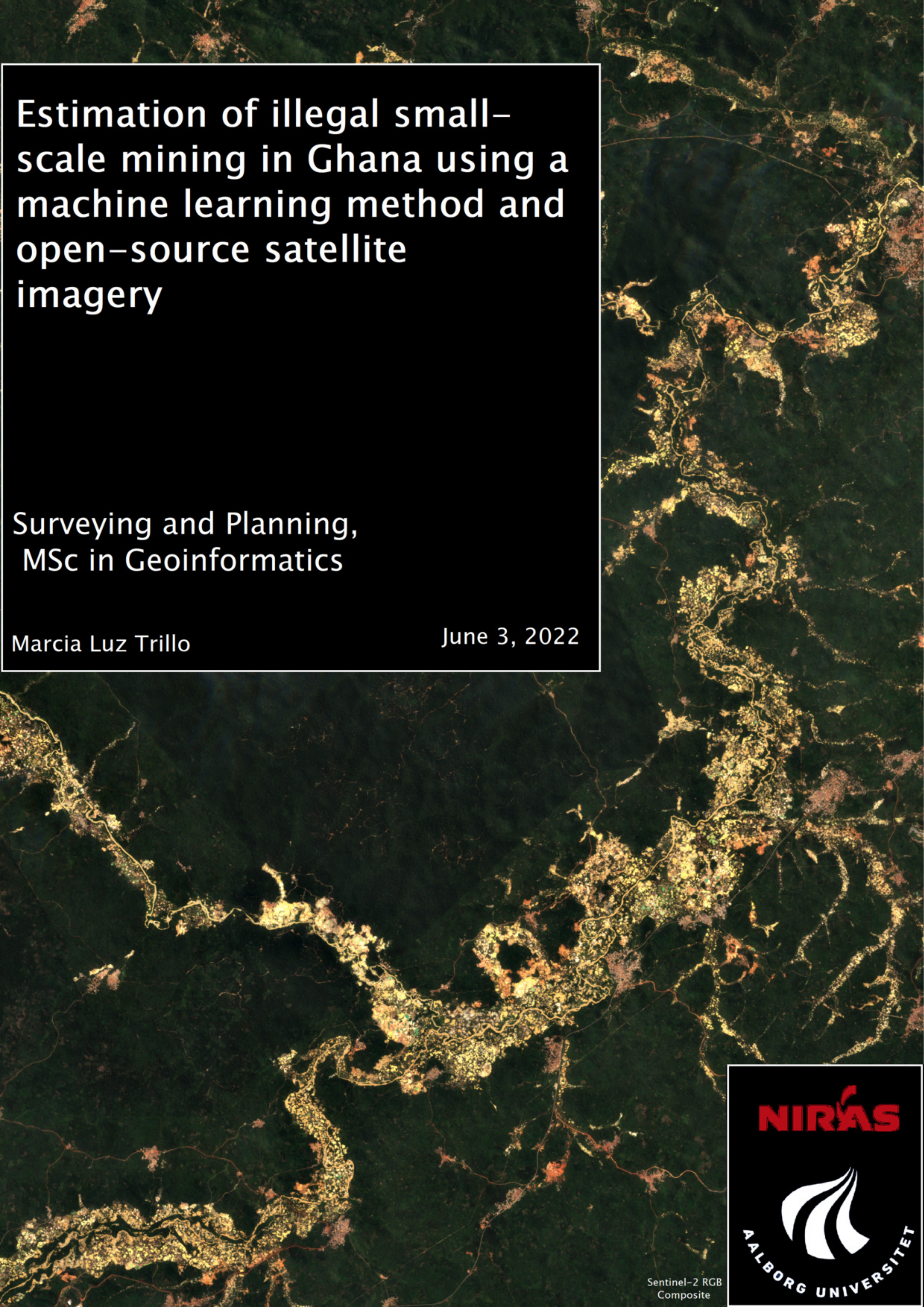
# Estimation of illegal small-scale mining in Ghana using a machine learning method and open-source satellite imagery

Surveying and Planning,
MSc in Geoinformatics

Marcia Luz Trillo

June 3, 2022



Sentinel-2 RGB Composite

## Abstract

Ghana is considered one of the largest gold-producing countries in the world and is ranked first in the African continent. This nation produces vast quantities of gold through Galamsey (illegal small-scale gold mining), which is a very popular but unregulated technique for mineral extraction in southern Ghana and is the main source of income for many Ghanaians. Over the past decade, Galamsey has grown tremendously, causing a concerning and noticeable degradation of the environment and posing a real threat to people's lives. Attempts have been made to detect and map these illicit mining operations in the past, but a high-quality map identifying the distribution patterns of Galamsey has not yet been conducted. In this study, a highly detailed Galamsey identification map of the entire country of Ghana was produced using an image recognition deep learning method [1]. More specifically, the regression ML approach to detect Galamsey features followed a Convolutional Neural Network (CNN) algorithm with an Inception- ResNet -like architecture. Predictions were computed using the current best available resolution for open-source satellite images (10m), and Sentinel-1 and -2 products were processed to train the ML model for a three months period (November 1- 2021 to February 1- 2022). Even though further studies should include more algorithm testing, this pixel-based method delivered good results, achieving a binary accuracy of nearly 90%. Model predictions have shown that illegal mining is concentrated in four main regions in Ghana, being Western Ghana the hotspot for unlicensed artisanal miners. Galamsey spatial distribution is characterized by clusters along the ramification of main water streams of the country, degrading forest reserves that are protected at a national level. The ML method presented in this study serves as a valuable tool for identifying unauthorized gold mining activities and is therefore of considerable significance for government decision-makers and stakeholders that are involved in law-making practices against Galamsey.

## Preface

This thesis was written in partial fulfillment of the requirements for the degree of Surveying and Planning, MSc in Technology: Geoinformatics, at Aalborg University CPH, Denmark.

This thesis aims to predict illegal small-scale mining in Ghana, using a Convolutional Neural Network method and open-source Sentinel imagery.

I would like to thank my supervisor Jamal Jokar Arsanjani for his guidance throughout this process. My colleagues at NIRAS in the AREA department also deserve my acknowledgments, in particular Casper Fibæk. I am extremely thankful for his support and encouragement which helped me grow as a professional.

Finally, I would like to thank my family. I feel very close to them, even though I have not seen them in a long time. It was only with their unconditional love and support that I have come this far.

