

A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: applying deep learning to shallow mining

Jane Gallwey^{1*}, Carlo Robiati¹, John Coggan¹, Declan Vogt¹ and Matthew Eyre¹

¹ Camborne School of Mines, University of Exeter, Tremough Campus, Penryn TR10 9FE, United Kingdom;
j.gallwey@exeter.ac.uk, c.robiati@exeter.ac.uk, j.coggan@exeter.ac.uk, d.vogt@exeter.ac.uk, m.eyre@exeter.ac.uk

* Correspondence: j.gallwey@exeter.ac.uk;

Abstract: Artisanal Small-scale Mining (ASM) is a critical source of livelihoods for large areas of the Global South but it can bring with it many problems, including deforestation, water pollution and low worker safety. Timely and comprehensive management of ASM is crucial to ensure that it can take place safely and cleanly, supporting sustainable development. The informal nature of the sector presents challenges related to documenting the locations of ASM. Remote sensing methods have been used to detect ASM, although difficulties with accuracy, resolution and persistent cloud cover have been encountered. This paper proposes a method of ASM detection using a deep convolutional neural network model applied to open source Sentinel-2 multispectral satellite imagery. Firstly, the model is evaluated against both existing ASM detection methods and visual inspection of randomly sampled points. Secondly, the model is used to map mining and urban land use changes over a dataset spanning four years and 6 million hectares of southern Ghana, demonstrating the ability of this method to process very large areas.

The omission and commission errors of less than 8% from the sampled points indicate that this model has achieved unprecedented levels of accuracy for the task of detecting ASM from satellite imagery. When applied to the case study area, the data on ASM trends over time demonstrate a correlation between the Ghanaian government's 2017 clampdown and ASM activities. The ASM land use category decreased by 6,000 hectares in 2017, despite a net increase of 15,000 hectares over the period 2015-2019. Additionally, the model was applied to quantify the extent of illegal mining related deforestation within Ghana's protected forests, measured at over 3,500 hectares, with 2,400 of these lost since 2015. The results demonstrate that this methodology can detect ASM in Ghana with a high degree of accuracy at a minimal cost in terms of financial and human resources. The model shows strong generalisation abilities, offering exciting potential for using this methodology to further monitor and analyse ASM related land use changes worldwide.

Keywords: Sentinel-2; deep learning; small-scale mining; sustainable development; deforestation

1 Introduction

Artisanal and small-scale mining (ASM) is a rapidly expanding source of livelihood for many in the Global South, particularly in rural areas. It is estimated that over 40 million people are directly working in ASM across 80 countries, with a further 150 million people dependent on ASM indirectly (Fritz et al., 2017). ASM activities can be considered to exist on a spectrum of formality (McQuilken and Garvin, 2016), from highly illegal mining either within nature preserves (Boadi et al., 2016) to fully licensed, environmentally compliant and formalised small-scale mining (Hilson, 2002).

Whilst ASM generates vital socioeconomic benefits for communities and countries it is also associated with environmental and social problems such as land degradation, deforestation, water pollution, illegal immigration, low worker safety and child labour (Corbett et al., 2017, Fritz et al., 2017, Hilson, 2002). The excavation of ground for mining causes widespread deforestation and leaves dangerous unstable pits that fill with standing water, creating breeding grounds for malarial mosquitos (Bansah et al., 2018). The unregulated mineral processing also leads to heavy metal pollution, especially increased mercury levels (Telmer and Stapper, 2007, Bansah et al., 2018).

Appropriate management of ASM activities is critical, with its benefits directly linked to 9 of the 17 UN Sustainable Development Goals (Hilson and Maconachie, 2019). However, effective ASM management is inhibited by factors such as the informal and undocumented nature of the sector, legacy of inappropriate policies, limited government resources and the remote locations of mine sites (Corbett et al., 2017, Hilson and Gatsinzi, 2014). Detailed, accurate and inexpensive geoinformation about ASM activities would aid legislative pathways by providing rapid mapping resources to support small scale licensing claims, one of the major barriers to effective legislation (McQuilken and Garvin, 2016). These datasets would also support the timely tackling of environmental problems by focusing enforcement and remediation efforts where they are most needed.

Remote sensing data sourced from earth observation satellites could generate this geoinformation, although in the past their applicability for mapping ASM activity has been hindered by the resolution of the sensors and the accuracy of the classification techniques (Asner et al., 2013). Landsat and MODIS, the most prominent satellite imagers of the 1990s and 2000s, have resolutions of 30m and 250m respectively, making identification of the often small ASM alluvial mine workings problematic as workings can cover less than half a pixel, even at Landsat resolution. ASM activity visible from space is primarily surface mining, usually alluvial in nature and presenting as ribbons of clustered pits (Snapir et al., 2017). They are normally located on or near watercourses, along with dry pits and bare earth waste piles (Owusu-Nimo et al., 2018), ranging in size from less than half a hectare to several hundred hectares for the larger clusters.

The past five years have seen the launch of advanced satellite imaging systems with greatly increased spatial resolutions. These include the European Space Agency's (ESA) Sentinel-2 platform (Berger et al., 2012) which provides multispectral imagery with 10m resolution in four visible and near infrared bands and 20m resolution in four further infrared bands (Drusch et al., 2012, European Space Agency, 2015). Spectrally, the alluvial workings are highly reflective in the near infrared bands between 700nm and 875nm and display the greatest contrast from vegetation, rooftops and open ground in the short-wave infrared bands between 1600nm and 2500nm.

Delineating ASM related deforestation is a subset of Land Use/Land Cover (LULC) mapping. The most common approach to LULC mapping performs pixel-based classification based on the spectral

signatures of the classes of interest, utilising machine learning classifiers. Pixel based classifiers have been used for mapping ASM activity in Ghana (Boakye et al., 2019, Kusimi, 2008, Obodai et al., 2019), Burkina Faso (Leroux et al., 2018) and Brazil (Lobo et al., 2018). The reported omission/commission errors range from 8-40% for the mining class, indicating a large variability in the accuracy of this method (Boakye et al., 2019, Obodai et al., 2019).

A subset of pixel-based methods, used primarily in deforestation studies, leverages spectral mixture analysis to detect sub pixel changes. Multiple studies have used this technique to detect ASM related deforestation (Asner et al., 2013, Asner and Tupayachi, 2017, Espejo et al., 2018). Typical omission and commission errors from these methods are in the region of 10-25% (Asner and Tupayachi, 2017, Espejo et al., 2018), with some difficulties encountered in correctly classifying the water pools within the alluvial mine sites (Espejo et al., 2018). The greatest weakness of the pixel-based classifiers is their lack of spatial context. Each pixel is considered individually, leading to a 'speckled' effect in the resulting classification maps (Blaschke et al., 2014) which can be mitigated to some degree by postprocessing but not wholly removed (Kelly et al., 2011).

Object based approaches attempt to address this lack of spatial context by first segmenting the image into regions based on spectral similarity then classifying the regions rather than the individual pixels (Blaschke et al., 2014). These approaches have seen relatively little application for detecting ASM, but a successful implementation for general rainforest LULC change analysis (including mining) is found in Souza-Filho et al. (2018) with commission/omission errors in the range of 10-30%. However, finding adequate parameters for the initial object segmentation can be labour intensive and relies on good domain knowledge (Nuijten et al., 2019). All methods to date have required a significant level of human input, including feature and segmentation parameter design, manual cluster selection and manual relabelling of incorrect pixels.

Outside the field of traditional remote sensing, techniques from computer science could provide a solution: since 2010, advances in the available processing power of Graphics Processing Units (GPUs) have allowed a type of deep artificial neural network known as a Convolutional Neural Network (CNN) (LeCun et al., 1990) to emerge dominant for most image processing tasks (Gu et al., 2018). CNNs are inspired by biological visual cortexes and work by adaptively and automatically learning spatial dependencies and hierarchies of features from gridded data. They can approximate highly nonlinear functions whilst maintaining spatial connectivity between pixels (Goodfellow et al., 2016).

Considerable research has been carried out in the broader remote sensing community as to how to modify CNNs for LULC tasks; a review is given in Ball et al. (2017). The main identified challenges to adoption were found to be the limited availability of large amounts of pre-labelled training data and the multiple channels found in multispectral and hyperspectral imaging systems (Nogueira et al., 2017, Signoroni et al., 2019). Deep learning based LULC applications have been successfully implemented

using Sentinel-2 data, but the images are most commonly clipped to contain only the first three or four bands, as seen in Kroupi et al. (2019) and Wurm et al. (2019), to maintain compatibility with non-multispectral deep learning models. Very few studies have modified deep learning models to ingest true multispectral data, with Kemker et al. (2018) describing the most comprehensive and promising study to date. No published work has described the application of true multispectral deep learning for classifying ASM and deforestation.

LULC mapping allows delineation of ASM activity, but to extend from delineation to monitoring, the model must include a temporal change detection element. Post-classification change detection is the method used by most ASM mapping studies to date (Asner and Tupayachi, 2017, Boakye et al., 2019, Kusimi, 2008), but it is sensitive to the errors in the original classification maps and can produce imprecise results especially for smaller land cover classes such as mining or urban. The improvements in classification accuracy achievable by using a CNN could circumvent the weaknesses of post-classification change detection methods and lead to simple and effective change mapping.

By bringing together recent advances in deep learning and satellite sensor technology, we propose a new method to automatically map the extent of alluvial ASM activities at a hitherto unprecedented level of accuracy and detail for minimal cost, allowing for effective ASM monitoring. The objectives of this paper are to: i) design a multispectral CNN model capable of distinguishing between mining, built/developed and vegetation land use classes from freely available Sentinel-2 imagery, ii) benchmark the model's performance against other classification methods, iii) demonstrate the performance of this model for mining area detection across a large spatial and temporal range of images, with minimal human input and iv) provide maps which could be used to analyse the impacts of ASM policies over the studied time period. The novel contributions of this work include: the development of a multispectral deep learning model which is significantly more accurate than existing techniques for detecting ASM, a full processing pipeline for monitoring ASM and urbanisation via spatial and temporal mosaicing, and the production of a large scale dataset showing the extent and expansion of ASM in the Ghanaian case study area from 2015-2019.

2 Methods

In this study we built a multispectral CNN designed to detect mining and built environments from Sentinel-2 satellite data. Alongside this, several traditional pixel-based machine learning classifiers were employed to benchmark the CNN method against current practices. We then investigated how the design of the CNN impacts performance across multiple validation patches. Once the model design was finalised, an area of more than 6 million hectares was analysed over 4 years, with a temporal sampling of once per year. To improve the classification reliability of the yearly maps, ensemble methods using probability averaging were used to generate the most probable class from multiple images per season.

2.1 Case Study Area

Ghana was chosen as the case study area as it is at the front line of the ongoing land use changes happening throughout mineral endowed areas of the Global South. Ghanaian rainforest is being deforested at the world's fastest rate, with over 60% more forest lost in 2018 than 2017, initially reported by Global Forest Watch (Weisse and Goldman, 2019) and verified using Sentinel-2 data by Dekker (2019). This is due to mining, agriculture, logging, fires, and urban expansion (Boadi et al., 2016).

Alongside deforestation, ecological issues arising from Artisanal Scale Gold Mining (ASGM) in Ghana include mercury, arsenic and other heavy metal contaminations, decreased water quality and land disturbances (Rajaei et al., 2015). Ghana's economy is heavily reliant on gold and cocoa exports, which often compete for the same land (Snapir et al., 2017). Mining comprised 35% of all national exports in 2014, of which 65% came from large scale mines and 35% from ASM (McQuilken and Garvin, 2016). This is one of the highest percentages attributable to ASM worldwide, largely related to the influx of Chinese miners since the mid-2000s, bringing with them heavy machinery and causing an accelerated rate of land degradation (Botchwey et al., 2018).

