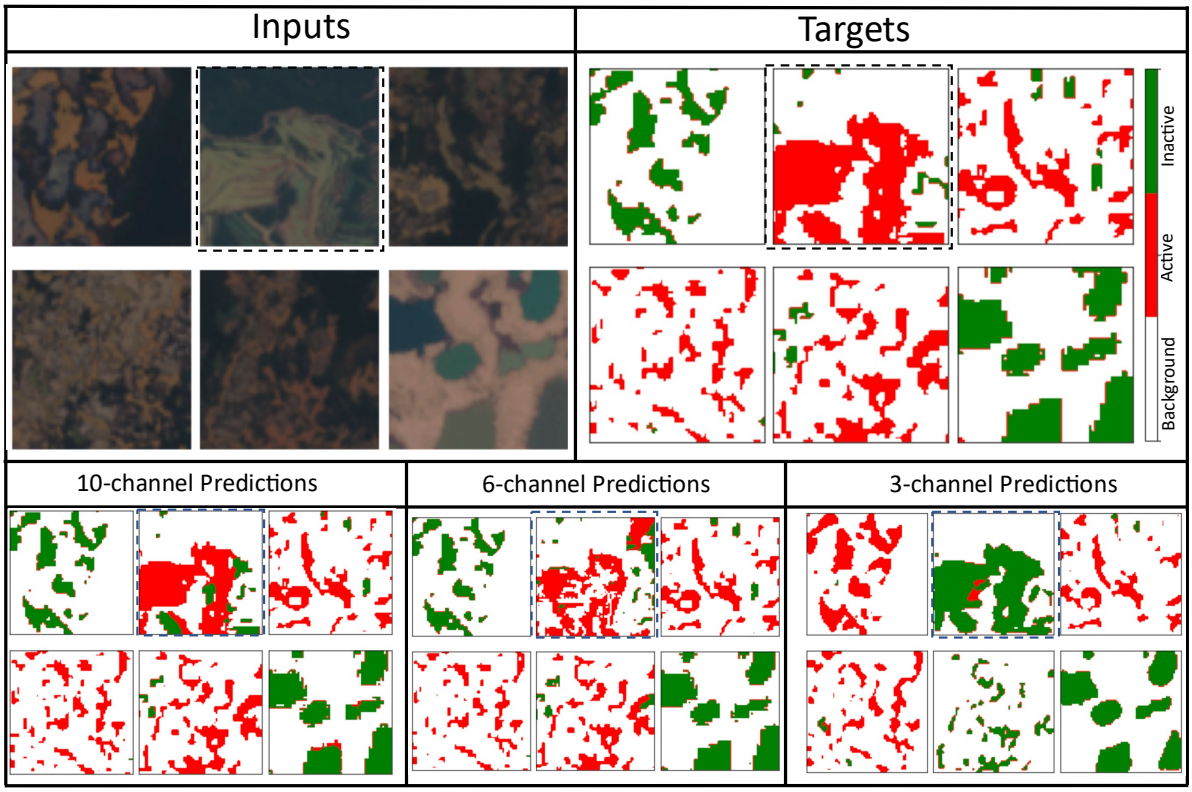
An exact comparison of developed models with Camalan's (2022) Re-CNN is not possible, as here ponds were identified and classified at a single timestep, while Camalan detected temporal changes. While the classes 'increase' and 'decrease' largely correspond to the ability of Re-CNN to detect ponds' shapes and types and relate to current 'active', 'transition', 'inactive', such comparison is not direct. Nevertheless, the best U-Net models developed in this research have F1-scores of 0.97, 0.81, 0.67 and 0.61 for background active, transition and inactive classes respectively, which could be considered as improvement from 'decrease' (0.56) and 'increase' (0.57) in Re-CNN by Calman (2022).

P-value from Kruskal-Wallis test determined that the difference of MFI distribution in four classes is statistically significant at 99% confidence level (p value = 4.82e-19), proving that risk of AMD is higher in active ponds. In active ponds most common value of MFI was 1.5, while for transition and inactive ponds it was 0.5, which is 3 times lower, indicating that sites where mining activity ceased have a lower risk of AMD. Three examples of how these findings can be used for monitoring mining activity and quantification of AMD risk will be discussed in the next section.