## Table of Contents

## List of Abbreviations

| | |
|---|---|
| ACC | Accuracy |
| ANN | Artificial Neural Networks |
| ASM | Artisanal Small Scale Mining |
| BACC | Balanced Accuracy |
| CNN | Convolutional Neural Networks |
| DL | Deep Learning |
| ESA | European Space Agency |
| EU | European Union |
| LSM | Large Scale Mining |
| MAE | Mean Absolute Error |
| ML | Machine Learning |
| MSE | Mean Squared Error |
| NDMI | Normalized Difference Moisture Index |
| NDVI | Normalized Difference Vegetation Index |
| PREC | Precision |
| REC | Recall |
| RMSE | Root Mean Square Error |
| SDGs | Sustainable Development Goals |
| SDSS | Spatial Decision Support Systems |
| SSM | Small Scale Mining |

## List of Figures

**List of Tables**



**List of Workflows**

# 1. Introduction

## 1.1. Galamsey in Ghana

Ghana is currently considered the 6th largest gold producer in the world and the continental leader of gold production in Africa, surpassing South Africa. According to the World Gold Council, Ghana is the only African country that has reached the top 10 leading gold countries in the world in 2020, as it manufactured approximately 139 tonnes of gold. [2][3]

Gold reserves in Ghana are mainly located in the Southern part of the country, and there are two predominantly gold mining activities: Large-Scale Mining (LSM) and Small-Scale or Artisanal Small-Scale Mining (SSM or ASM). The latter is characterized by low mineral production levels, minimal capital investment, and is limited to 25 acres. Although a mining license from the Minerals Commission of Ghana [4] is required for small-scale mining, the lack of enforcement has led to the proliferation of illegal activities [5].

"Galamsey" is a popular local Ghanaian term derived from the phrase "gather them and sell" and describes collecting minerals from the soil surface in order to then sell them illegally. [6]. "Galamsey" falls into the category of small-scale mining and "Galamseyers" are the workers who perform this illegal activity in this sub- Saharan African country. They are mainly involved in labor-driven activities such as digging small tunnels and working pits to extract gold. Although these small-scale miners should first get approval from the corresponding government authorities (Water Resources Commission [7], the Minerals Commission of Ghana [4], Environmental Protection Agency [8], and the Forestry Commission [9], among others) they work without regulatory licenses, avoid taxes, and are mainly located in government-protected zones (such as forest reserves) and culturally sensitive areas or residential districts.

In spite of the fact that illegal mining has become a major source of income for local residents, these illicit mining operations have negatively affected the environment [10], agricultural productivity [11][12], human health [13], and safety of the local communities [14]. In most cases, rivers are the main source of drinking water for mining local communities and they are heavily polluted because Galamseyers need to be near water to perform artisanal gold mining. However, they do not have the necessary resources to avoid environmental degradation, and pollutants such as Mercury (Hg) are present in water bodies that are utilized by the local communities. Land degradation and deforestation are also negative impacts of Galamsey, leading to biodiversity loss and contributing to climate change variation [15]. It is not uncommon to find abandoned and flooded mining pits, deforested lands, and garbage piles near the ASM areas [16]. This poses a life-threatening situation to the local communities, as the abandoned pits have a high risk of collapsing and a lot of Galamseyers have died in an attempt to dig out gold deposits [17] [18].

Poor regulatory enforcement around artisanal mining areas, the lack of job opportunities in Ghana, and the absence of education in environmental protection have led to the presence of this unmonitored activity, which poses a real threat to the country's mining sector and people's lives [13].

Although the stakeholders involved in the mining sector have tried to regulate these illegal operations, there is still not enough information about the type and scale of Galamsey areas.

Despite industrial mines covering more area than Galamsey individual activities, the illegal artisanal mining footprint is bigger in South West Ghana (which is a part of the country that has national forest reserves and major water bodies) [19].

Although the government authorities of Ghana have tried to have more control over these illegal activities in the past years, Galamsey sites are very difficult to find. These small gold mines are generally located in remote areas with poor road networks, which makes it challenging for the local security agencies to access. Furthermore, Galamseyers work in densely vegetated areas such as forest reserves, and these small-scale mines are in most cases just a few acres big. These small-scale miners also work in abandoned underground tunnels and land-locked areas, which makes it difficult for the law enforcement authorities to track [20]. There are two types of Galamsey ('panning' and 'selection') that are even harder to pin down, due to their inconsistent nature of constantly changing their working environment [21].

Not only these clandestine sites are arduous to locate for local security authorities, but they also represent a major challenge in the remote sensing field. Differentiating between legal and illegal mining sites can be very challenging, as they have similar spectral signatures and there is generally a high concentration of Galamsey near large-scale gold mining companies, as well as in urban centers. While the majority of Galamsey sites are characterized by clusters, some other ones are more scattered in space and play a more stand-alone role [20].

Surveying small-scale mines requires a big effort, due to the labor-driven and time-consuming tasks involved. What is more, Galamsey patterns are constantly changing in space and time, making it difficult for surveyors to track. As a consequence, little is known about the short and long-term cumulative impact of illicit mining activities in Ghana. Earth observation technology helps to identify complex features and to monitor some of the United Nations (UN) Sustainable Development Goals (SDGs) [22]. Clandestine mining detection in Western Africa is of special interest as it interferes with complying with some of the SDGs, such as including zero hunger, good health and well-being, and clean water and sanitation.

Although manual digitization could reach high levels of accuracy, it is not optimal for large-scale projects as it is considered to be costly and time-consuming. Despite conducting thorough research in the past years [21][19][20][23] [24], a detailed digital map of Galamsey activities using the current best available resolution for free satellite imagery has not yet been completely undertaken. Thus, this study aims to estimate Galamsey footprints for the period Nov 1, 2021- Feb 1, 2022 and to create a high-resolution digital map of Galamsey for the entire country of Ghana. A machine learning algorithm (i.e. a Convolutional Neural Network (CNN) regression approach based on image recognition) will be implemented, using freely available and open-source satellite images [25].

### 1.2. Deep learning

Machine learning is a type of Artificial Intelligence (AI) that enables computing systems to continuously “learn” from data and gradually improve (in terms of accuracy), developing pattern recognition and achieving automation. Deep learning (DL) is a part of machine learning that works with Artificial Neural Networks (ANN) where complex algorithms are executed to try to “imitate the human brain”. Different disciplines have been using ANN technology in a wide variety of applications, from speech recognition to DNA mutation [42], [43]. In remote sensing, image recognition has been extremely beneficial in different fields: from addressing environmental degradation [43] to estimating population growth [44].

Convolutional Neural Networks (CNNs or ConVet) fall into the category of DL, and they are widely used in image recognition and computer-based visualization applications. These complex algorithms allow computers to extract valuable information from digital images (such as satellite products) or videos. Machine learning CNNs take a set of images as input and “learn” from them, assigning importance to objects, to then make predictions. The ConVet architecture is characterized by different stages [42]. These stages contain convolutional layers, pooling layers, and fully connected layers organized in feature maps. The main objective of CNNs is to extract features from an image, to then predict the specified feature from a new feature map. The convolutional layer is crucial in the network system, as it is involved with most of the computational processes. In this stage, a ‘kernel’ performs a dot product between matrices, to extract ‘features’ from a set of images. A kernel, spatially smaller than the input image, is basically an array of weights, that multiplied by the input generates an “enhanced” output (known as a ‘feature map’). Since the output does not need to connect directly to each input pixel, convolutional layers are ‘partially connected layers’. After completing this convolution operation, a non-linear Rectified Linear Unit (ReLU) transformation is included into the model [45]. The training performance of the model (or the kernels) is monitored by the loss function. As the model trains, the loss values should improve, reaching a point where the difference between the input (ground truth data) and the output product is minimal. This optimization is generally achieved with gradient descent and backpropagation. A pooling layer then introduces another filter to the output generated by the convolutional layer, but this time instead of having weights, the filter is an aggregation function, which helps with overfitting. This step can be characterized by ‘Max pooling’ (the filter populates the output array with the maximum pixel value of the input) or ‘Average pooling’ (where the filter computes the average value of the input pixels). Finally, the fully-connected layer (meaning that each node in the output layer is linked to a node in the previous layer) maps the features that were created in the convolutional and pooling layers using a classifying function, which is generally an activation function.