Government policy towards ASM has been varied. Initially Ghana was a global leader in ASM formalisation with the Small-Scale Gold Mining Law in 1989 (McQuilken and Garvin, 2016); however, the licensing process was bureaucratically challenging and precluded access to the system for most poverty driven artisanal miners, propelling the sector into increasing informality (Hilson, 2001). To address the issues associated with ASM, in 2013 the government set up a National Task Force to curb operations; however, it was perceived by many as a façade to placate the media and the public (Hilson et al., 2014). All small-scale mining including legal operations was banned in April 2017 as part of the Task Force's Operation Vanguard (Botchwey et al., 2018, Ministry of Lands and Natural Resources, 2019), with the ban partially lifted in December 2018 (Bansah et al., 2018).

The study area boundary chosen is the Ghanaian territory corresponding to the Precambrian West African Craton (Labou et al., 2020), shown in **Fig. 1**. This area covers the major gold belts of Ghana and corresponds to all of the Western and Central provinces and large parts of the Brong Ahafo, Ashanti and Eastern Regions. Additionally, the area contains the ten districts inspected by Owusu-Nimo et al. (2018) in a fieldwork based study, allowing results from our model to be examined within the context of existing ground truth data. The findings from Owusu-Nimo et al. (2018) can provide valuable supplementary information for interrogating the results of our model, despite the scales and study types not being directly comparable.

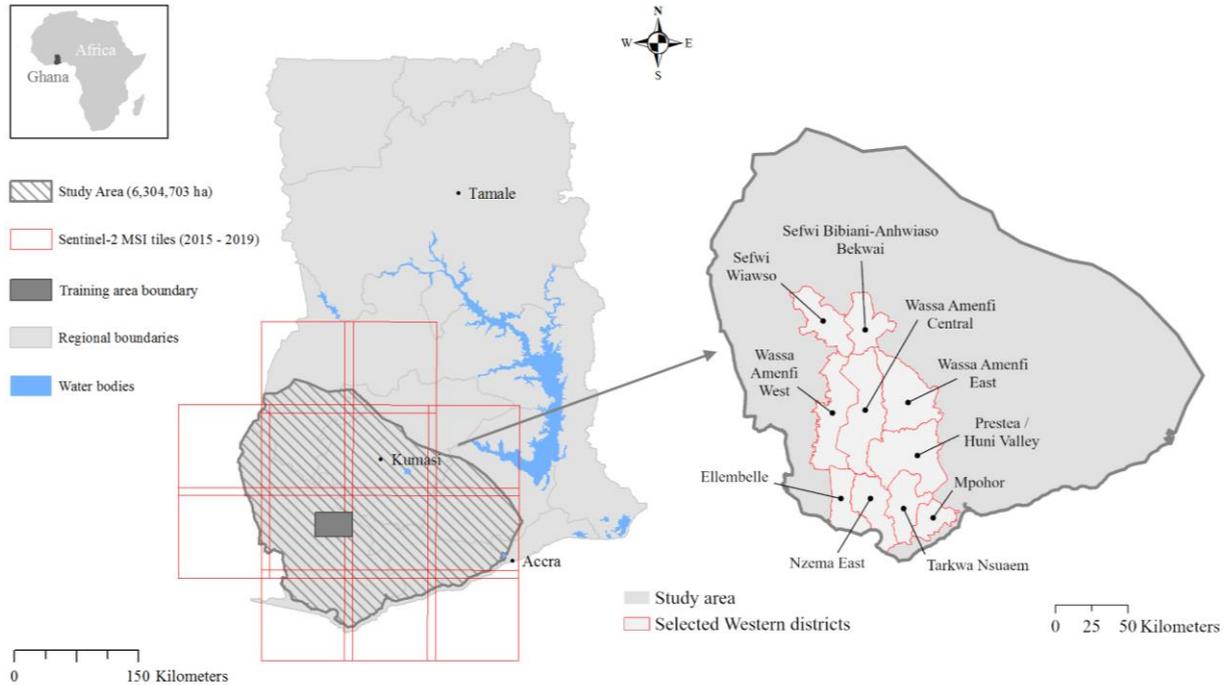


Fig. 1. Overview map of study area, Sentinel-2 tile coverage and training data boundary. Righthand sub-image shows the ten districts common to the Owusu-Nimo et al. (2018) study.

2.2 Datasets

The imagery chosen for this research was sourced from the ESA’s Sentinel-2 MultiSpectral Instrument (MSI). This platform was selected because: i) it offers spatial resolutions of 10m or 20m per pixel (depending on band) in the visible and infrared, adequate to detect most alluvial mining sites; ii) at these spatial resolutions it provides ten spectral bands, ranging from 490nm to 2200nm, delivering sufficient spectral resolution, particularly in the short wave infrared range; iii) it is ‘future-proof’ with operation planned until 2025 with potential for extension to 2030; and most importantly iv) it is open access data (European Space Agency, 2015, van der Meer et al., 2014). Forkuor et al. (2020) state that open access satellite data can be valuable in assisting data-scarce developing countries to measure progress towards SDG targets. Whilst other sensors such as Planet’s PlanetScope satellites offer faster revisit times and higher spatial resolutions, as explored by Shendryk et al. (2019), it was deemed important to avoid potentially costly commitments to private companies when designing a tool to be used for sustainable development purposes. This, alongside its higher spectral resolution informed our choice of Sentinel-2 data.

The Sentinel-2 data was downloaded as 100km x 100km UTM registered orthorectified tiles, known as ‘granules’ in the Sentinel-2 nomenclature (European Space Agency, 2015) in either L1C (top of atmosphere) or L2A (bottom of atmosphere) formats. The total study area partially or wholly intersected 13 different tile footprints. Between three and five granules containing less than 20% cloud coverage were downloaded for every dry season in Ghana (November - March), hereafter known as the temporal

period. Three tile footprints intersected the study area by less than 4%. To reduce dataset size only one granule per temporal period was downloaded for these areas, as it was possible to manually select a granule with completely clear skies over the small area of interest. In total 211 granules were downloaded, details of which can be found in supplementary file S1. The ESA command line program `sen2cor` was used to convert the L1C tiles to L2A (Gascon et al., 2017). The six bands with spatial resolutions of 20m were upsampled to 10m using bicubic interpolation (Vaiopoulos and Karantzas, 2016). For clarity, pixels corresponding to lakes, known large scale mining operations and a 5km buffer along the Atlantic coast were manually masked out and do not contribute further to the analysis.

To provide training data the model, a 30km x 45km area centred at 2°12'W 5°55'N in the Western province was digitised, location shown in **Fig. 2**. This area has high mining activity, helping to mitigate the issues arising from imbalanced classes, where results can be biased due to the larger number of examples in one class in relation to another (Peng et al. 2019, Shendryk et al., 2019, Wurm et al., 2019). Initially, the Sentinel-2 granule from the 23rd January 2016 was manually digitised into the three macroclasses of mines, built areas¹ and vegetation. This granule was chosen as it was the earliest cloud free granule available, allowing high resolution RGB imagery (0.31-0.5m pixel size) imaged by DigitalGlobe's WorldView satellites in 2015 to be used for additional reference (DigitalGlobe, 2015). In the Sentinel-2 imagery, the visibility of mines and settlements generally was sufficient to allow confident human digitisation; where confusion areas occurred the higher resolution imagery was consulted, with due consideration given to the temporal shift between the datasets. The initial 2016 digitisation was used as a starting point for digitising two further training tiles covering the same geographic area but imaged in on the 11th of February and the 2nd of April 2019. These additional temporal tiles were included to obtain training examples under differing radiometric and atmospheric conditions, improving the final model's generalisation ability. Training area labels were digitised by one operator and checked by another. This check indicated high reliability of generated labels, with only small areas and class boundaries displaying disagreements, mainly arising from the subjective nature of delineating precise mine boundaries. When generating the training boundaries, the emphasis was placed on delineating land use rather than pixel level land class, therefore, small patches of vegetation within mined areas were included in the mining class rather than the vegetation class. This strategy more closely represents the actual land area degraded by mining and also generates smoother class boundaries for later interpretation.

For validation, two 5km x 5km areas were digitised using the same methodology. Validation Tile A was generated to the north of the training area using the same 2016 granule, while Validation Tile B was generated from a 2018 granule not used for training. **Fig. 2** shows the of training and validation areas. These two validation tiles allow performance to be examined under both known and unknown

¹ The built class includes areas of bare earth along roads and around buildings but does not generally include fallow fields unless contiguous with other dwellings.

radiometric conditions. For testing, eight additional 5km x 5km tiles were generated using the same workflow. Three of these tiles were extracted from the same larger 2016 Sentinel-2 granule used for training; these make up the known radiometric test set. Another three were extracted from different years and locations within the study area, these make up the unseen Ghana test set. Lastly, two tiles, from Surinam and Indonesia, were created to investigate the model’s global generalisation ability. All tiles are representative of the type of mining landscape found across the region, containing clear examples of all classes and minimal cloud occlusion. Full details of the test tiles are found in supplementary file **S2**.

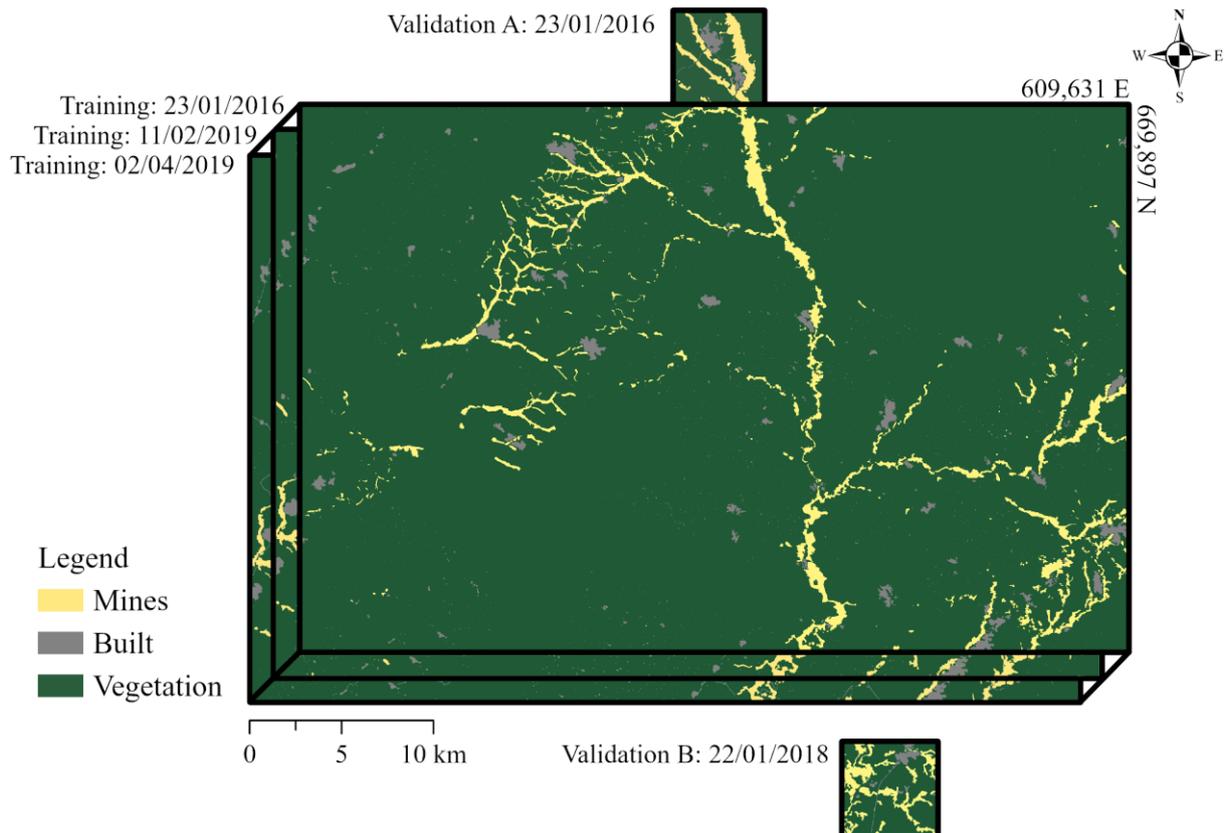


Fig. 2. Locations of the manually digitised training and validation areas. Dates refer to the acquisition date of the imagery, the training area boundary is identical for all years, the offset is used to indicate temporal change.