A



B



C

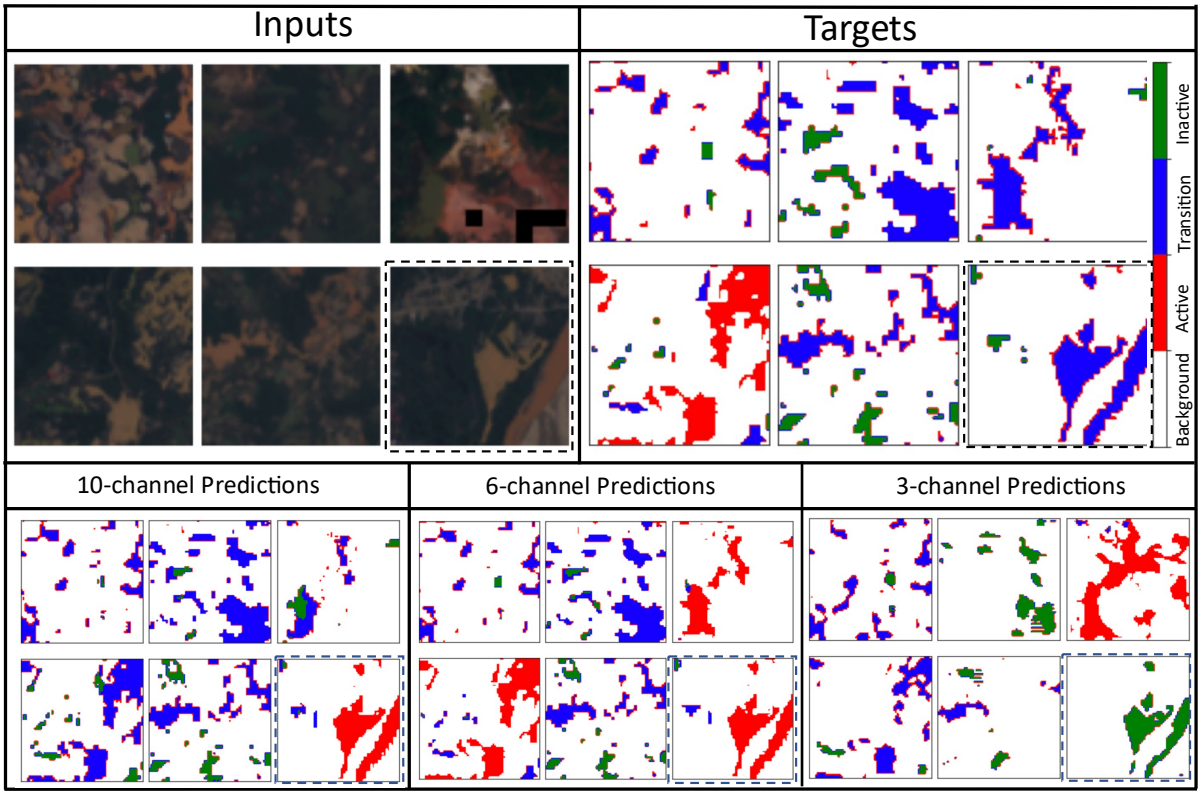


Figure 4: Visualisation of models' predictions, inputs, and targets on 6 samples of 64x64 pixels from test set. A- Binary classification; B - Three-class classification; C – Four-class classification.

4 Discussion

4.1 Influence of spectral bands on model performance

Models with 10 bands performed best in all classifications, indicating that complexity of scenes needed more parameters to explain data variance. The characteristics of gold mining wetlands vary around the globe – in Madre de Dios, Peru gold mining takes place within Amazon Forest. Hence most scenes used for model training depicted simple landscapes where only ponds, bare land and dense vegetation were captured land types. In Myanmar, Indonesia and Venezuela gold mining is more complex, including urban areas, roads, vegetation, and various soils. Thus, to segment ponds in these complicated scenes more parameters were required.