In remote sensing applications, deep learning methods such as CNNs have been widely used to understand different image recognition-related topics with a high level of accuracy [46], such as land cover classification using aerial photographs [47]. If the satellite images that are being processed are not characterized by a very high resolution, extracting patches instead of single pixels is a better approach as medium resolution images do not have a high level of data granularity [48].

Deep learning models have been already implemented in Ghana to address financial inclusion and other related geospatial SDGs [49]. In this study, a CNN-based regression method was the approach used to identify mining activities in Ghana.

### 1.3. NIRAS

This research was carried out with the help of NIRAS, in partial fulfillment of the degree of Master of Science in Technology at Aalborg University[26]. NIRAS is an engineering consultancy company that has offices in many countries, including Denmark, England, Belgium, and Australia, among others. This firm conducts work in a wide range of disciplines such as climate, urban development, water supply, and GIS and remote sensing.

NIRAS' mission is to secure sustainable progress. This private firm has integrated the United Nations SDGs into its engineering strategies. This vision was key when conducting this study, as this illegal mining identification in Ghana could contribute to achieving some of these global sustainability objectives.

This work is supported by the NIRAS Mapping and GIS (MAGI) area, which is part of the 'Data, Analytics, and Planning' (AREA) department. MAGI leads different geospatial consulting services such as building modeling, remote sensing, and GIS mapping. The MAGI Head of Department is Laurids Rolighed Larsen, and Buteo Toolbox [1] was developed by Casper Samsø Fibæk, an industrial PhD candidate at Aalborg University and colleague at NIRAS.

## 2. Tools and technologies

QGIS [27]

QGIS is a freely available software that stands for Quantum Geographic Information Systems. It is widely used in remote sensing and geospatial-related fields. A license is not required for installation, which makes it easy for any user to execute. What is good about QGIS is that it has a wide variety of raster and vector tools, as well as several plug-ins that can be easily installed and executed for different mapping-related tasks. What is more, any user can add more functionalities by developing new plug-ins, that can be shared with the entire QGIS community. QGIS is better at data processing times and it can handle multiple large datasets simultaneously when comparing it with other geospatial technologies of the same kind. Some useful plug-ins were downloaded for the aims of this study, such as the 'HCMGIS' plug-in [28]which contains different satellite base maps (which were very beneficial when it came to spotting the mining sites in Ghana). In this project, QGIS was mostly utilized for the vector pre-processing steps such as the creation of the model's training sites and the digitization of the LSM and ASM mining areas. QGIS was also used for raster visualization and for custom map creation.

GitHub [29]

GitHub is a non-profit and cloud-based website repository hosting service mainly used for software development. Users can store their code and track and make changes. Users can also have multiple personal or shared repositories, which allows them to cooperatively and simultaneously edit scripts. GitHub is built on Version Control, which helps developers safely work through tracking and managing changes to a target's repository, and on Git, a freely available distributed version control system developed by Linus Torvalds. GitHub has become very popular in software programming since it was founded in 2008, as its interface is user-friendly and anyone can create an account and host a public repository at no cost. GitHub is used in this project to host Buteo Toolbox [1].

Buteo Toolbox [1]

Buteo Toolbox is a non-profit repository hosted on GitHub [29] that was used in this project to identify the spatial coverage of mining activities in Ghana. This Python-based toolbox was developed by Casper Fibæk, and is described as a 'series of modules that ease the creation of Earth Observation Driven Spatial Decision Support Systems (SDSS)'. SDSS are computer-based geospatial technologies that contain both spatial and conventional referenced data. SDSS are very helpful for decision-makers, as they assist organizations to make important land use-related decisions.

Buteo Toolbox has multiple folders that serve different purposes. Each folder contains Python scripts with a set of modules to process many types of datasets. These folders are split into vector and raster formats, as well as into earth observation and artificial intelligence sections. Buteo Toolbox makes it easy for the user when it comes to data processing examples. This open-source repository contains some Python files that were used to execute different geospatial tasks, from downloading and processing satellite data to training a Convolutional Neural Network (CNN) model.

Python [30]

Buteo Toolbox [1] was written in Python. Python is a widely known open-source programming language that can be used for many different applications. Due to the availability of free resources to learn this Python, it has become very popular. It's based on English syntax and it's considered to be very straightforward to learn. It includes several third-party modules that are beneficial in a wide range of fields such as data science. Python has an object-oriented approach, it has libraries that are easy to import and execute, and it can be used for small tasks (such as simple data processing), or large-scale projects (such as executing complex codes for machine learning or software development). A big advantage of Python is that it has a huge community, which makes it easier for the user to solve programming issues, as codes and tutorials are openly shared.

TensorFlow [31]

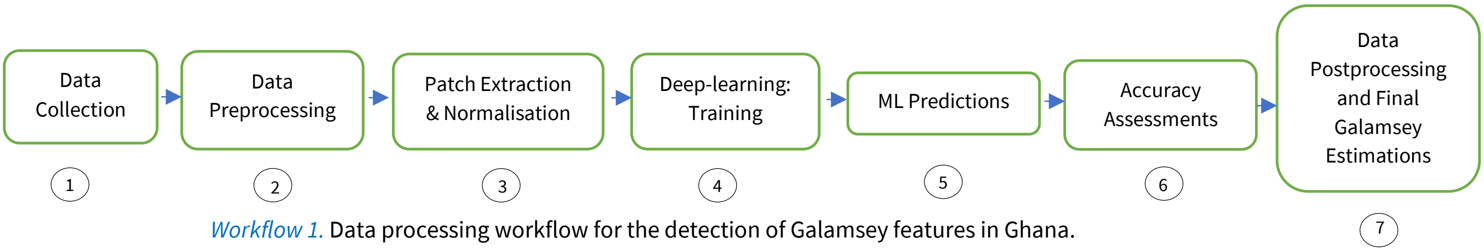
The CNN model that was used for predicting mining coverage in Ghana is TensorFlow-based, using an image recognition approach. TensorFlow is an open-source and Python-friendly library created by Google [32] that is used for artificial intelligence and machine learning applications. Nodes and tensors in TensorFlow are Python objects. Because TensorFlow is built on a flexible architecture, it can be run on different platforms such as local machines, a cluster in the cloud, GPUs or CPUs, and other devices. TensorFlow contains different useful deep learning tools, libraries, and an enormous community that can help build and deploy Machine Learning (ML) applications. TensorFlow data inputs are in the format of multi-dimensional arrays, that are called tensors. After the data is fed into the neural network system, computations are executed in the form of graphs. TensorFlow supports GPUs computing devices, hence a NIRAS eGPU was used to allow better processing times, especially in the CNN model training and making predictions map.