The contiguous training and validation coverages were then converted to a patch-based structure, similar to that described by Wang et al. (2019). This structure was chosen for its ability to ingest large scenes in a memory efficient manner, whilst decreasing overfitting tendencies. From the full labelled training scene of 9,000 x 4,500 pixels, 16,000 images of size 256 x 256 pixels were randomly extracted. During training, these images were augmented with random X and Y reflections and random rotations from 0° to 90° every epoch. From this augmentation and patching strategy, it is possible to feed the model 480,000 individual training patches generated from the initial 9,000 x 4,500 reference image, leading to robust generalisation from limited human annotated source data.

2.3 Cloud Removal

In areas of the tropics where much of the world's ASM activities are located, it is not uncommon to obtain only one completely cloud free image per year, despite weekly satellite revisit times. The area of southern Ghana chosen for this study is one of the cloudiest regions of the world (Coulter et al., 2016); therefore, utilising partially cloudy images is essential for any robust ASM monitoring system. Sentinel-2 software comes with in-built cloud removal tools; however, these are based on the high water reflectance of clouds in certain bands (European Space Agency, 2019) which also occur in alluvial ASM ponds. When tested, the Sentinel-2 cloud removal tools were unable to distinguish between clouds and ASM, removing large areas of valid pixels. To replace the Sentinel-2 cloud removal tool, pixels exceeding a set reflectance threshold in all three of the visible wavelength bands were masked out, as the main differentiable feature between the clouds and the ponds is their hue in the visual spectrum. This method was effective on the majority of the isolated cumulus clouds found during the dry season. Cloud edge boundaries and high cirrus clouds were not removed by this technique but eliminated later in the ensemble averaging stage, allowing potential detections to be made under light cloud obscuration.

2.4 Model Development

The model chosen for this research is a type of U-Net, a successful encoder-decoder network designed by Ronneberger et al. (2015) for biomedical image segmentation. It has an encoder path consisting of multiple convolutions, ReLU activations and maxpooling operations, followed by a decoder path which upsamples the lower level stacks with the aid of skip connections from the higher resolution layers, combined with further convolutions and ReLU activations. It has been proven to be a highly effective architecture, particularly for problems with limited training data (Ronneberger et al., 2015) and has been the model of choice for multiple remote sensing applications (Iglovikov et al., 2017, Peng et al., 2019, Zhao et al., 2019). The implementation used here is adapted from a multispectral U-Net available online² in MATLAB format (MathWorks, 2019). Our version, which is modified to ingest Sentinel-2 data is available in supplementary file **S3** and is illustrated in **Fig. 3**. The model was created and trained with MATLAB on a desktop computer with a single NVIDIA Titan X GPU and 64 GB of RAM. Using this hardware setup, training time was approximately 16 hours.

² <https://uk.mathworks.com/help/images/multispectral-semantic-segmentation-using-deep-learning.html>

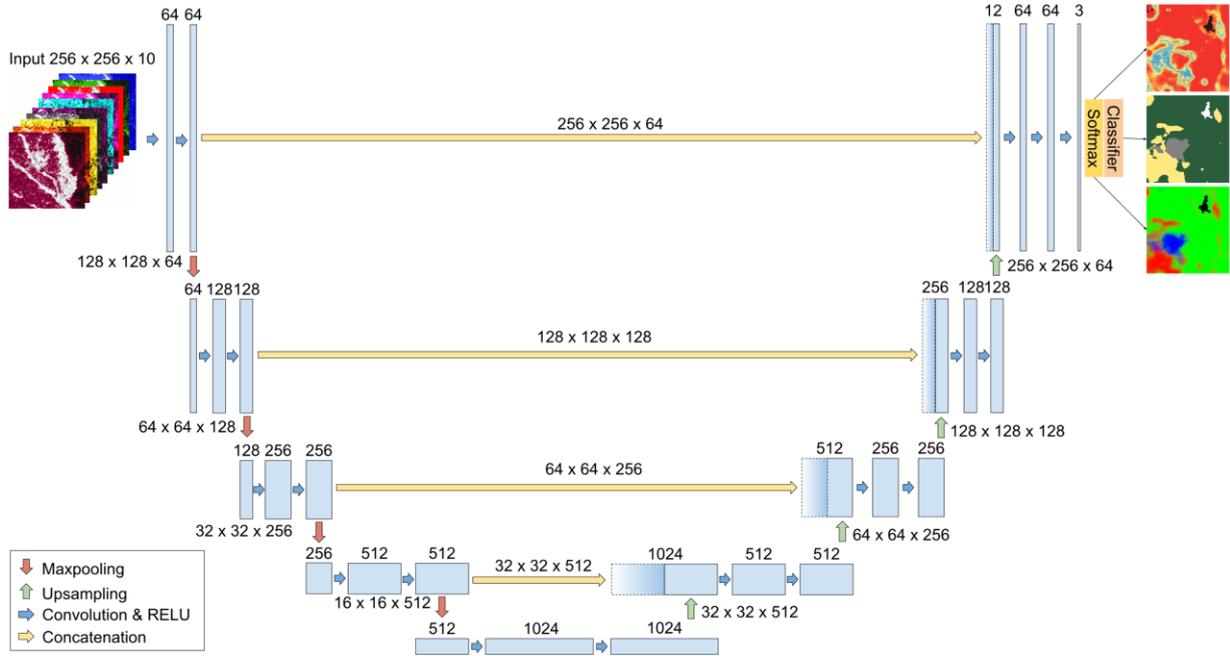


Fig. 3. Schematic of the U-Net architecture used for the model. The input is a 10 channel multispectral image of 256 x 256 pixels and the outputs include: a 1 channel prediction mask, a 1 channel prediction probability layer for the highest class probability and a 3 channel prediction probability layer for all classes, all at the same resolution as the input image.

Due to the small training set size, strong regularisation strategies were employed to minimise overfitting. Alongside the data augmentation discussed in **Section 2.2**, two 50% dropout layers (Srivastava et al., 2014) were added, one at the end of the encoder path and another at the model’s midpoint. Stochastic Gradient Descent with Momentum (SGDM) was chosen as the optimiser as it has been shown to provide better generalisation abilities than adaptive optimisation methods (Wilson et al., 2017). The momentum hyperparameter was fixed at 0.9; this value is typical and effective in practice (Géron, 2017, Goodfellow et al., 2016, Hinton, 2012). Further hyperparameters were determined by training multiple models; during training the validation patch loss was monitored and after training a manual inspection of the output prediction masks for the two validation tiles was carried out. An adaptive learning rate was used, initiated at $5e-2$ and reduced by a factor of 10 every 10 epochs. This was chosen iteratively by examining loss curves during training. Weight decay was set to $1e-4$. Increasing this value to $5e-4$ did not improve model performance. With the hardware configuration described above, it was found that a field of view of 256 x 256 and a mini-batch size of 16 fitted into memory and provided high quality results. Neither increasing the field of view to 512 x 512 nor increasing the mini-batch size to 128 improved model accuracy, with both requiring longer training times. The mini-batches were shuffled after each epoch to increase convergence and improve accuracy (Bengio, 2012). Automated early stopping was not used, however, human monitoring of validation loss during training showed a divergence in training versus validation accuracy after 30 epochs; therefore, the number of training epochs was set to 30.

Loss was calculated using binary cross-entropy, as used in the original U-Net implementation (Ronneberger et al., 2015). Dice loss was investigated; however, this led to unstable training and poor convergence. As the dataset contains highly unbalanced classes, experiments were carried out to obtain the best class weighting strategy for the loss function. Initially, inverse proportional weighting was trialled (Huang et al., 2016); however, this resulted in undue importance being given to minimising omission errors in the rarer classes, resulting in notably decreased accuracy. Halving the weight of the majority vegetation class provided the most balanced results and was the strategy used for the final trained model.

2.5 Pixel based method comparison

To compare the results obtained from the U-Net model, benchmarking was carried out against several pixel-based machine learning classifiers. The first classifier tested is Spectral Angle Mapping (SAM), implemented in the Supervised Classification Plugin (SCP) (Congedo 2016) of the opensource QGIS software. Processing was carried out using the 2016 training data following the workflows described in Congedo (2016), Boakye et al. (2019) and Obodai et al. (2019). Alongside the SCP workflow, another experiment was undertaken to determine whether stronger machine learning classifiers such as a Multi-Layer Perceptron (MLP) or a Random Forest (Breiman, 2001) were better able to model the pixelwise relationships between the spectral reflectances and the land cover classes.

To test this hypothesis, MLP and Random Forest classifiers were built with the Scikit-learn libraries in Python (Pedregosa et al., 2011). The MLP used is a simple shallow model with two fully connected hidden layers of 10 and 5 nodes respectively, ReLU activations, an adaptive learning rate and ADAM solver. For the Random Forest classifier, 100 trees were used for the model. For both these models, class rebalancing was carried out to reduce the vegetation class to five times the mining class. A hyperparameter search was undertaken using 4 fold cross-validation on the training data; however, negligible differences were observed. Processing involved unrolling each image into a single $n \times m$ vector for input to the classifiers, where: $n = \text{image width} \times \text{image height}$, $m = 10$. Post-classification, the prediction masks were reshaped back to the original image dimensions to assess their accuracy against the human generated masks.

2.6 Post processing

Following the CNN model prediction stage, a result stacking strategy was developed to combine predictions across temporal periods. Ensemble methods such as stacking have been shown to improve the performance of most machine learning classifiers with the greatest gains inversely proportional to model correlation (Dietterich, 2000). Prediction stacking was also necessary for this application in order to mosaic together multiple partially cloud-occluded prediction maps. The U-Net model was modified to output both the classifications from the final layer and the per class prediction probabilities from the softmax layer (**Fig. 4**. Model outputs showing a) the Sentinel-2 RGB data as a reference, b) the single predicted class labels, c) the model's confidence in its predictions and d) the prediction probabilities for

every class mapped to the RGB colour channels). As there are only three land cover classes for this application, added value was gained by mapping the per class probabilities to each of the colour channels, allowing intuitive visual interrogation of the model's predictions. This probability visualisation map goes some way towards addressing a criticism often levelled at deep learning models that they are a 'black box' solution that does not reveal the processes which led to the results. The probability mapping provides insights into how the model thinks and why it makes mistakes. Muddled colours represent areas of confusion, where the model has predicted similar probabilities for multiple classes, as seen in **Fig. 4. (d)**.