Waterbodies (including mining ponds) are least reflective at visible red and NIR bands, hence they are widely used in water detection tasks. Conversely, the spectral difference between water from non-water is smallest at short wave infrared bands (SWIR1 & SWIR2) (Sepulcre-Cantó et al., 2006). As SWIR1 and SWIR2 were included in both 6 and 10 band combinations, it could be the source of errors. For future refinement of this project tests with the exclusion of short-wave infrared bands could be run to analyse how they affect the performance.

Algal blooms, largely determining the class of ponds, are related to the presence of chlorophyll-a in waterbodies. A normalised difference chlorophyll index (NDCI) developed by Mishra et al., (2012) could have been included as a band to enhance the importance of algae colonisation in pond categorisation. However, NDCI is calculated as a normalised ratio of vegetation reg edge 1 and visible red bands in Sentinel 2 (Delegido et al., 2011), that were already included in 6 and 10 channel images. Adding those bands individually and as an index could lead to overfitting.

Increasing the number of input bands significantly improved the model's performance. The shapefiles of labelled regions were not available, so it was not possible to re-download scenes and use all 13 spectral bands of Sentinel-2 to potentially reach maximum F1-scores. Alternatively, building a binary model for each class for the same band combinations could improve the result, which has been a proven effective strategy for classification problems with unbalanced data (Ulmas and Liiv, 2020).

4.2 Spatial & Temporal applications of U-Net for monitoring of minor ferric iron

We speculate that beyond the tracking of intensification and cessation of mining activity, developed models can be used for the estimation of AMD risks. Kruskal-Wallis test proved that the concentration of MFI decreases when algae colonises abandoned ponds. Hence, for planning and evaluating environmental restoration projects it is important to quantify how much AMD risk decrease is due to the suspension of mining activity and how much due to external efforts. Moreover, developed tools can help to prioritise sites requiring environmental protection and suspension of mining or highlight small-scale mining, with efficient practices, where mining activity is in progress, but the risk of AMD is low. Three applications for such investigations are proposed:

1) Comparison of sites (e.g., with different environmental restoration strategies).
2) The seasonal cycle of mining activities and ferric iron fluctuation (to indicate time of highest and lowest risk of AMD)
3) Monitoring of temporal changes in pond type proportions vs MFI concentration.

Most mining in South Ghana takes place along the banks of the Ofin river, thus four sites were selected from that area to showcase application 1. 10-channel U-Net model was used to compute pond predictions (Figure 5). Applications 2 and 3 were computed using Planet.com basemaps, as cloud free images were unavailable for most of the months from Sentinel-2. Only red, green, and blue bands were common for Planet and Sentinel-2, so 3-channel U-Net had to be used. It is acknowledged that 3-channel models resulted in a high error rate, especially in inactive and transition classes, but accuracy had to be traded off for data availability.