Visual Studio Code (VSC) [33]

Visual Studio Code was preferred for the code editing part of this study. This includes debugging, tracking errors, and committing and merging changes (as VSC can be linked with GitHub, allowing the user the possibility of having direct interaction with the target repository). VSC is an open-source project developed by Microsoft [34] and it can be used in Windows, Linux, and macOS operating systems. Due to its easy-to-follow syntax highlighting with different colors, it is not complex for the user to identify variables and special characters. VSC supports several programming languages such as Python, JavaScript, CSS, and HTML, among others.

## 3. Galamsey Analysis Workflow

In order to detect Galamsey areas in Ghana, a thorough analysis consisting of seven main steps was followed *(Workflow 1)*.




*Workflow 1.* Data processing workflow for the detection of Galamsey features in Ghana.

### 3.1. Data Collection

#### 3.1.1. Vector Data

OpenStreetMap (OSM) [35]

OSM was utilized to retrieve the main water bodies, waterways, and roads of Ghana. The open water features are of special interest, as Galamseyers need water in the gold mining process, hence and there is more chance to find ASM sites around rivers. Quarries (open-pit mines) were also downloaded to help in the mining digitization process, but they were scattered in space and not properly mapped, so this vector layer was partly disregarded.

OSM is an open-source and collaborative project where any user can edit and add geographic items to a world map. Although OSM data might have some degree of bias and there's no quality check procedure, sometimes the level of detail is better than official datasets from the government or commercial organizations. In addition, OSM is free to use and easy to download through the QuickOSM plug-in [36] in QGIS, or the Geofabrik Download Server [37]. OSM is updated by volunteers/users in real-time and it offers high-resolution geographic vector data.

Google Maps [38]

Although an official list of operating mines was retrieved from the Minerals Commission of Ghana [4], their exact coordinates were not available. Hence, Google Maps was very beneficial when it came to locating these licensed companies with their labels. Google

Maps is a web-based service developed by Google [32]. Google maps can be used to get directions, retrieve businesses' information, and check public transportation routes, among other helpful functionalities. It's very easy to navigate through, and it can be installed on mobile devices or accessed through a web browser. It offers visualization not only in a vector format but the user can also look at a map through satellite images and aerial photography. Google Satellite products and Google Maps combined were advantageous when identifying Galamsey activities, as in some cases there are more chances to find illegal mining activity close to declared and mapped legal mining infrastructure. Although Google Satellite was not updated in some areas of the country, it helped visualize the main mining areas.

Minerals Commission of Ghana [4]

The Minerals Commission of Ghana is a Government agency in charge of the regulation and management of the mineral resources in Ghana. This organization is also involved in the policy-making process and is responsible for the implementation and monitoring of mining policies. This government entity works with stakeholders to make sure that sustainable development is achieved through mining regulations. Small-Scale Mining or Artisanal Mining stimulates the local economy and contributes to the socio-economic development of the country, as the government of Ghana employs about one million small-scale miners. A Small-Scale Mining License is only emitted to Ghanaians who are over 18 years old, and it allows them to mine up to 25 acres in designated areas [39]. A list of major operating mines in Ghana was downloaded from the Minerals Commission of Ghana's official website [40]*(Figure 1A)*. This dataset helped create a points layer of the location of the large-scale mining companies in the country. In addition, a map of all mining licenses in Ghana was accessed through the Ghana Mining Repository website [41] and exported as a vector file. This polygon dataset was crucial in the postprocessing part of the analysis, as it contained an attribute table with the active licenses in the country *(Figure 1B)*.

Galamsey Spatial Distribution in the Western Region Dataset

An excel spreadsheet consisting of 868 Galamsey sightings was provided by Owusu-Nimo et. al [20]. This dataset contains the survey dates, the point coordinates of the surveys, the Galamsey operational type, and the operational status, among other important characteristics. Given the lack of geospatial data in Ghana, this dataset was crucial in recognizing Galamsey agglomerations when digitizing sites for the model training *(Figure 2)*.

### 3.1.2. Raster Data

Copernicus [25]is a European Union (EU) [51] initiative for developing satellite-based products with Earth Observation data. The implementation of Copernicus is in charge of the European Commission (EC) [51] which is supported by the European Space Agency (ESA) [52]. Copernicus makes its products available to all citizens and organizations across the globe, on an open-source, user-friendly, and free-of-charge platform. Due to Sentinel's open data policy and worldwide coverage, these satellite products are widely used for a variety of multi-disciplinary earth observation applications.

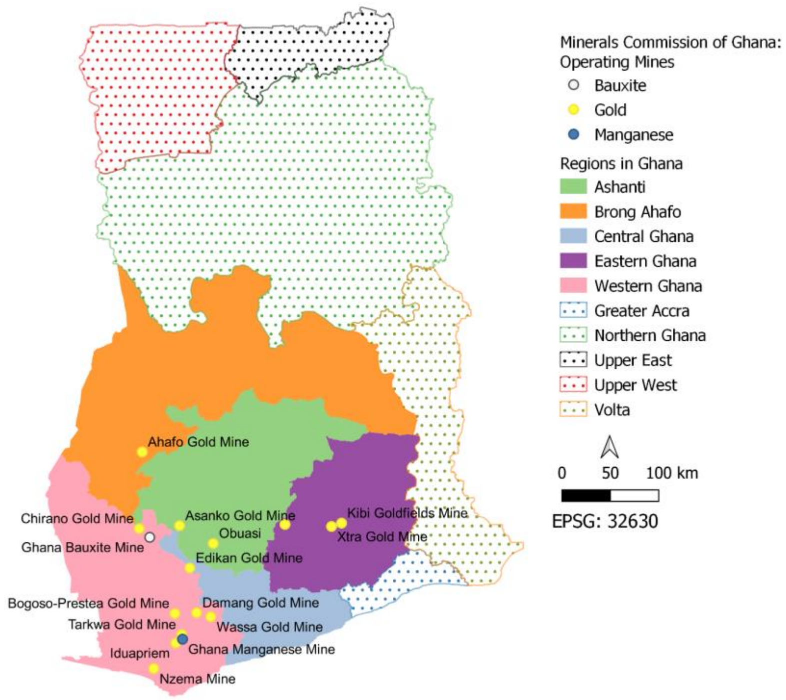



*Figure 1A.* Large-scale operating mines in Ghana- Minerals Commission of Ghana: Point layer.