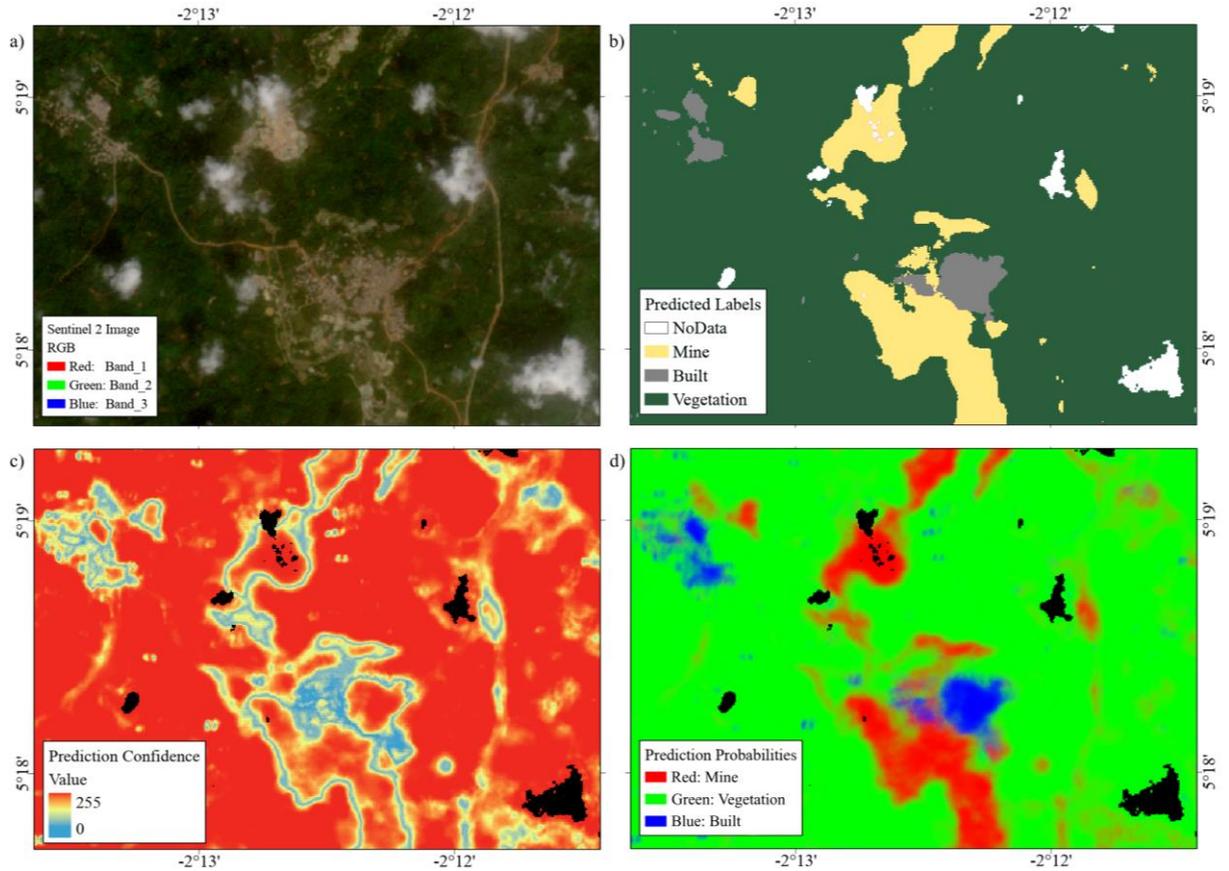


Fig. 4. Model outputs showing a) the Sentinel-2 RGB data as a reference, b) the single predicted class labels, c) the model's confidence in its predictions and d) the prediction probabilities for every class mapped to the RGB colour channels

Whilst the simplest model ensemble would combine the final classified images for each temporal period, utilising the full prediction probability images allows greater weight to be given to the more confident predictions. The models were combined as shown in Equations 1-3. For a dataset of l number of images, each with 3 probability classes mapped to the channels r, g, b and pixels i, j :

$$result(i, j) = \max(r_{(i,j)}, g_{(i,j)}, b_{(i,j)}) \quad (1)$$

$$where \quad r_{(i,j)} = \frac{\sum_1^l dataset(i,j,r,l)}{valid_{(i,j)}} \quad g_{(i,j)} = \frac{\sum_1^l dataset(i,j,g,l)}{valid_{(i,j)}} \quad b_{(i,j)} = \frac{\sum_1^l dataset(i,j,b,l)}{valid_{(i,j)}} \quad (2)$$

$$and \quad where \quad valid_{(i,j)} = \text{number of valid pixels at location } (i,j) \text{ across dataset } l \quad (3)$$

The final pixel class is determined by averaging the probabilities from each prediction image per class over every pixel, then taking the class with the highest mean probability score. Null values from occluded pixels did not contribute to the averages. This method rewards predictions with high confidence and minimises the contribution of pixels with high confusion. It is particularly effective at removing noise around the boundaries of clouds, as the confused pixel is only present in one of the temporal period images. After this temporal ensembling, the image tiles are geographically mosaiced using the maximum class value in overlap areas to prioritise vegetation and minimise clouds. The final model outputs are four images of predicted class probabilities over the entire study area, corresponding to each dry season of the study period (winter 2015/16 – winter 2018/19).

2.7 ASM monitoring

From the prediction maps, several datasets were generated to further the geospatial understanding of ASM patterns in Ghana. Firstly, yearly change maps are generated by subtracting the overall prediction maps using a simple integer change mapping schema, detailed in supplementary file **S4**. Secondly, to visualise the mining related changes across the entire study area, a kernel density heatmap was generated showing the density of new ASM pixels (100m²) per square kilometre for each year. Lastly, to study how illegal mining is encroaching on Ghana's forest reserves, the total number of new mining pixels detected inside or adjacent to protected forests over the course of the study was calculated. The geospatial data for the forest reserve boundaries was sourced from the Ghana Open Data Initiative portal (Forestry Commission, 2010).

2.8 Model evaluation

Accuracy assessment techniques from both the computer vision and remote sensing communities were used to evaluate the model. Pixel masks commonly used in computer vision applications were used in the first two assessments. These assessments compared the raw prediction masks output by the model to a manually digitised reference mask. The first assessment used tiles extracted from the same larger 2016 Sentinel-2 granule used for model training, as the SCP classification methodology is not designed to generalise to different satellite images acquired on different days due to the changes in radiometric values (Congedo, 2016). Validation Tile A was used by all classifiers for model tuning, Test Tiles B-D are fully unseen. All classifiers in the first test were trained on only the 2016 training dataset. For the second assessment, three further test tiles from within the Ghanaian case study area but acquired on different years were used, along with two international test tiles chosen to investigate the model's global applicability to other spectrally similar ASM sites. The classifiers used in the second assessment were trained on both the 2016 and 2019 training datasets.

The metrics used for evaluating these assessments are mean class accuracy, mine class accuracy, mean class Intersection Over Union (IoU) and mine class IoU. These metrics are derived from the relationships between True Positives (correctly detected pixels, TP), False Positives (incorrectly detected pixels, FP)

and False Negatives (undetected pixels, FN). Mean class accuracy is defined as the mean of the per class accuracy, where the per class accuracy is the number of true positives (TP) per class divided by the total number of pixels per class (Equations 4 and 5).

$$Acc_{mines} = \frac{TP_{mines}}{(TP_{mines}+FN_{mines})} \quad Acc_{built} = \frac{TP_{built}}{(TP_{built}+FN_{built})} \quad Acc_{veg} = \frac{TP_{veg}}{(TP_{veg}+FN_{veg})} \quad (4)$$

$$MeanClassAccuracy = \frac{Acc_{mines}+Acc_{built}+Acc_{veg}}{3} \quad (5)$$

Mean class accuracy, though intuitive, can create misleading results, especially in the presence of many false positives. The IoU score is a robust and commonly used metric for semantic segmentation problems as it provides a statistical accuracy metric which penalises both false positives and false negatives. Each class IoU is calculated as the number of true positives divided by the number of true positives, false positives, and false negatives, shown in Equation 6. The mean class IoU is the average of the IoU scores for each class.

$$IoU_{class_n} = \frac{TP_{class_n}}{(TP_{class_n}+FN_{class_n}+FP_{class_n})} \quad (6)$$

These pixel level assessments are standard in computer science image segmentation problems and provide a challenging test of the model's abilities. However, they are not directly comparable to the point based stratified random sampling accuracy assessments described in Olofsson et al. (2014) and used in related remote sensing works such as Snapir et al. (2017), Obodai et al. (2019) and Espejo et al. (2018). In order to more closely compare our CNN method to prior work, a third assessment, based on stratified random sampling, was carried out over the entire study area for each year using the post processed prediction maps. The sample points were chosen according to Equation 7 (Cochran, 1977)

$$\text{described in Olofsson et al. (2014): } n = \frac{(\sum W_i S_i)^2}{(S(O))^2 + \left(\frac{1}{N}\right) \sum W_i S_i^2} \approx \left(\frac{\sum W_i S_i}{S(O)}\right)^2 \quad (7)$$

where N = total number of pixels, $S(O)$ is the standard error of desired accuracy estimate, W_i is the proportional area of each class and S_i is the standard deviation of each class, calculated from the user's accuracy U_i as $S_i = \sqrt{U_i(1 - U_i)}$. Using the formula from Equation (7) and a desired standard error of 0.01 gives a suggested value of 509 sample points. For point class allocation, the points were divided into 50% for the largest class and 25% each for the smaller classes, giving totals of 254, 127 and 127. These numbers were increased to 300, 150 and 150 to allow for loss due to falling within masked areas and pixels which were unidentifiable to a human from the available Sentinel-2 data. All 600 points were manually compared against the corresponding year's Sentinel-2 RGB images for every temporal epoch of the study. Higher resolution data was consulted when available for the precise temporal period. The metrics used to report the results from this accuracy assessment are overall accuracy, class accuracy, commission errors and omission errors.

3 Results

The experiments carried out in this study aim to address two key aspects of the research. The first is concerned with determining the accuracy of the CNN model and comparing it to other published methods. The second examines how the model can then be used to effectively monitor and manage ASM, particularly to support sustainable development in the Global South.

3.1 Accuracy assessments

The first assessment compares the results from the initial CNN model against three other machine learning classification methods. The test tiles for this assessment are radiometrically similar to the training data as they are extracted from the same larger Sentinel-2 granule. The results from this assessment are shown in **Table 1**. The second assessment evaluates the model’s generalisation ability against radiometrically unseen test data from both within the study area and from ASM operations in Surinam and Indonesia. The results from the second assessment are shown in **Table 2** and **Fig 5**. The SAM classifier was not included in the second assessment as the SCP workflow is not designed for processing radiometrically unseen granules.

Table 1. Results from the first assesement. This assesemnt evaluated the results from each classifier on the radiometrically similar test tiles. Full explanation of accuracy metrics is provided in Section 2.8.

	Mean Class Accuracy	Mine Class Accuracy	Mean IoU	Mine Class IoU
Spectral Angle Mapping				
Validation Tile A	0.71	0.82	0.65	0.73
Test Tile B	0.67	0.75	0.62	0.68
Test Tile C	0.69	0.73	0.63	0.64
Test Tile D	0.75	0.75	0.66	0.56
<i>Average</i>	0.72	0.79	0.65	0.68
Random Forest				
Validation Tile A	0.87	0.90	0.67	0.63
Test Tile B	0.85	0.93	0.81	0.86
Test Tile C	0.85	0.90	0.73	0.73
Test Tile D	0.88	0.85	0.75	0.63
<i>Average</i>	0.86	0.90	0.74	0.71
Multi-Layer Perceptron				
Validation Tile A	0.86	0.89	0.67	0.64
Test Tile B	0.84	0.93	0.80	0.87
Test Tile C	0.86	0.90	0.75	0.75
Test Tile D	0.88	0.85	0.76	0.67
<i>Average</i>	0.86	0.89	0.75	0.73
Convolutional Neural Network				
Validation Tile A	0.92	0.93	0.88	0.89
Test Tile B	0.93	0.95	0.89	0.90
Test Tile C	0.92	0.95	0.86	0.88
Test Tile D	0.94	0.91	0.83	0.78
<i>Average</i>	0.93	0.94	0.87	0.86

Table 2. Results from the second assesement. This assesemnt evaluated the results from each classifier on the radiometrically unseen test tiles. All test tiles are from the different granules than those used for training. Full explanation of accuracy metrics is provided in Section 2.8.