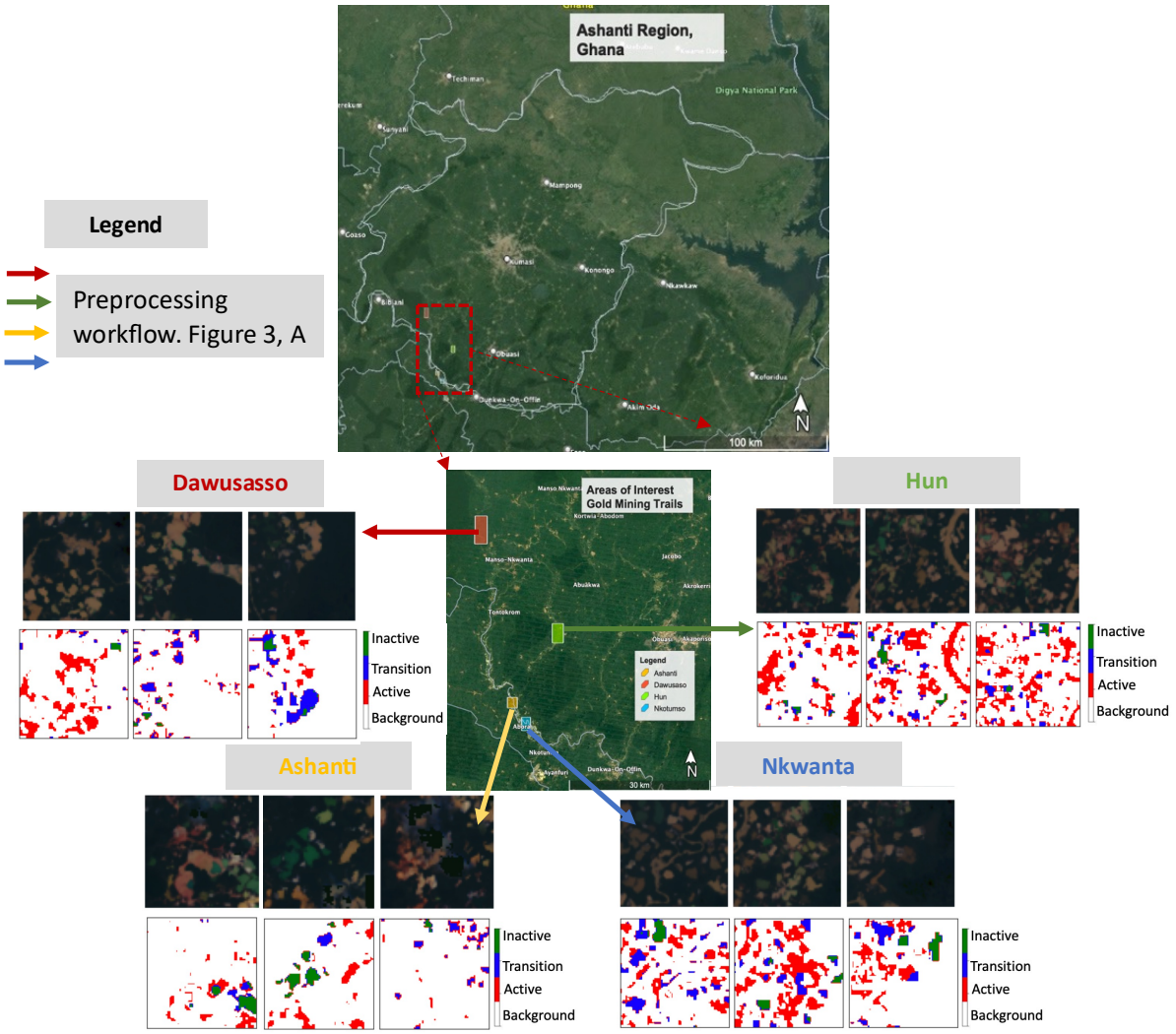


Figure 5: Locations of Sentinel-2 images of Ghana downloaded from GEE. For each image a 10-channel model predicting four classes (background, active, transition and inactive) was applied. Input and output pairs are demonstrated for each of the selected regions.

Figure 6, A presents application 1, comparing Ashanti, Hun, Dawusasso and Nkwanta sites for the level of mining activity and corresponding MFI concentration. Ashanti region had the highest proportion of active ponds (88%), and the distribution of MFI was skewed to larger values (Figure 6, A). Dawusaso had most ponds in transition or inactive state, hence lower mean of MFI and highest interquartile range, as more ponds have low values of MFI concentration. Thus, Ashanti was at the most risk of AMD and should be a priority site for environmental conservation and protection projects. These 4 regions were chosen arbitrarily, but the presented framework can be used for comparison of restoration policies success.

Figure 6, B illustrates seasonal cycle of mining in Aniamoa in 2020. Despite no unavailability of data for July and September (almost full coverage with clouds) it is still evident that mining intensity rises during colder winter month - November to March, as the proportion of active ponds is largest. In spring some ponds are abandoned and go in transition phase, becoming inactive by June. Correspondingly, mean MFI lowers from 1.7 January -March to 1.3 in June (~24% reduction). Such lag in lowering of MFI concentration between initial mining activity slowdown in March to lowest mean value in June, implies that site stays acidic even after mining activity has cessed. With mining accelerating back in August the site is constantly exposed to pollutants. This cycle is alarming as rebuilding the environment to previous state in such areas can be very difficult.