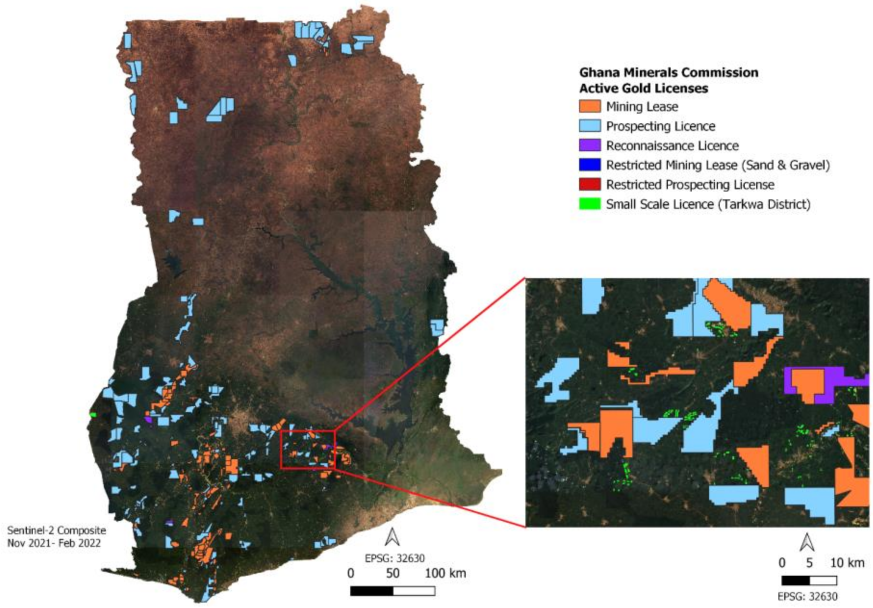



*Figure 1B.* Mining licenses in Ghana (including small-scale and large scale-mines)- Minerals Commission of Ghana

For the purposes of this project, satellite imagery from the Copernicus Sentinel-1 [53]and Sentinel-2 [54]missions were used. Sentinel- 1 was launched in 2014 and it provides temporally dense and high-resolution images through Synthetic Aperture Radar (SAR), returning day and night data that can be retrieved in all atmospheric conditions. Due to its unique sensing characteristics and high spatial resolution, Sentinel-1 has been widely used in many remote sensing research studies [55], [56]. Since cloud cover can be penetrated, working with Sentinel-1 products is advantageous as clouds are considered unwanted noise that would have to be removed by conducting extra preprocessing work. It is possible to operate the SAR instrument in dual-polarity, transmitting a signal that can switch between the horizontal (H) and the vertical (V) plane. Sentinel-1 has four distinct acquisition modes and the

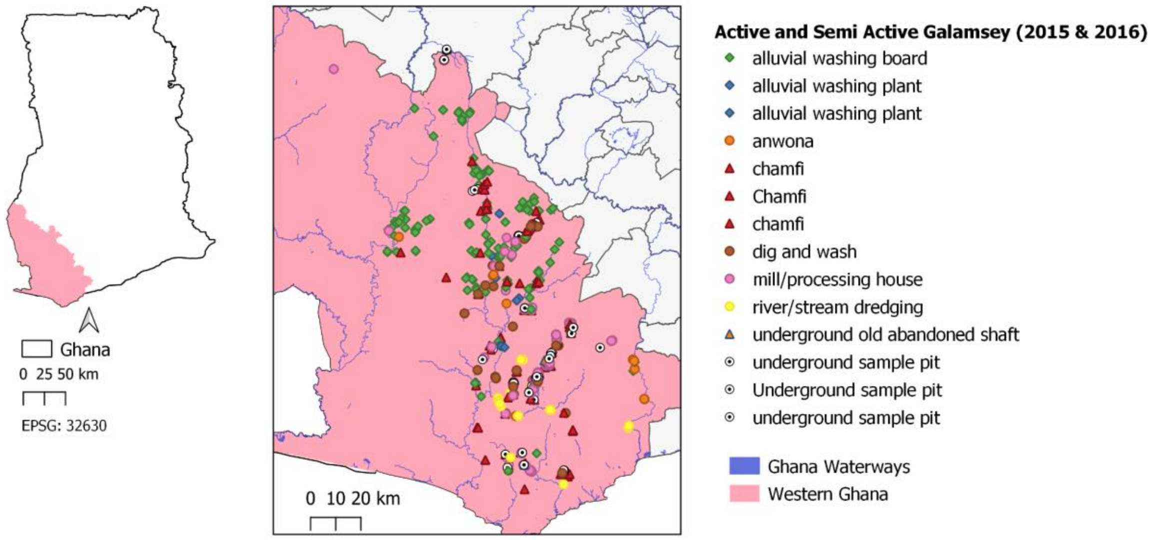



*Figure 2.* Galamsey surveys carried out in the Western Region of Ghana in 2015 and 2016, by Owusu-Nimo et. al.

10m resolution images that were processed in this study were level-1 Ground Range Detected (GRD)- this means that they were previously projected to ground range based on an ellipsoid model of the Earth.

Copernicus Sentinel-2 consists of a constellation of two satellites that are flying in the same orbit and are synchronously phased at an angle of 180 degrees. They carry onboard multi-spectral scanners and this mission has been providing new insights into monitoring the Earth's surface for seven years. Sentinel-2 has been widely used among the remote sensing scientific community in a variety of projects in the fields of agriculture, forestry management, land cover classification, and water quality monitoring [57][58] Sentinel-2 mission offers a wide variety of products at different spatial resolutions and bands 2, 3, 4, and 8 have the highest level of detail (10m by 10m square pixels). Images can also be retrieved at a resolution of 20m (bands 5, 6, 7, 8A, 11, and 12) or at a 60m resolution (bands 1, 9, and 10). For the purposes of this project, images with a cloud cover coverage of less than 20% were acquired, and Sentinel-2 images with bands at a 10m and 20m spatial resolution were selected.

## 3.2. Data Preprocessing

### 3.2.1. Vector Data

Creating the Training Dataset

Data quality plays an important role in the model's performance. In order to make accurate predictions, a detailed and comprehensive vector dataset of the features of interest in the study area is required. However, illegal activities are not generally mapped or acknowledged as they are not officially registered by government authorities. It is also very difficult to find good geodata material in Africa, especially in developing countries like Ghana. However, numerous Galamsey mining sites were required for the machine learning model training and were digitized as ‘samples’ that the model learned from. Although the Owusu-Nimo et. al [20] dataset was a vector layer with points and did not have any information on the extent of Galamsey sites, it was useful for gaining a general understanding of the location of the target features. Previous Galamsey studies were also of great importance to assist the digitization of training data, as they described different types of Galamsey, as well as some other physical characteristics that can be easily observed with satellites from space [10] [21]. Additionally, the Minerals Commission of Ghana’s database of active mining licenses and large scale-operating mines [40][41] also served as a valuable guide when locating illicit small-scale mining, by ruling out the areas where mining was permitted.

Types of Galamsey

A Galamsey activity can be classified as a gold processing operation (Mill House Galamsey), both mining and gold extraction (Placer/Alluvial Galamsey and Chamfi/Surface Galamsey ), or as a stand-alone mining operation (primarily Underground and Selection Galamsey) [20], [21].

Mill Houses are “structures” that are usually located by the side of roads and near water sources. In these structures, miners crush and break the rocks (that contain gold) into smaller pieces, to then wash the smoothened material [21]. Human-made structures were not included in the illegal mining vector dataset created for model training, as there is not enough land use geodata in Ghana to identify this Galamsey type. However, the visual negative impacts of this type of uncontrolled operation were included in the digitization process: surface water pollution such as water ponds contaminated with Mercury, oil spills, and waste piles. Selection or “Pilfering Mining” Galamsey is a clandestine method in which a Galamseyer selects medium or high-grade ore from a waste rock disposal area at a licensed LSM or ASM/SSM site. In most cases, visits occur at night or when it’s raining and Galamseyers avoid wearing bright color clothes. This type of illicit mining is considered to be the simplest and easiest among all types of Galamsey, and Ghanaians refer to it as a fast way of becoming wealthy [21]. Selection Galamsey does not involve the physical mining of gold, so the overall negative impact is considered “low” and similar to the Mill-House Galamsey.