		Mean Class Accuracy	Mine Class Accuracy	Mean IoU	Mine Class IoU
Random Forest					
Eastern	2017/03/28	0.72	0.97	0.49	0.31
Kumasi	2017/01/27	0.82	0.79	0.75	0.73
Obuasi	2019/01/22	0.70	0.70	0.60	0.65
<i>Ghana Average</i>		0.75	0.82	0.61	0.56
Surinam	2019/12/01	0.65	0.53	0.52	0.50
Indonesia	2019/09/04	0.65	0.58	0.49	0.53
Multi-Layer Perceptron					
Eastern	2017/03/28	0.78	0.96	0.56	0.36
Kumasi	2017/01/27	0.83	0.80	0.76	0.75
Obuasi	2019/01/22	0.78	0.96	0.58	0.61
<i>Ghana Average</i>		0.80	0.91	0.63	0.57
Surinam	2019/12/01	0.67	0.55	0.53	0.51
Indonesia	2019/09/04	0.66	0.43	0.44	0.40
Convolutional Neural Network					
Eastern	2017/03/28	0.92	0.96	0.86	0.83
Kumasi	2017/01/27	0.94	0.97	0.83	0.78
Obuasi	2019/01/22	0.87	0.96	0.79	0.84
<i>Ghana Average</i>		0.91	0.96	0.83	0.82
Surinam	2019/12/01	0.73	0.83	0.61	0.71
Indonesia	2019/09/04	0.84	0.82	0.57	0.70

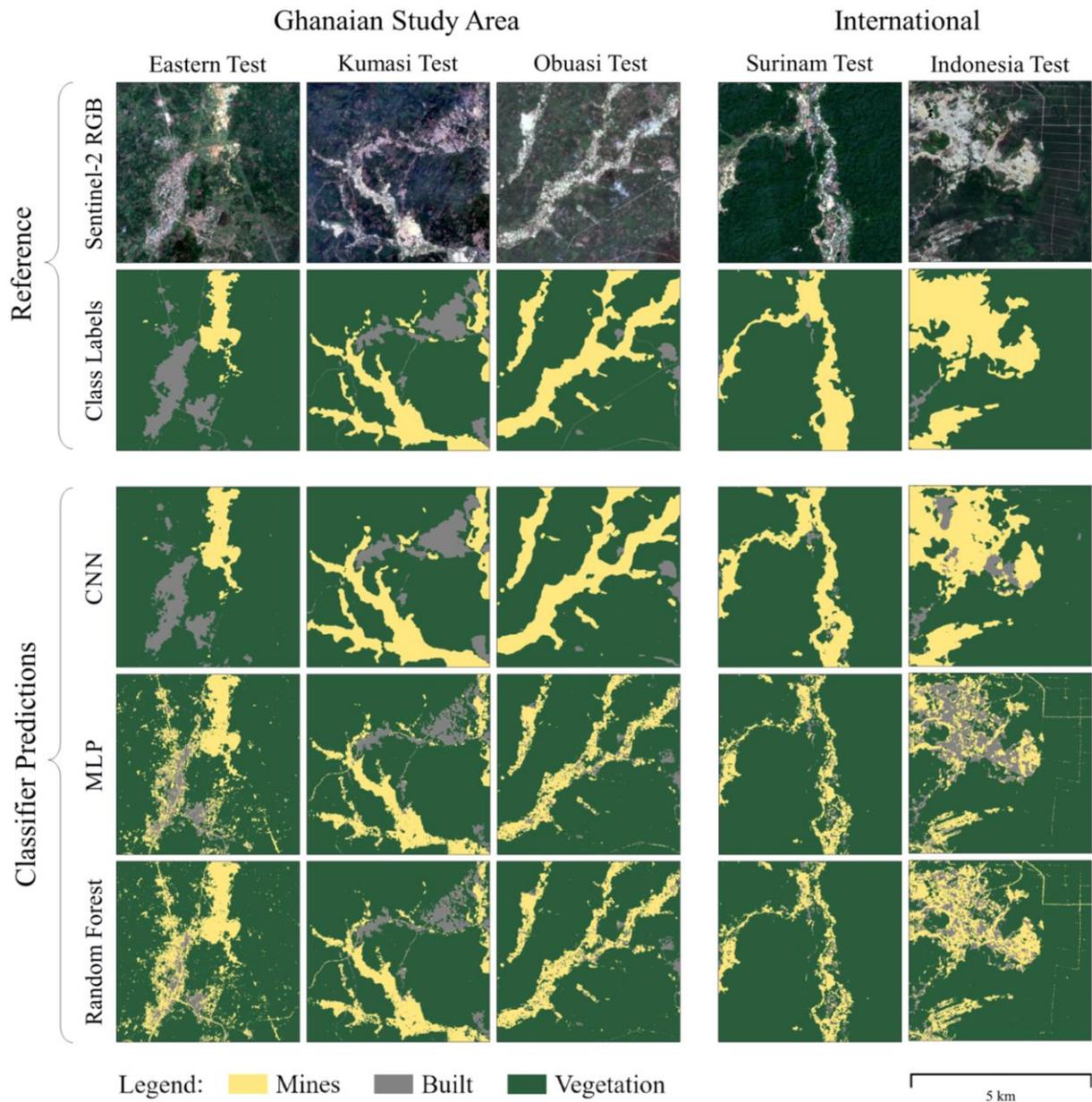


Fig 5. Qualitative view of the results from assessment two. Reference information for all five unseen test tiles is displayed alongside the corresponding prediction masks from the various machine learning classifiers.

The first two accuracy assessments (**Table 1** and **Table 2**) were carried out using the raw prediction results from the CNN model on small individual test tiles, allowing the prediction results to be compared to the labels generated by an operator from a single Sentinel-2 reference image for every pixel. To evaluate the performance over the entire nationwide study area for each temporal period, a third accuracy assessment was carried out, this time based on point based random sampling. This assessment used the post processed class prediction maps, as described in **Section 2.8**. The results assessment three are given in **Table 3**. The class accuracies obtained here are higher than the class accuracies from the pixelwise assessments, due to both the probability voting at the ensembling stage and the lower chances of the randomly sampled points coinciding with confusion areas along class boundaries.

Table 3. Confusion matrices and accuracy metrics for the results from the point based stratified random sampling assessment. The metrics used are class accuracy, overall accuracy, omission error and commission error.

		Predicted			Class Accuracy	Omission Error	Commission Error
		Mines	Built	Vegetation			
Winter 2015-16							
True	Mines	117	1	0	0.99	1%	6%
	Built	2	126	1	0.98	2%	13%
	Veg	5	18	299	0.93	7%	0%
					<i>Overall Accuracy</i>		<i>0.95</i>
Winter 2016-17							
True	Mines	107	0	2	0.98	2%	4%
	Built	4	119	5	0.93	7%	3%
	Veg	0	4	327	0.99	1%	2%
					<i>Overall Accuracy</i>		<i>0.97</i>
Winter 2017-18							
True	Mines	101	2	4	0.94	6%	0%
	Built	0	125	3	0.98	2%	5%
	Veg	0	4	331	0.99	1%	2%
					<i>Overall Accuracy</i>		<i>0.98</i>
Winter 2018-19							
True	Mines	107	0	0	1.00	0%	2%
	Built	1	122	7	0.94	6%	2%
	Veg	1	3	329	0.99	1%	2%
					<i>Overall Accuracy</i>		<i>0.98</i>

3.2 Applied results

The results obtained in **Sections 3.1** demonstrate that the CNN model is reliably able to locate ASM activity from the Sentinel-2 data to a very high degree of accuracy. Using the full classified area prediction maps and change maps as described in **Section 2.7**, the extents and changes in ASM activities are mapped over time. **Fig. 6** shows the overall land use classes across the whole study area at the start of the study period in winter 2015-16, with insets from the district of Wassa Amenfi East, one of the districts of highest ASM concentrations. The insets show the changes during 2016 alongside the baseline values. The largest scale inset clearly shows the increase in ASM activities both from expansion of existing operations and creation of new ones. Built areas have also increased around the mines, likely due to workforce migration. **Fig. 7** plots the changes in land use categories over the four years of the study. Mining areas increased in 2016 and 2018 but decreased in 2017, built areas increased except in 2016, and vegetation declined steadily each year.

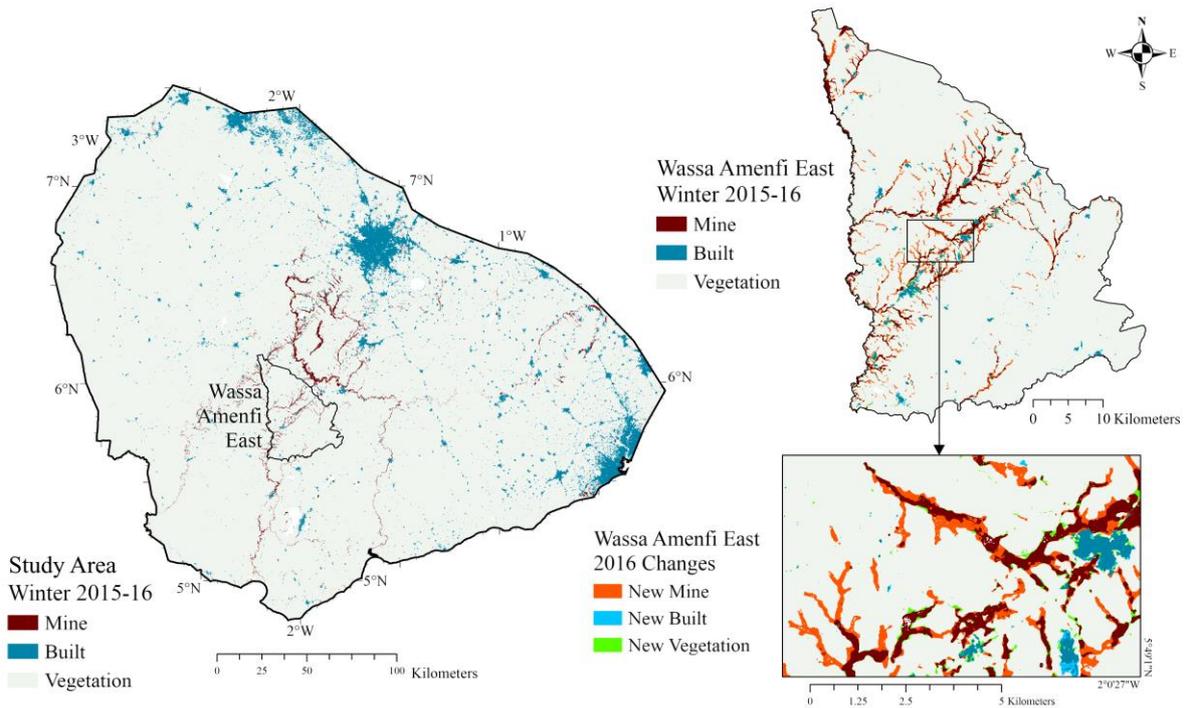


Fig. 6. Prediction map results displayed for the first year of the study period. The main image shows the full study area at the end of the first temporal period. The city of Kumasi is the large area of built up land in the north and the outskirts of Accra are visible in the far south east. The upper sub-image shows the district of Wassa Amenfi East with both the winter 2015-16 baseline data and the 2016 changes. The lower sub-image displays a close-up of some of the changes within Wassa Amenfi East detected during 2016.

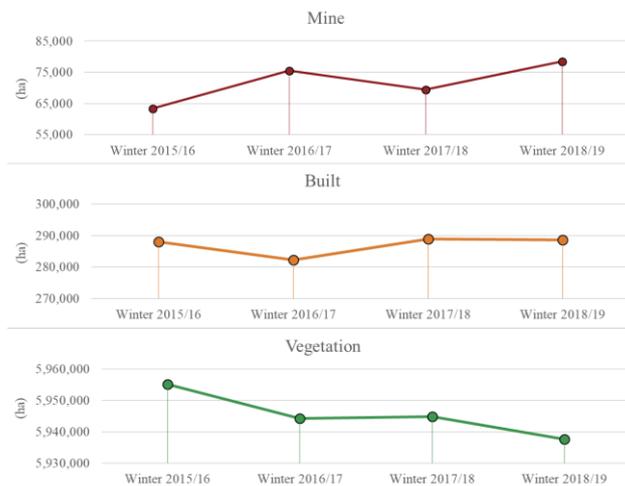


Fig. 7. Total land cover changes per class observed over the entire study area. Y-axis scale is constant between sub-graphs.