Figure 6, C shows framework 3 – temporal dynamics of Aniamoa region during peak mining season – January-March from 2018 to 2022. This site is not under any conservation program, however there is a clear downtrend in active mining ponds from 2018-2021. Distribution of MFI shifts to lower values in 2021, but the process is slow with only 10% decrease in mean. The site is likely getting re-visited for mining in 2022, as proportion of active ponds rises again, as does MFI concentration. When abandoned ponds are revisited, miners who do so expose themselves to high risk of malaria infection. Paintsil et al., (2019) reported that residents of rural settlements in Ashanti region had been most effected by malaria and had lowest access level to healthcare. By regulating and restoring the abandoned artificial wetlands, both AMD and social health risks can be reduced.

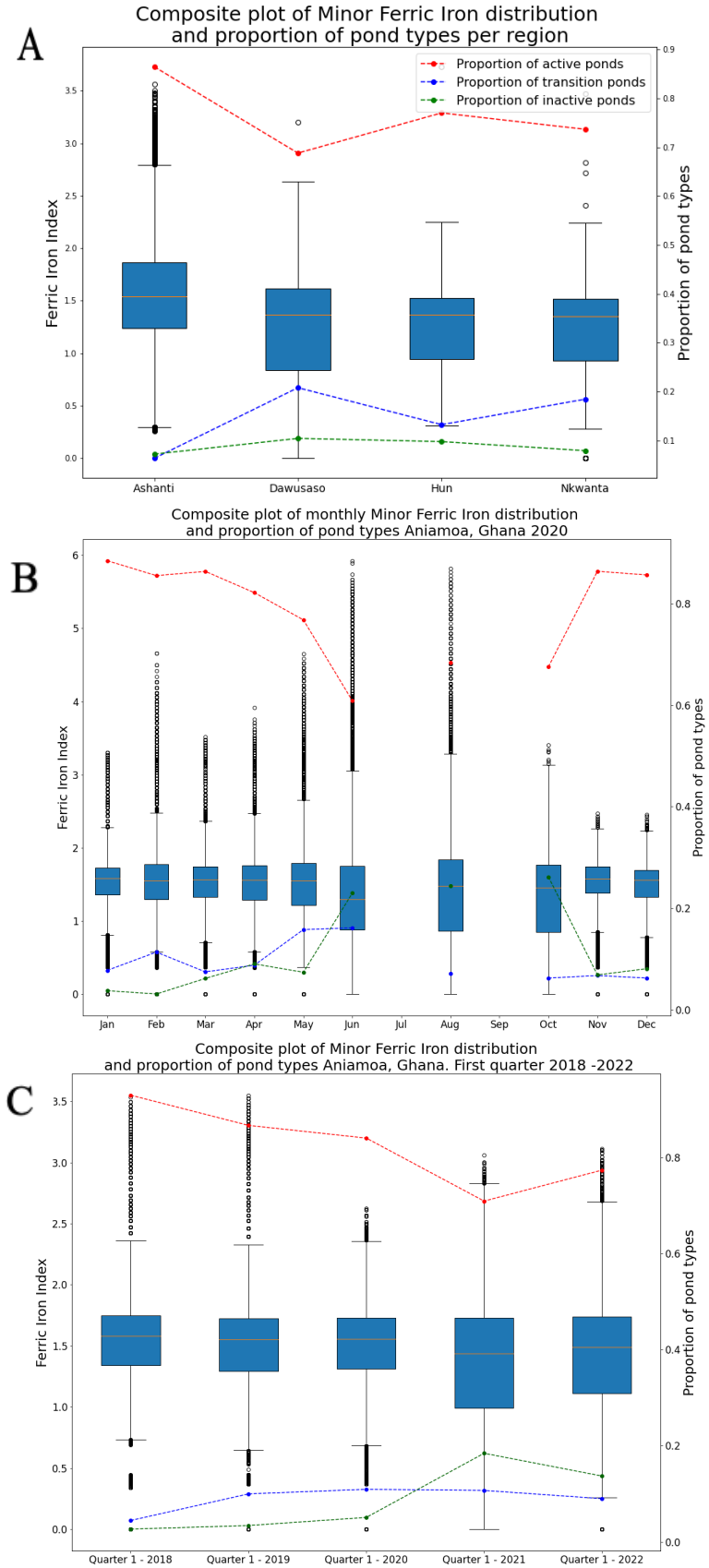


Figure 6: Composite plots of active, transition and inactive ponds' proportions and MFI distributions. A – site comparison framework showing varying intensity of mining and MFI concentrations in 4 regions; B – investigation of seasonal mining cycles and response of MFI distributions; C – timeline mining intensity and MFI concentration at first quarter (January-March) of 2018-2022.

Conclusion

By applying the novel image-segmentation algorithm U-Net on a unique dataset created by Camalan (2022) it was possible to identify different types of mining ponds – active, transition and inactive. The best models were those using 10-input channels demonstrating high capability in identifying ponds and distinguishing their class. While F-1 scores for transition and inactive classes were lower (0.67- 0.67), the model's results significantly improved to 0.80 when those were combined in a single class (inactive). For studies where the level of colonisation of algae is less important it is recommended to use 3 class model.

Models can be used to identify the cessation or intensification of mining activity to track the effectiveness of environmental restoration strategies. This study presents three such applications by uncovering how MFI concentration changes during the transition of ponds from active to inactive. The developed tool can be used in conjunction with U-Net by Gallwey (2020), which segments the whole mining site to understand a proportion of land used for gold mining out of the total area degraded and deforested. The spatial distribution, placement and size of ponds or lower proportion of ponds versus total site area can all be helpful signs for distinguishing legal sites versus illegal (Labbe, 2021).