Placer or Alluvial Galamsey activities generally require large pieces of land, and are usually found along river banks, where reliable water supply can be guaranteed. The process of Chamfi or Surface Galamsey entails mining ore from one to a couple of meters into the ground and it can be done manually or mechanically (with the use of big equipment such as excavators and loaders). This is the most cost-effective method for mining lower-grade ores in the Western Region of Ghana. A typical set-up is normally a shelter, sandbags, a mine waste disposal area, a dump for processing waste, and a good water supply system. Some of the visual characteristics of this type of Galamsey include river pollution, waste piles, and hydrocarbon spills [21]. Re-entering abandoned shafts and sinking sample holes (by taking advantage of the abandoned shafts’ ventilation and water draining systems) were observed as the two most frequent Underground Galamsey operations[21]. A typical sample hole has a diameter of approximately one meter, and the deep learning model produces a 10-m resolution output raster. It is therefore difficult to capture this type of Galamsey. Due to its intrinsic nature, Underground Galamsey was disregarded in the illegal mining digitization process, as these illicit operations are carried out below the surface, and hence they are hard to detect with remote sensing techniques.

Visual characteristics of Galamsey sites

*Open Pits (Active and Abandoned)*

Licensed mining companies build mine pits with big machinery such as excavators. Depending on the gold extracting area, pits can range from approximately 37 $m^2$ to 372 $m^2$, and from 2 to 9 meters in depth [10]. Galamsey operators also construct pits for mineral extraction but they do it in a more rudimentary way (with basic elements such as pickaxes and shovels) and without the necessary equipment. This leads to a more labor-intensive task, and the size of the pits is generally smaller (this small-scale mining spatial characteristic was key in identifying Galamsey from satellite images during the digitization process). Although the focus was on identifying pits with small land coverage, some Galamseyers risk their lives trying to find gold in abandoned pits from the big mining licensed firms. When creating the vector dataset for the deep learning model, pits that were excavated by licensed LSM, hence large in size, were also used as a reference when looking for unlicensed mining areas.

*Deforested Areas*

Open-pit mining is the most common global technique for mineral extraction. This method is preferred when minerals are found close to the soil surface. During its construction process, vegetation has to be removed, leaving deforested patches of land *(Figure 3)*. This geospatial pattern can be easily observed from satellite imagery. Although large deforested areas were used as a reference in the digitization process, Galamsey activities are carried out with elementary tools and thus the areas with no trees or vegetation are smaller in size.

*Tailing Ponds*

Residual materials are common during ore extraction. After the ore is processed, it is 'washed' (mixed with water- this chemical combination is called a slurry, which is easier to transport) and is stored in tailing ponds. Tailing ponds contain wastewater from these mining operations and are generally toxic as they have sulfide minerals. After the washing process is completed, some companies discharge the tailings into the previously excavated open pits, so that this wastewater effluent can be reused as many times as possible before discharging it into the close-by rivers. However, the majority of licensed miners discharge tailings directly into the surrounding water bodies without any type of chemical or physical treatment, producing pollution that changes the natural color of water in rivers [10]. These hazardous and highly toxic products are also discharged by unlicensed miners into the previously excavated pits. However, these open pits are smaller in size (as they are built in a more primitive manner and without the necessary equipment) and they can be observed through satellite images, as shown in *(Figure 3)*.

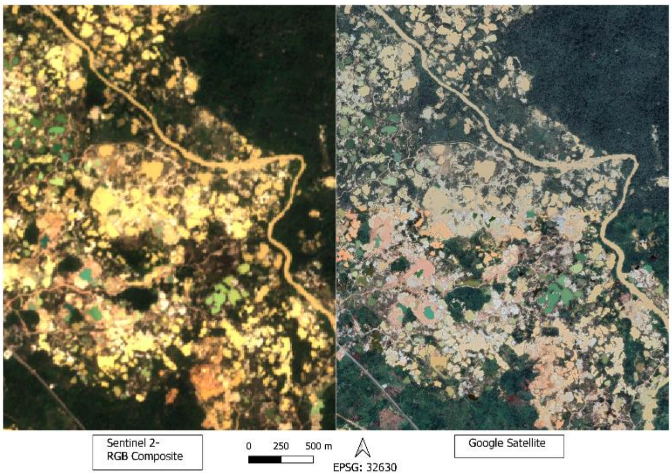



*Figure 3.* Galamsey tailing ponds and deforested areas along the Ofin river in Upper Denkyira West District, Ghana.

*Land Color*

Heavy metals such as iron are associated with the presence of gold deposits. Land color can be also an indicator of the presence of gold: areas characterized by red or black soil that were close to mining features were also used as a reference to spot Galamsey operations.

*Proximity to Rivers*

The majority of Galamseyers are located near rivers because they require large amounts of clean water for their gold mining operations (such as the "washing" part of the process, to be able to transport the ore). Illegal miners are also found close to main water streams due to the abundance of alluvial deposits, where they can easily encounter minerals. The "alluvial washing board" is the most popular type of Galamsey in the Western region of Ghana [20]. This unlicensed method is the most efficient and profitable way to extract low-grade ore for commercial purposes. Even though this technique's equipment is a bit difficult to transport (because big trucks are needed), it is not only relatively easy to install but does not require special training to operate, making it a very popular source of income for many Ghanaians. The main water streams of Ghana were retrieved from OSM [35] in a vector format, and they were used as a reference when digitizing illegal mining activities.

*Distance to Roads*

Contrary to some types of Galamsey (such as the "Mill House" operation) which can be located in the center of villages or towns, the most popular type of illegal operation among Ghanaians is typically found in remote areas where security and enforcement rules are inconsistent[20] . It is difficult for security authorities to gain access to these remote and unlicensed mining areas opens because they are generally far from roads. Hence, national security rules are limited in these rural areas, or even nonexistent. This allows the free movement of the "washing board" equipment such as big excavators and helps the Galamseyers to work without restraints [20]. The main roads of Ghana were also accessed through the OSM platform [35], and the proximity to roads was taken into account when visually examining Galemsey agglomerations.

*Forest Reserves*

The majority of alluvial Galamseyers generally seek densely vegetated areas that are hard to spot. Typically, they work in rural settings and within forest reserves that are protected at a national level (such as Bonsa, Ekumfi, Neung South and Neung North). This poses a threat not only to their own safety but to the environment as well, increasing the likelihood of degradation. Forest reserves are vast, and due to the abundance of trees, it is hard for security enforcement to detect illegal mining operations. A dataset of protected areas in Ghana was downloaded from The World Database on Protected Areas (WDPA)[50], and the forest reserves were shortlisted to conduct a thorough visual assessment of Galamsey sites.