To examine the changes in ASM activity in more detail, the total mined area and the yearly mined area changes were computed for ten districts in the Western region. These districts were chosen as they correspond with those studied by Owusu-Nimo et al. (2018), allowing our results to be viewed within the context of their work. The total mined areas and the changes in mined areas are plotted in **Fig. 8**, showing a substantial increase in mining in 2016, followed by a decrease across most areas in 2017, followed by a final smaller increase in 2018.



Fig. 8. Cumulative mined area changes a) and yearly mined area changes b) measured across ten Western districts.

The kernel density heatmap, shown in **Fig. 9**, allows observation of the distribution of new ASM pixels each year. It can be seen that in 2016 the greatest focus was in the Western Region, with other hotspots around Accra, due to sand mining and stone quarrying. In 2017 there is an overall decrease, alongside a general migration north-eastward into the Ashanti Region from the Western Region. There was also an increase in the northern Dormaa districts of Brong Ahafo. In 2018 the intensity resumes in the Western district though at a lower level than seen in 2016. **Fig. 10** indicates the effect that ASM is having on Ghana’s protected forests. This shows the locations of all the forest reserves within the study area. The different colours refer to the hectares of new mining detected within their borders from 2015-2019. The inset shows the yearly mining expansion in the most affected reserve of Upper Wassaw.

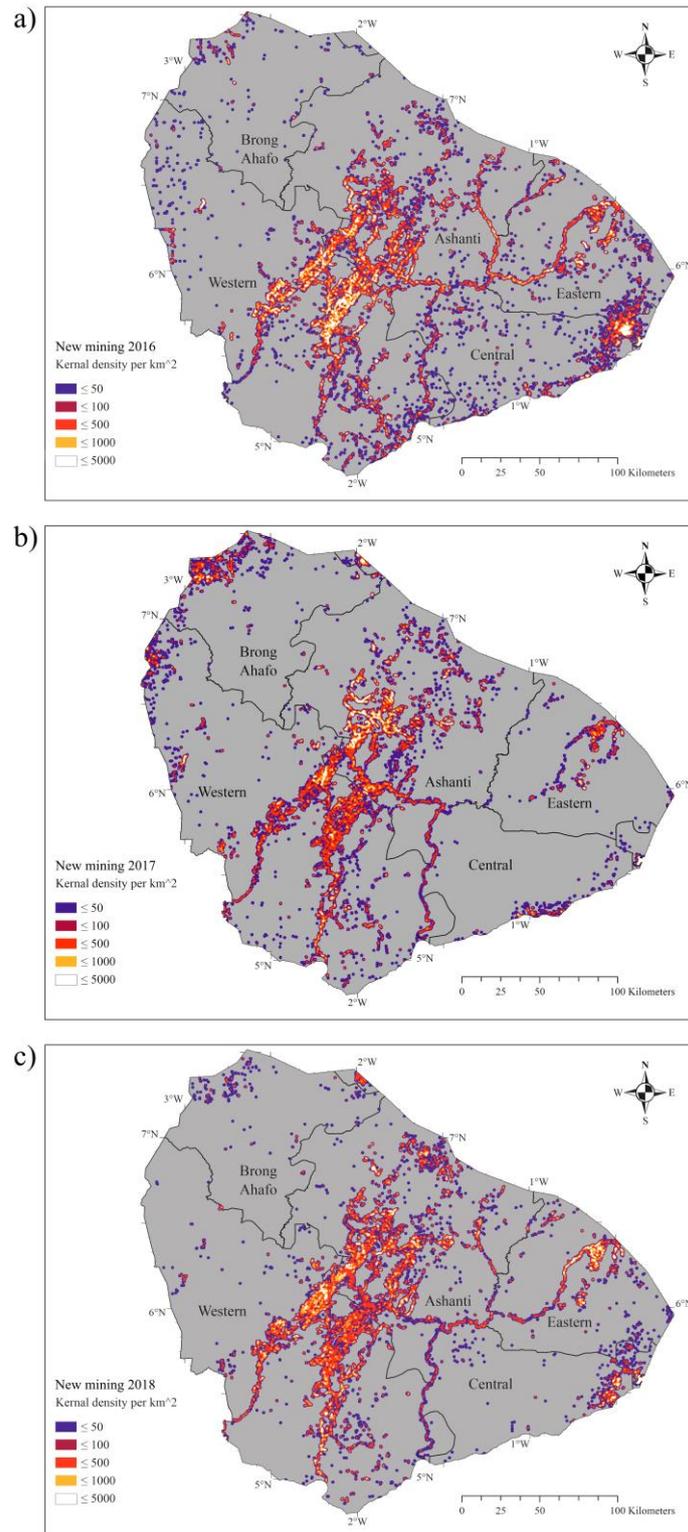


Fig. 9. Kernel density heatmap showing the concentrations of new ASM activity over the study area over the four years. The areas surrounding Accra in the south east are related to sand mining and quarrying, these classifications are outside the scope of the original training data and are provided for visual reference only.

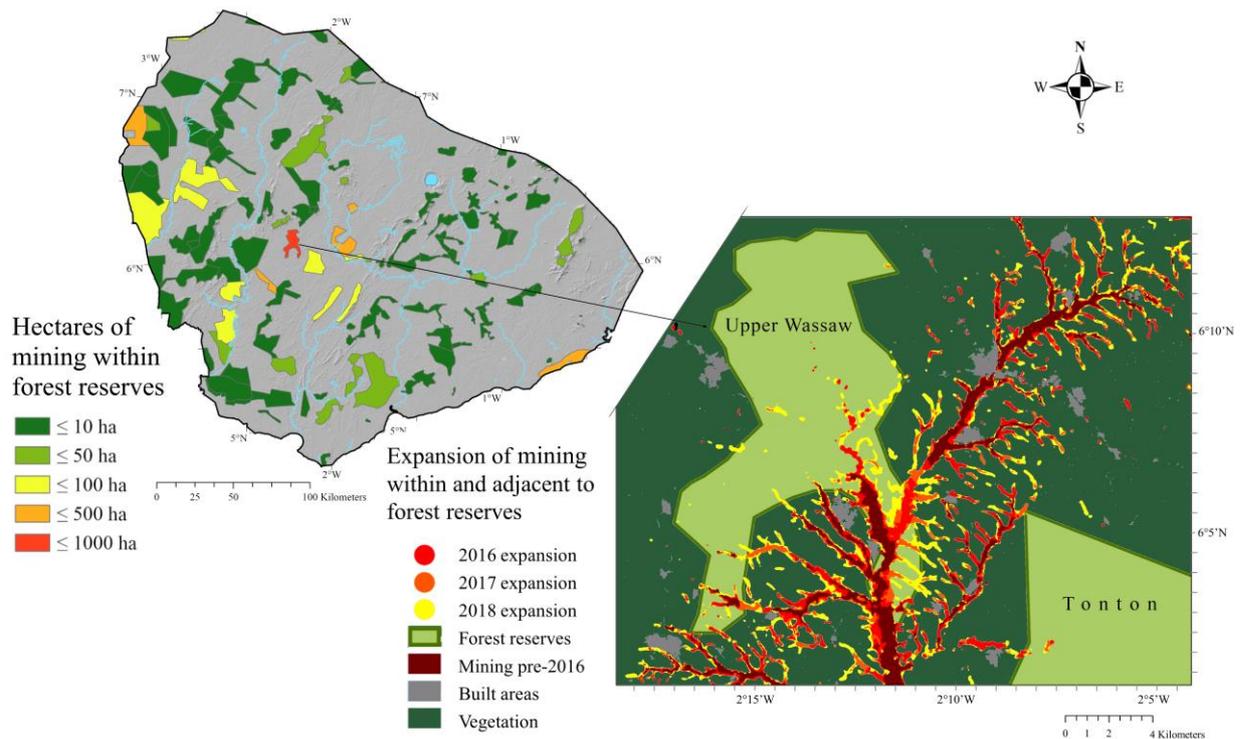


Fig. 10. Maps showing the hectares of mining within Ghana’s national protected forests. The inset shows expansion within the Upper Wassaw reserve. This reserve has the highest rate of mining deforestation at over 800 new hectares lost since 2015. Forest reserve boundaries from Forestry Commission (2010), background GMTED2010 from U.S. Geological Survey.

4 Discussion

The results from the series of accuracy assessments prove that using a CNN is a highly effective method for detecting ASM and its related deforestation. The CNN model substantially outperforms the other machine learning classifiers when both false negative and false positives are considered (**Tables 1 and 2**) and does not suffer from the pixel level noise associated with the other methods (**Fig 5**). The CNN is particularly strong at generalisation; the mean IoU score of the CNN changes from 0.87 for the radiometrically similar data to 0.83 for the radiometrically unseen data. In comparison, the mean IoU drops from 0.75 to 0.63 for the MLP and 0.74 to 0.61 for the Random Forest when moving to the unseen granules. It is likely that the greater generalisation ability observed from the CNN is due to its knowledge of pixel context; even under differing radiometric conditions the patterns between the spectral signatures of adjacent pixels can be recognised. The results from the two global test tiles show promise for generalising to alluvial ASM sites globally, despite different vegetation types and mining styles. This indicates that the model could be applied successfully in other countries with the addition of appropriate local training data.

The stratified random sampling assessment (**Table 3**) confirms the high accuracy of the classification maps, with omission and commission errors consistently below 8% except for a solitary instance. These errors are lower than others previously reported in the literature. It is suspected that the single high commission error of 13% for the built class in 2015-16 is due to the El Nino related drought in the

furthest north region of the study area during that year (Kwadwo, 2019). This resulted in image granules with areas of drought affected fields, which the model was not exposed to during training. These were misclassified as built areas due to their similarity to dirt roads and bare earth around dwellings. If those northern points are removed, the commission error drops to 5%, in line with the other results.

During this assessment it was found that the model predicted some sand mining and quarrying as the mining class, despite not being exposed to these land use classes in training. Only one sampling point from the assessment fell within this confusion area; for clarity, this point was removed from the error matrix table as it is not obvious whether it should be considered a correct or incorrect class for this point. Another anomaly detected from visual inspection of the prediction maps is that polluted rivers with a high sediment load are classified as mines; this is due to having only three possible land use classes in the model. Under this categorisation scheme, polluted rivers most closely resemble alluvial mines and are therefore categorised as such. Whilst the sand mining, drought and the waterway anomalies can be considered errors, they also show that the model is successfully learning from what it has been taught: both land cover types have been assigned to the most similar class based on the extent of the model's prior knowledge. Future studies could incorporate these rarer classes into the training data to avoid such anomalies; however, machine learning algorithms will always be susceptible to new classes that arise after training has been carried out.

The distribution and magnitude of ASM activities observed across the selected Western districts (**Fig. 8**) broadly follow those observed during site visits by Owusu-Nimo et al. (2018). The districts of greatest activity are primarily in agreement, with the only substantial difference found in the district of Tarkwa Nsuaem. This is likely due to both the large number of underground illegal mining operations in this district that are not visible to satellite monitoring systems and also because a large portion of the illegal mining observed by Owusu-Nimo et al. (2018) was within or adjacent to Tarkwa Gold Fields large scale mine which has been masked from this study's results. This agreement between the results from the CNN model and the results from a fieldwork-based study are another indicator that the model is capable of correctly detecting ASM.