Using the developed tool, made it is possible to generate accessible, accurate and free data on mining sites composition and proposed applications can be helpful for the development of sustainable mining practices. Cloud cover is a primary restriction for such research in South Ghana, however 3-channel models can be used on Planel.com basemaps, that provide more frequent cloud-free composite images. As the performance of 3-channel models is weaker than 6 and 10-channel, retraining models on original Planet's bands is a recommended extension of this project.

References

Akcil, A. and Koldas, S., (2006). Acid Mine Drainage (AMD): causes, treatment and case studies. Journal of cleaner production, 14(12-13), pp.1139-1145. doi:10.1016/j.jclepro.2004.09.006

Bastarrika, A., Chuvieco, E. and Martín, M.P., (2011). Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: Balancing omission and commission errors. Remote sensing of Environment, 115(4), pp.1003-1012. doi:10.1016/j.rse.2010.12.005

Barenblitt, A., Payton, A., Lagomasino, D., Fatoyinbo, L., Asare, K., Aidoo, K., Pigott, H., Som, C.K., Smeets, L., Seidu, O. and Wood, D., (2021). The large footprint of small-scale artisanal gold mining in Ghana. Science of the Total Environment, 781, p.146644. doi:10.1016/j.scitotenv.2021.146644

Bisong, E., (2019). Building machine learning and deep learning models on Google cloud platform (pp. 59-64). Berkeley, CA: Apress.

Brownlee J., (2017) Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. Deep Learning Performance [online] Available at: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/#:~:text=Adam%20is%20a%20replacement%20optimization,sparse%20gradients%20on%20noisy%20problems. Accessed on 10.08.2022

Cai, J., Chen, T., (2018). Recent advances in convolutional neural networks. Pattern Recogn. 77, 354–377. doi:10.1016/j.patcog.2017.10.013

Cayuela, L., Golicher, J.D., Rey, J.S. and Benayas, J.R., (2006). Classification of a complex landscape using Dempster–Shafer theory of evidence. International Journal of Remote Sensing, 27(10), pp.1951-1971. doi:10.1080/01431160500181788