Digitization of Galamsey features

This part of the geospatial analysis was conducted using QGIS. Ghana was assigned the European Petroleum Survey Group (EPSG) code of 32630, which refers to the geographic coordinate system (GCS) WGS 84, and projected coordinate system (PCS) UTM Zone 30N. An initial step involved creating a grid (aligned with the satellite imagery) covering the entire study area with 320m vertical and horizontal spacing. As a second step, a polygon layer containing 35 rectangles was created,  representing the boundaries of the training sites for the deep learning model.  A total of five test sites were also produced, as they are required to assess the accuracy of the deep learning algorithm. Although each training site was required to have a minimum area of  640 by 640 meters (or 409600m$^2$), the majority of the sites exceeded the minimum requirement (since having more training data generally improves the deep learning algorithm's performance). After delimiting the boundaries, Galamsey features were created using the RGB Sentinel 2 Composite as a reference and then clipped to the training sites *(Figure 4)*.

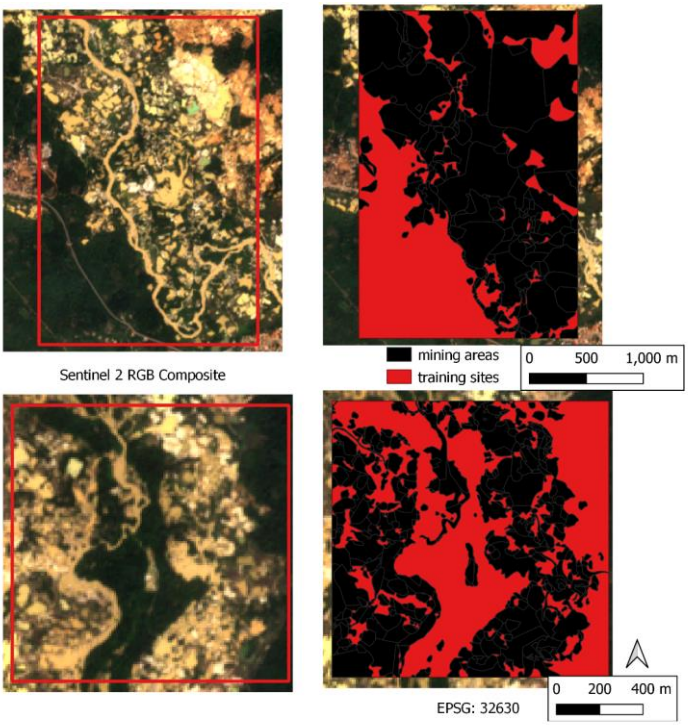
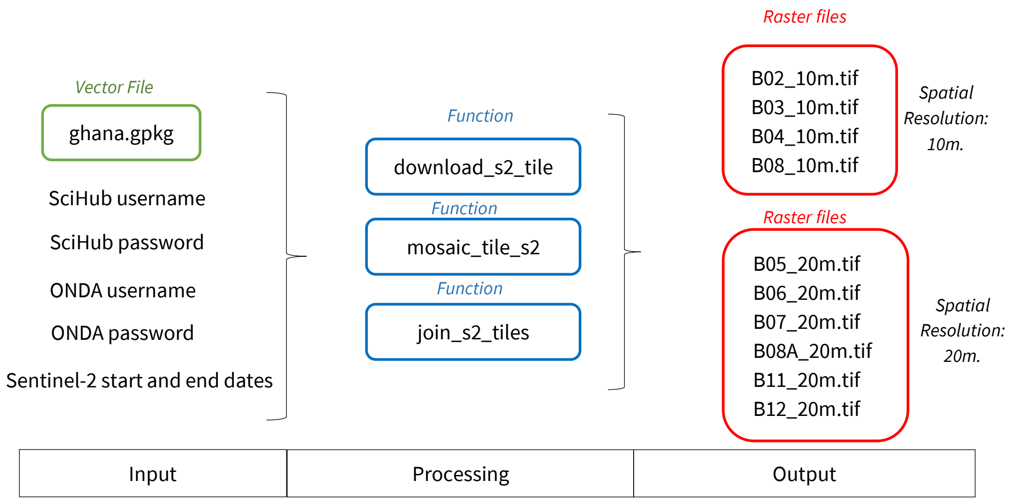



*Figure 4.* Examples of training data: digitized Galamsey mining areas with their corresponding boundaries.

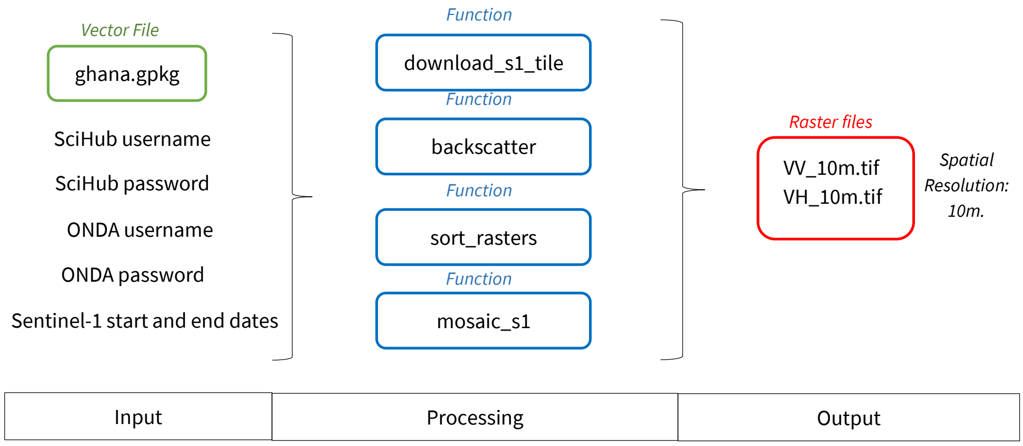
### 3.2.2. Raster Data

Download and process satellite images

Satellite imagery was acquired and processed using using python-based toolchains with Buteo Toolbox. Scripts were modified and customized to meet this project's specific needs. It was necessary to install some important modules and libraries before executing the downloading and raster processing functions. 'Sys', 'os', 'glob' and 'shutil' modules were imported. 'Sys' [59]is an integral part of Python's runtime environment, it contains several functions and variables that are used to interact with the interpreter. Among Python's standard utility modules is the 'OS' module [60], which offers a wide range of functions to interact with the operating system. The 'glob' module [61]is part of the standard library in Python and it allows to find file paths and folders whose names follow a particular pattern. Using the 'Shutil' module[62], which is part of Python' standard utility modules, complex operations like creating and remotely manipulating files can be performed. 'Shutil' enables automated operations such as copying and deleting files or directories. Shutil.rmtree() command deletes a directory tree, and it is very useful to optimize memory usage. Some other important Earth Observation functions were imported to help process the Copernicus satellite data.




*Workflow 2.* Sentinel-2 imagery processing steps




*Workflow 3.* Sentinel-1 imagery processing steps

The optical and radar data processing analysis is described in *Workflow 2* and *Workflow 3*. A vector file delimiting the study area, which comprised Ghana in a GeoPackage format, was retrieved and used as a reference when downloading the Sentinel-2 tiles. Automated folder creation was one of the first steps of the process, together with determining the geographic and projected coordinate system of the project (EPSG: 32630).