The trends observed over the entire study area (**Fig. 7**) and the selected Western districts (**Fig. 8**) show that whilst mining is increasing overall there was a marked decrease in 2017. This corresponds to the launch of the government's Operation Vanguard (See **Section 2.1**). This decrease can also be seen in **Fig. 9**, where the greatest hotspots from the Tarkwa area in 2016 show significant dispersal by 2017. The Upper Pra and Birim Rivers also show minimal new ASM activity in 2017, though the Dormaa area in Brong Ahafo Region shows an increase, possibly due to migration of miners from areas of higher enforcement elsewhere. From analysis of the intersection of illegal mining activities and Ghana's protected forest reserves (**Fig. 10**) it is clear that significant areas of natural rainforest are either at risk from mining or have already been destroyed. In total, over 3,500 hectares of forest preserve had been

deforested by mining in the final land cover map from winter 2018-19. Upper Wassaw is the most affected reserve, with 1,300 hectares of mining deforestation; this is more than 11% of its total area.

The applied results from Section 3.2 give a snapshot of the potential applications for which stakeholders could use this model. These datasets also could be used for other applications such as improving the formalisation pathways for small-scale miners, a vital part of improving ASM management as described in Bansah et al. (2018). They also could be used to verify licenses when combined with a layer detailing the locations of licensed small-scale mines. Additionally, the prediction maps could be used to measure progress towards sustainable development targets such as the Reducing Emissions from Deforestation and Degradation target (REDD+). These potential benefits to the formalisation process are amplified by significant savings in terms of capital and workforce resources.

Moving from research to implementation, it is envisioned that the method described here could be re-coded with fully open-source libraries and a user interface to create a software package accessible to GIS professionals worldwide. Generating training and validation patches is straightforward and can be carried out locally by a GIS technician. The workflow is heavily front-loaded in terms of human and computing resources. The initial model training requires deep learning knowledge and a high-powered computer; however, after this stage the processing of large area prediction maps can be accomplished in a single step by a local GIS technician using a standard workstation. A relatively small manually digitised training area of 140,000 hectares can be used to create a model capable of making predictions on a national scale, provided the essential spectral morphologies remain consistent. Most importantly, the resulting model is capable of making generalised predictions on any past or future image from the same sensor, eliminating the need to retrain the classifier for new images, a major time cost for traditional methods. Furthermore, as it uses open source satellite data, there is zero data purchasing cost, making it an attractive alternative to drone-based methods, particularly in its ability to cover very large areas.

5 Conclusion

This study explored the recommendations from Espejo et al. (2018) that higher resolution imagery and artificial intelligence-based methods would be the key to reducing misclassifications and improving the accuracy of automated methods of ASM detection. Utilising powerful deep convolutional neural networks and high-resolution Sentinel-2 data it was possible to robustly learn the spatial and spectral characteristics of alluvial small-scale mining. This research developed the first published multispectral CNN model for this task. It is highly capable of detecting both mined and built areas from Sentinel-2 open source multispectral satellite imagery, alongside clearly distinguishing between the two categories, a task that has proved problematic in the past (Snapir et al., 2017).

The CNN model has been subjected to a series of accuracy assessments to evaluate its abilities. The performance of the model is state of the art, surpassing previously published accuracy figures. The

model is able to generalise well, with minimal accuracy loss observed between seen and unseen radiometric data. The future incorporation of additional training data from varying climates and land use classes could be used to further extend its generalisation abilities. The model is robust to radiometric noise and accurately follows the boundaries of the mined areas, whilst also reducing the issues arising from extensive cloud cover. The processing pipeline developed here rapidly evaluated an area of over 6 million hectares, proving that this methodology can be scaled up to national level for countries to remotely map and monitor small-scale mining.

The datasets produced in this study show how using CNN satellite-based monitoring could provide governments with rapid and detailed knowledge of small-scale mining changes within their jurisdictions. The temporal resolution achievable is dependent on cloud cover, with near real-time weekly updates possible during the dry season. Improved understanding of the spatiotemporal patterns of small-scale mining could be used to track the effectiveness of a range of mitigation strategies employed across different districts. In the Ghanaian study area, these results show a clear link between the establishment of the anti-illegal mining taskforce Operation Vanguard and a decrease in mining activities in 2017. The datasets also enable monitoring of mining-related deforestation within Ghana's protected forests, allowing stakeholders to identify the reserves most at risk and to prioritise conservation. As well as acting as a deterrent for illegal mining, the methodology also has potential benefits in supporting environmentally sound small-scale mining, by identifying sites with good practices and earmarking them for support. Overall, the methodology has been shown to provide accessible, accurate and inexpensive data on ASM regionally, which can be used to create more sustainable mining practices.

Acknowledgements

This research was carried out as part of an PhD studentship funded by the University of Exeter and received no external funding.

References

Asner, G.P., Llactayo, W., Tupayachi, R., Luna, E.R., 2013. Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18454–18459. <https://doi.org/10.1073/pnas.1318271110>

Asner, G.P., Tupayachi, R., 2017. Accelerated losses of protected forests from gold mining in the Peruvian Amazon. *Environ. Res. Lett.* 12. <https://doi.org/10.1088/1748-9326/aa7dab>

Ball, J.E., Anderson, D.T., Chan, C.S., Ball, J.E., Anderson, D.T., Chan, C.S., 2017. Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools and Challenges for the Community. *J. Appl. Remote Sens.* 11. <https://doi.org/10.1117/1.JRS.11.042609>

Bansah, K.J., Dumakor-Dupey, N.K., Kansake, B.A., Assan, E., Bekui, P., 2018. Socioeconomic and environmental assessment of informal artisanal and small-scale mining in Ghana. *J. Clean. Prod.* 202, 465–475. <https://doi.org/10.1016/j.jclepro.2018.08.150>

Bengio, Y., 2012. Practical recommendations for gradient-based training of deep architectures, in: Montavon, G., Orr, G.B., Müller, K.-R. (Eds.), *Neural Networks: Tricks of the Trade*. Springer, Berlin.

Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F., 2012. ESA's sentinel missions in support of Earth system science. *Remote Sens. Environ.* 120, 84–90. <https://doi.org/https://doi.org/10.1016/j.rse.2011.07.023>

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* 87, 180–191. <https://doi.org/10.1016/j.isprsjprs.2013.09.014>

Boadi, S., Nsor, C.A., Antobre, O.O., Acquah, E., 2016. An analysis of illegal mining on the Offin shelterbelt forest reserve, Ghana: Implications on community livelihood. *J. Sustain. Min.* 15, 115–119. <https://doi.org/10.1016/j.jsm.2016.12.001>

Boakye, E., Anyemedu, F.O.K., Quaye-Ballard, J.A., Donkor, E.A., 2019. Spatio-temporal analysis of land use/cover changes in the Pra River Basin, Ghana. *Appl. Geomatics*. <https://doi.org/10.1007/s12518-019-00278-3>

Botchwey, G., Crawford, G., Loubere, N., Lu, J., 2018. South-South Irregular Migration: The Impacts of China's Informal Gold Rush in Ghana. *Int. Migr.* 57. <https://doi.org/10.1111/imig.12518>

Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>

Cochran, W.G., 1977. *Sampling Techniques*, 3rd ed. John Wiley & Sons, New York, NY.

Congedo, L., 2016. Semi-Automatic Classification Plugin Documentation. Release 6.0.1.1. <https://doi.org/10.13140/RG.2.2.29474.02242/1>

Corbett, T., O'Faircheallaigh, C., Regan, A., 2017. 'Designated areas' and the regulation of artisanal and small-scale mining. *Land use policy* 68, 393–401. <https://doi.org/10.1016/j.landusepol.2017.08.004>

Coulter, L.L., Stow, D.A., Tsai, Y.H., Ibanez, N., Shih, H. Chien, Kerr, A., Benza, M., Weeks, J.R., Mensah, F., 2016. Classification and assessment of land cover and land use change in southern Ghana using dense stacks of Landsat 7 ETM+ imagery. *Remote Sens. Environ.* 184, 396–409.

<https://doi.org/10.1016/j.rse.2016.07.016>

Dekker, F., 2019. Cocoa not main cause of deforestation in Ghana [WWW Document]. *Satelligence*. URL <https://satelligence.com/news/2019/5/17/cocoa-not-main-cause-of-deforestation-in-ghana> (accessed 12.4.19).

Dietterich T.G. (2000) Ensemble Methods in Machine Learning. In: Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science, vol 1857. Springer, Berlin, Heidelberg

DigitalGlobe, 2015. World Imagery Basemap.

Drusch, M., Bello, U. Del, Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* 120, 25–36. <https://doi.org/https://doi.org/10.1016/j.rse.2011.11.026>

Espejo, J.C., Messinger, M., Román-Dañobeytia, F., Ascorra, C., Fernandez, L.E., Silman, M., 2018. Deforestation and forest degradation due to gold mining in the Peruvian Amazon: A 34-year perspective. *Remote Sens.* 10, 1–17. <https://doi.org/10.3390/rs10121903>

European Space Agency, 2015. Sentinel-2 User Handbook, ESA documentation. [WWW Document] URL https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

European Space Agency, 2019. Level-2A. Prod. Types User Guid. [WWW Document] URL <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a> (accessed 12.5.19).

Forestry Commission, 2010. IUCN, Switzerland. Shapefiles of Forest and Game Reserves in Ghana 2010 [WWW Document] URL <https://data.gov.gh/dataset/shapefiles-forest-and-game-reserves-ghana-2010> (accessed 11.9.19)

Forkuor, G., Benewinde Zoungrana, J.B., Dimobe, K., Ouattara, B., Vadrevu, K.P., Tondoh, J.E., 2020. Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - A case study. *Remote Sens. Environ.* 236, 111496. <https://doi.org/10.1016/j.rse.2019.111496>

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. Intergovernmental Forum on Mining, Minerals, Metals and Sustainable Development (IGF). Winnipeg: IISD.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., et al., 2017. Copernicus Sentinel-2A calibration and products validation status. *Remote Sens.* 9. <https://doi.org/10.3390/rs9060584>

Géron, A., 2017. *Hands-On Machine Learning with Scikit-Learn & Tensorflow*, 1st ed. O'Reilly, Sebastopol.

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

Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) Digital Object Identifier (DOI) number: /10.5066/F7J38R2N

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. *Pattern Recognit.* 77, 354–377. <https://doi.org/10.1016/j.patcog.2017.10.013>

Hilson, G., 2001. A contextual review of the Ghanaian small-scale mining industry. *Mining, Miner. Sustain. Dev.* 76.

Hilson, G., 2002. The future of small-scale mining: environmental and socioeconomic perspectives. *Futures* 34, 863–872. [https://doi.org/10.1016/S0016-3287\(02\)00044-7](https://doi.org/10.1016/S0016-3287(02)00044-7)

Hilson, G., Gatsinzi, A., 2014. A rocky road ahead? Critical reflections on the futures of small-scale mining in sub-Saharan Africa. *Futures* 62, 1–9. <https://doi.org/10.1016/j.futures.2014.05.006>

Hilson, G., Hilson, A., Adu-Darko, E., 2014. Chinese participation in Ghana's informal gold mining economy: Drivers, implications and clarifications. *J. Rural Stud.* 34, 292–303. <https://doi.org/10.1016/j.jrurstud.2014.03.001>

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. <https://doi.org/10.1016/j.geoforum.2019.09.006>

Hinton, G.E., 2012. A Practical Guide to Training Restricted Boltzmann Machines, in: Montavon, G., Orr, G.B., Müller, K.-R. (Eds.), *Neural Networks: Tricks of the Trade*. Springer, Berlin.