Delegido, J., Verrelst, J., Alonso, L. and Moreno, J., (2011). Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors, 11(7), pp.7063-7081. doi:10.3390/s110707063

Driscoll, C. T. , Mason, R. P., Chan, H. M., Jacob, D. J., Pirrone, N., (2013) Mercury as a global pollutant: Sources, pathways, and effects. Environ. Sci. Technol. 47, 4967–4983, doi:10.1021/es305071v

Fisher, R.P., Hobgen, S.E., Haleberek, K., Sula, N. and Mandaya, I., (2018). Free satellite imagery and digital elevation model analyses enabling natural resource management in the developing world: Case studies from Eastern Indonesia. Singapore Journal of Tropical Geography, 39(1), pp.45-61. doi:10.1111/sjtg.12210

Fritz, M., McQuilken, J., Collins, N., Weldegiorgis, F., (2017). 'Global trends in artisanal and small-scale mining (ASM): A review of key numbers and issues.' In: Intergovernmental Forum on Mining, Minerals, Metals and Sustainable Development (IGF). Winnipeg: IISD.

Goodfellow, I., Bengio, Y., Courville, A., (2016). Deep Learning. MIT Press, Cambridge MA.

Heiden, U., Segl, K., Roessner, S. and Kaufmann, H., (2007). Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data. Remote Sensing of Environment, 111(4), pp.537-552. doi:10.1016/j.rse.2007.04.008

Hilson, G., Maconachie, R., (2019). Artisanal and small-scale mining and the sustainable development goals: opportunities and new directions for sub-Saharan Africa. Geoforum 1–17. doi:10.1016/j.geoforum.2019.09.006

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W. and Wu, J., (2020), May. Unet 3+: A full-scale connected unet for medical image segmentation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1055-1059). IEEE. doi:10.1109/ICASSP40776.2020.9053405

Jiao, L., Huo, L., Hu, C. and Tang, P., (2020). Refined Unet: Unet-based refinement network for cloud and shadow precise segmentation. Remote Sensing, 12(12), p.2001. doi:10.3390/rs12122001

Jogin, M., Madhulika, M.S., Divya, G.D., Meghana, R.K. and Apoorva, S., (2018), May. Feature extraction using convolution neural networks (CNN) and deep learning. In 2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT) (pp. 2319-2323). IEEE. doi:10.1109/RTEICT42901.2018.9012507

Kim, M., Jung, M. and Kim, Y., (2019). Histogram matching of Sentinel-2 spectral information to enhance Planetscope imagery for effective wildfire damage assessment. Korean Journal of Remote Sensing, 35(4), pp.517-534. doi:10.7780/kjrs.2019.35.4.3

Kotaridis, I. and Lazaridou, M., (2021). Remote sensing image segmentation advances: A meta-analysis. ISPRS Journal of Photogrammetry and Remote Sensing, 173, pp.309-322. doi:10.1016/j.isprsjprs.2021.01.020

Kumar A. (2020), Micro-average & Macro-average Scoring Metrics – Python. Data Analytics [online], Available at: https://vitalflux.com/micro-average-macro-average-scoring-metrics-multi-class-classification-python/ Accessed on: 20.08.2022

LeCun, Y., Boser, B., Denker, J., Henderson, D., (1990). Handwritten digit recognition with a back-propagation network. In: Adv. Neural Inf. Process. Syst. 2. D. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 396–404.

Lever, J., Krzywinski, M. and Altman, N., (2016). Points of significance: model selection and overfitting. Nature methods, 13(9), pp.703-705.

Liao, Z. and Carneiro, G., (2016), March. On the importance of normalisation layers in deep learning with piecewise linear activation units. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1-8). IEEE. doi:10.1109/WACV.2016.7477624

Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P., (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

Lupa, M & Lesniak, A (2015), 'New prospects for automated geological mapping with Sentinel-2 multispectral instrument' in 16th Czech - Polish workshop ON RECENT GEODYNAMICS OF THE SUDETY MTS. AND ADJACENT AREAS.

McGlinchy, J., Johnson, B., Muller, B., Joseph, M. and Diaz, J., (2019), July. Application of Unet fully convolutional neural network to impervious surface segmentation in urban environment from high resolution satellite imagery. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 3915-3918). IEEE. doi:10.1109/IGARSS.2019.8900453

McKight, P.E. and Najab, J., (2010). Kruskal-wallis test. The corsini encyclopedia of psychology, pp.1-1. doi:10.1002/9780470479216.corpsy0491

McQuilken, J., & Hilson, G. (2016). Artisanal and small-scale gold mining in Ghana. Evidence to inform an 'action dialogue'. IIED.

Miyoshi, G.T., Arruda, M.D.S., Osco, L.P., Marcato Junior, J., Gonçalves, D.N., Imai, N.N., Tommaselli, A.M.G., Honkavaara, E. and Gonçalves, W.N., (2020). A novel deep learning method to identify single tree species in UAV-based hyperspectral images. Remote Sensing, 12(8), p.1294. doi:10.3390/rs12081294

Mishra, S. and Mishra, D.R., (2012). Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sensing of Environment, 117, pp.394-406. doi:10.1016/j.rse.2011.10.016

Nicholls, E., Ely, A., Birkin, L., Basu, P., & Goulson, D. (2020). The contribution of small-scale food produc- tion in urban areas to the sustainable development goals: A review and case study. Sustainability Science, 15(6), 1585–1599. doi:10.1007/s11625-020-00792-z

Paintsil, E.K., Omari-Sasu, A.Y., Addo, M.G. and Boateng, M.A., 2019. Analysis of haematological parameters as predictors of malaria infection using a logistic regression model: A case study of a hospital in the Ashanti Region of Ghana. Malaria Research and Treatment, 2019. doi:10.1155/2019/1486370

Ri, W.A.N.G., Xiong, C.Y., Yong, X.I.E., Ming-Jie, H.A.N., Yu-Hao, X.U., Chao, B.I.A.N. and Shan-Hong, X.I.A., 2022. Electrochemical sensor based on MoS2 nanosheets and DNA hybridization for trace mercury detection. Chinese Journal of Analytical Chemistry, 50(3), p.100066. doi:10.1016/j.cjac.2022.100066

Sasu, D. D., (2021) Annual contributions of gold to GDP in Ghana 2013-2020 [online] Avaiable at: https://www.statista.com/statistics/1272300/annual-contributions-of-gold-to-gdp-in-ghana/ Accessed on 20.06.2022

Sepulcre-Cantó, G., Zarco-Tejada, P.J., Jiménez-Muñoz, J.C., Sobrino, J.A., De Miguel, E. and Villalobos, F.J., 2006. Detection of water stress in an olive orchard with thermal remote sensing imagery. Agricultural and Forest meteorology, 136(1-2), pp.31-44. doi:10.1016/j.agrformet.2006.01.008

Signoroni, A., Savardi, M., Baronio, A. and Benini, S., 2019. Deep learning meets hyperspectral image analysis: A multidisciplinary review. Journal of Imaging, 5(5), p.52.

Varma P., (2020) Hands-On Guide To Implement Batch Normalization in Deep Learning Models, DEVELOPERS CORNER [online], Available at https://analyticsindiamag.com/hands-on-guide-to-implement-batch-normalization-in-deep-learning-models/#:~:text=Batch%20Normalization%20is%20also%20a,of%20overfitting%20and%20works%20well, Accessed on 20.08.2022.

Wang, S., Chen, W., Xie, S. M., Azzari, G., & Lobell, D. B. (2020). Weakly supervised deep learning for segmentation of remote sensing imagery. Remote Sensing, 12(2), 1–25. doi:10.3390/rs12020207

Yan, C., Fan, X., Fan, J. and Wang, N., 2022. Improved U-Net Remote Sensing Classification Algorithm Based on Multi-Feature Fusion Perception. Remote Sensing, 14(5), p.1118. doi:10.3390/rs14051118