- Huang, C., Li, Y., Loy, C.C., Tang, X., 2016. Learning Deep Representation for Imbalanced Classification, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5375–5384. <https://doi.org/10.1109/CVPR.2016.580>
- Huang, Z., Cheng, G., Wang, H., Li, H., Shi, L., Pan, C., 2016. Building extraction from multi-source remote sensing images via deep deconvolution neural networks. *Int. Geosci. Remote Sens. Symp.* 2016-Novem, 1835–1838. <https://doi.org/10.1109/IGARSS.2016.7729471>
- Iglovikov, V., Mushinskiy, S., Osin, V., 2017. Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition. arXiv:1706.06169
- Kelly, M., Blanchard, S.D., Kersten, E., Koy, K., 2011. Terrestrial Remotely Sensed Imagery in Support of Public Health: New Avenues of Research Using Object-Based Image Analysis. *Remote Sens.* 3, 2321–2345. <https://doi.org/10.3390/rs3112321>
- Kemker, R., Salvaggio, C., Kanan, C., 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS J. Photogramm. Remote Sens.* 145, 60–77. <https://doi.org/10.1016/j.isprsjprs.2018.04.014>
- Kroupi, E., Kesa, M., Navarro-Sánchez, V.D., Saeed, S., Pelloquin, C., Alhaddad, B., Moreno, L., Soria-Frisch, A., Ruffini, G., 2019. Deep convolutional neural networks for land-cover classification with Sentinel-2 images. *J. Appl. Remote Sens.* 13, 1. <https://doi.org/10.1117/1.jrs.13.024525>
- Kusimi, J.M., 2008. Assessing land use and land cover change in the Wassa West District of Ghana using remote sensing. *GeoJournal* 71, 249–259. <https://doi.org/10.1007/s10708-008-9172-6>
- Kwadwo, O., 2019. The effects of 2015 El Nino on smallholder maize production in the transitional ecological zone of Ghana. *Int. J. Clim. Chang. Strateg. Manag.* 11, 609–621. <https://doi.org/10.1108/IJCCSM-02-2018-0014>
- Labou, I., Benoit, M., Baratoux, L., Grégoire, M., Ndiaye, P.M., Thebaud, N., Béziat, D., Debat, P., 2020. Petrological and geochemical study of Birimian ultramafic rocks within the West African Craton: Insights from Mako (Senegal) and Loraboué (Burkina Faso) lherzolite/harzburgerite/wehrliche associations. *J. African Earth Sci.* 162, 103677. <https://doi.org/10.1016/j.jafrearsci.2019.103677>
- LeCun, Y., Boser, B., Denker, J., Henderson, D., 1990. Handwritten digit recognition with a back-propagation network. *Adv. Neural Inf. Process. Syst.* 2. D. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 396–404.

- Leroux, L., Congedo, L., Bellón, B., Gaetano, R., Bégué, A., 2018. Land Cover Mapping Using Sentinel-2 Images and the Semi-Automatic Classification Plugin: A Northern Burkina Faso Case Study, in: QGIS and Applications in Agriculture and Forest. John Wiley & Sons, Ltd, pp. 119–151. <https://doi.org/10.1002/9781119457107.ch4>
- Lobo, F. de L., Souza-Filho, P.W.M., Novo, E.M.L. de M., Carlos, F.M., Barbosa, C.C.F., 2018. Mapping mining areas in the Brazilian amazon using MSI/Sentinel-2 imagery (2017). *Remote Sens.* 10. <https://doi.org/10.3390/rs10081178>
- MathWorks, 2019. Semantic Segmentation of Multispectral Images Using Deep Learning [WWW Document]. R2019b Doc. Examples. URL <https://uk.mathworks.com/help/images/multispectral-semantic-segmentation-using-deep-learning.html> (accessed 12.10.19).
- McQuilken, J. and Garvin, H., 2016. Artisanal and small-scale gold mining in Ghana. Evidence to inform an ‘action dialogue,’ IIED, London. <https://doi.org/10.13140/RG.2.2.36435.99368>
- Ministry of Lands and Natural Resources, 2019. Government committed to the fight against illegal mining [WWW Document]. Online Artic. URL <http://mlnr.gov.gh/index.php/government-committed-to-the-fight-against-illegal-mining/> (accessed 2.9.20).
- Nogueira, K., Penatti, O.A.B.B., dos Santos, J.A., Santos, J.A. dos, 2017. Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification. *Pattern Recognit.* 61, 539–556. <https://doi.org/10.1016/j.patcog.2016.07.001>
- Nuijten, R.J.G., Kooistra, L., De Deyn, G.B., 2019. Using Unmanned Aerial Systems (UAS) and Object-Based Image Analysis (OBIA) for Measuring Plant-Soil Feedback Effects on Crop Productivity. *Drones* 3, 54. <https://doi.org/10.3390/drones3030054>
- Obodai, J., Adjei, K.A., Odai, S.N., Lumor, M., 2019. Land use/land cover dynamics using Landsat data in a gold mining basin-the Ankobra, Ghana. *Remote Sens. Appl. Soc. Environ.* 13, 247–256. <https://doi.org/10.1016/j.rsase.2018.10.007>
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S. V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148, 42–57. <https://doi.org/10.1016/j.rse.2014.02.015>
- Owusu-Nimo, F., Mantey, J., Nyarko, K.B., Appiah-Effah, E., Aubynn, A., 2018. Spatial distribution patterns of illegal artisanal small scale gold mining (Galamsey) operations in Ghana: A focus on the Western Region. *Heliyon*, 4. <https://doi.org/10.1016/j.heliyon.2018.e00534>

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peng, D., Zhang, Y., Guan, H., 2019. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sens.* 11. <https://doi.org/10.3390/rs11111382>
- Rajae, M., Obiri, S., Green, A., Long, R., Cobbina, S.J., Nartey, V., Buck, D., Antwi, E., Basu, N., 2015. Integrated Assessment of Artisanal and Small-Scale Gold Mining In Ghana—Part 2: Natural Sciences Review. *Int. J. Environ. Res. Public Health* 12, 8971–9011. <https://doi.org/10.3390/ijerph120808971>
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 9351, 234–241.
- Schuegraf, P., Bittner, K., 2019. Automatic building footprint extraction from multi-resolution remote sensing images using a hybrid FCN. *ISPRS Int. J. Geo-Information* 8, 1–16. <https://doi.org/10.3390/ijgi8040191>
- Shendryk, Y., Rist, Y., Ticehurst, C., Thorburn, P., 2019. Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* 157, 124–136. <https://doi.org/10.1016/j.isprsjprs.2019.08.018>
- Signoroni, A., Savardi, M., Baronio, A., Benini, S., 2019. Deep learning meets hyperspectral image analysis: A multidisciplinary review. *J. Imaging* 5. <https://doi.org/10.3390/jimaging5050052>
- Snapir, B., Simms, D.M., Waine, T.W., 2017. Mapping the expansion of galamsey gold mines in the cocoa growing area of Ghana using optical remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 58, 225–233. <https://doi.org/10.1016/j.jag.2017.02.009>
- Souza-Filho, P.W.M., Nascimento, W.R., Santos, D.C., Weber, E.J., Silva, R.O., Siqueira, J.O., 2018. A GEOBIA approach for multitemporal land-cover and land-use change analysis in a tropical watershed in the southeastern Amazon. *Remote Sens.* 10. <https://doi.org/10.3390/rs10111683>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.

Telmer, K.H., Stapper, D., 2007. UNIDO Project EG/GLO/01/G34 Final Report Evaluating and Monitoring Small Scale Gold Mining and Mercury Use: Building a Knowledge-base with Satellite Imagery and Field Work 49.

van der Meer, F.D., van der Werff, H.M.A., van Ruitenbeek, F.J.A., 2014. Potential of ESA's Sentinel-2 for geological applications. *Remote Sens. Environ.* 148, 124–133.
<https://doi.org/https://doi.org/10.1016/j.rse.2014.03.022>

Vaiopoulos, A.D., Karantzalos, K., 2016. Pansharpening on the narrow VNIR and SWIR spectral bands of Sentinel-2. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* 41, 723–730.
<https://doi.org/10.5194/isprsarchives-XLI-B7-723-2016>

Wang, C., Zhao, Z., Ren, Q., Xu, Y., Yu, Y., 2019. Dense U-net based on patch-based learning for retinal vessel segmentation. *Entropy* 21, 1–15. <https://doi.org/10.3390/e21020168>

Weisse, M., Goldman, E.D., 2019. The World Lost a Belgium-sized Area of Primary Rainforests Last Year [WWW Document]. *World Resour. Inst.* URL https://www.wri.org/blog/2019/04/world-lost-belgium-sized-area-primary-rainforests-last-year?utm_campaign=GFW&source=socialmediakit&utm_medium=gfwsocial&utm_term=2018tcl_4_2019 (accessed 12.4.19).

Wilson, A.C., Roelofs, R., Stern, M., Srebro, N., Recht, B., 2017. The Marginal Value of Adaptive Gradient Methods in Machine Learning, in: Guyon, I., Luxburg, U. V, Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems* 30. Curran Associates, Inc., pp. 4148–4158.

Wurm, M., Stark, T., Zhu, X.X., Weigand, M., Taubenböck, H., 2019. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* 150, 59–69. <https://doi.org/10.1016/j.isprsjprs.2019.02.006>

Zhang, X., Xiao, Z., Li, D., Fan, M., Zhao, L., 2019. Semantic Segmentation of Remote Sensing Images Using Multiscale Decoding Network. *IEEE Geosci. Remote Sens. Lett.* 16, 1492–1496.
<https://doi.org/10.1109/lgrs.2019.2901592>

Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., Zhang, D., 2019. Use of Unmanned Aerial Vehicle Imagery and Deep Learning UNet to Extract Rice Lodging. *Sensors* 19, 3859.
<https://doi.org/10.3390/s19183859>

List of Figure Captions

Fig. 1. Overview map of study area, Sentinel-2 tile coverage and training data boundary. Righthand sub-image shows the ten districts common to the Owusu-Nimo et al. (2018) study.

Fig. 2. Locations of the manually digitised training and validation areas. Dates refer to the acquisition date of the imagery, the training area boundary is identical for all years, the offset is used to indicate temporal change.

Fig. 3. Schematic of the U-Net architecture used for the model. The input is a 10 channel multispectral image of 256 x 256 pixels and the outputs include: a 1 channel prediction mask, a 1 channel prediction probability layer for the highest class probability and a 3 channel prediction probability layer for all classes, all at the same resolution as the input image.

Fig. 4. Model outputs showing a) the Sentinel-2 RGB data as a reference, b) the single predicted class labels, c) the model's confidence in its predictions and d) the prediction probabilities for every class mapped to the RGB colour channels

Fig 5. Qualitative view of the results from assessment two. Reference information for all five unseen test tiles is displayed alongside the corresponding prediction masks from the various machine learning classifiers.

Fig. 6. Prediction map results displayed for the first year of the study period. The main image shows the full study area at the end of the first temporal period. The city of Kumasi is the large area of built up land in the north and the outskirts of Accra are visible in the far south east. The upper sub-image shows the district of Wassa Amenfi East with both the winter 2015-16 baseline data and the 2016 changes. The lower sub-image displays a close-up of some of the changes within Wassa Amenfi East detected during 2016.

Fig. 7. Total land cover changes per class observed over the entire study area. Y-axis scale is constant between sub-graphs.

Fig. 8. Cumulative mined area changes a) and yearly mined area changes b) measured across ten Western districts.

Fig. 9. Kernel density heatmap showing the concentrations of new ASM activity over the study area over the four years. The areas surrounding Accra in the south east are related to sand mining and quarrying, these classifications are outside the scope of the original training data and are provided for visual reference only.

Fig. 10. Maps showing the hectares of mining within Ghana's national protected forests. The inset shows expansion within the Upper Wassaw reserve. This reserve has the highest rate of mining deforestation at over 800 new hectares lost since 2015. Forest reserve boundaries from Forestry Commission (2010), background GMTED2010 from U.S. Geological Survey.