As clouds block Copernicus Sentinel's view of the Earth, images may not always be usable in Ghana, despite the satellite's six-day frequency over the country. Sentinel-2 Copernicus products with less than 20% cloud coverage were retrieved to obtain an almost cloud-free representation of Ghana. To compensate for the occurrence of clouds, Sentinel-2 mosaics of several images were created for a period of three months from November 1, 2021 to February 1,2022. Sentinel-1 active sensors are able to penetrate through clouds, so they do not have atmospheric limitations. To avoid imagery redundancy, the two middle weeks of December 2021 were considered for the Sentinel-1 downloading part of the process. A set of variables were created to allow easy coding and avoid cumbersome syntax in the processing functions:

```
#Project dates for Sentinel-2
s2_project_start = "20211101"  # yyyy-mm-dd
s2_project_end = "20220201"  # yyyy-mm-dd
#Project dates for Sentinel-1
s1_project_start = "20211206"  # yyyy-mm-dd
s1_project_end = "20211220"  # yyyy-mm-dd
```

Copernicus's full archives of Sentinel missions are available in a data catalog through ONDA [63]. ONDA is a hosting service where users can access geospatial data and build cloud-based applications. In order to get access to the full imagery archive, SciHub [64]and ONDA accounts need to be created. A connection with API Hub Access [65]was generated to allow an automated downloading process.

A series of Buteo functions were customized according to this project's needs and then executed to download the Sentinel-1 and -2 tiles and create the final raster mosaics.

The 'download_s2_tile' function requires a certain number of arguments: the SciHub and ONDA usernames and passwords, the destination folder directory, the Sentinel tiles that intersect with the input extent (project area), the project dates, and the maximum cloud cover percentage.

The 'download_s1_tile' function requires the same arguments, but the Sentinel-1 images that intersected the Sentinel-2 tiles were downloaded, and a minimum overlap of 10% with the Sentinel-2 images was required.

After completing the downloading steps, the Copernicus raw files were unzipped and the mosaicking process took place- through a weighting-based optimization. Sentinel-2 10m and 20m tiles were joined and harmonized separately, clipped to the project extent, feathered, and reprojected to match Ghana's coordinate system. As the script was running, temporal files were being deleted simultaneously to increase memory optimization. The final Sentinel-2 RGB composite of Ghana is shown in *Figure 5*.

Radar images are processed quite differently when compared with optical satellite products, as Sentinel-1 instruments use particular remote sensing techniques. Copernicus raw files were processed in "chunks", meaning that the project area was split into smaller polygons in order to avoid computer memory difficulties. The ESA suggests the use of the Science Toolbox Exploitation Platform [66]together with the Sentinel Application Platform (SNAP) Desktop [67] tools when processing Sentinel-1 images. There are several user guides on this platform created by the ESA and Skywatch [68]that has easy-to-follow steps demonstrating how to process SAR images. However, the SNAP software is best suited when working with small projects, as processing several images requires manual work and it can be very time-consuming. When working with big areas, such as the entire country of Ghana, it is better to use an automated approach to process radar data. The SAR images' final output

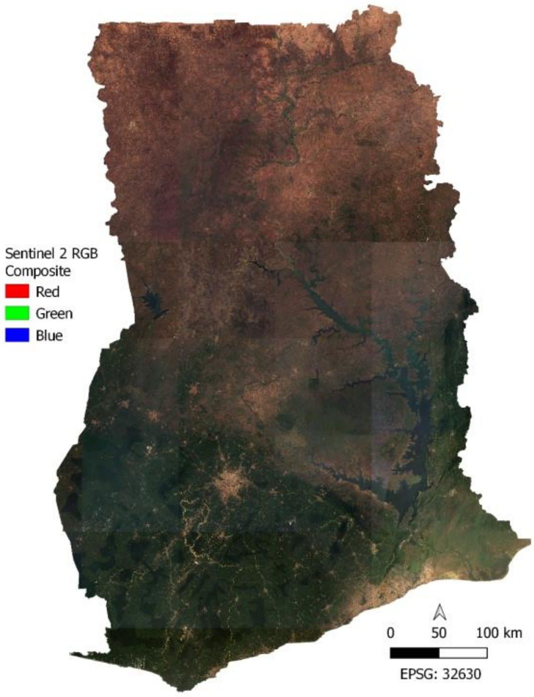



*Figure 5.* Ghana Sentinel-2 RGB Composite (November 1, 2021- February 1, 2022).

consisted of two raster images characterized by two radar polarization modes: Vertical Transmit- Vertical Receive (VV) and Vertical Transmit- Horizontal Receive(VH). The 'backscatter' function in Buteo follows the backscattering steps suggested by the ESA, which are observed in *Workflow 4*, and a python-based script was created to process the SAR images. The following backscattering workflow steps were followed:

*Apply orbit file:* a number of sensors are used to detect the orbiting track for satellites. Applying the orbit file improves geocoding and helps with getting a more accurate satellite position and velocity values.

*Remove GRD border noise:* Range compression causes radiometric artifacts and the sampling start time should be corrected to account for the curvature of the Earth. It is necessary to remove noise that is characterized by low intensity and also to delete invalid data at the borders of the images.

*Remove thermal noise:* Thermal noise occurs when the electrons of the satellite's devices move randomly due to thermal motion, disturbing the Sentinel-1 image intensity. Removing thermal noise normalizes the backscatter signal and improves the overall image's quality.

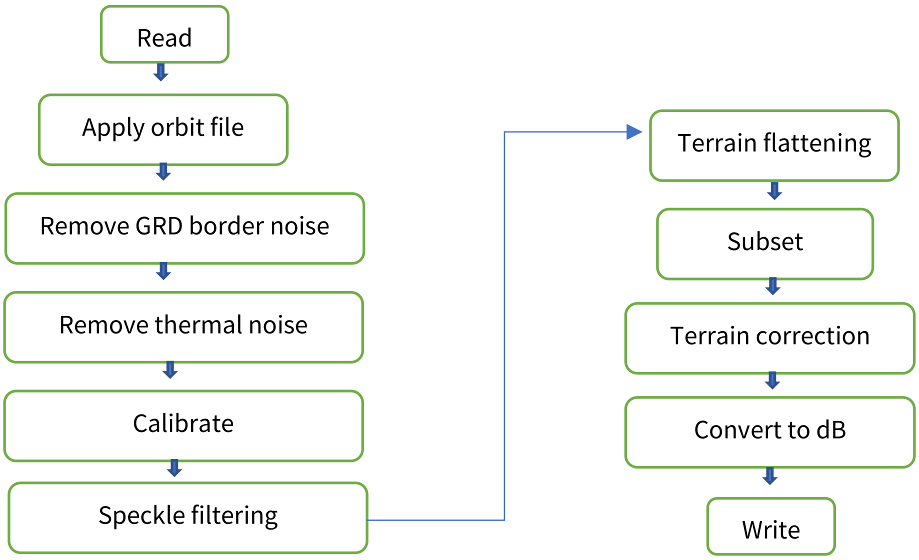
*Calibrate:* During the calibration process, radar images are corrected so that their digital pixel values accurately characterize the surface's reflected radiometric backscatter energy.

*Speckle filtering:* Speckle is present in SAR images in the form of granular noise and it is caused by the interaction of the waves that are out of phase and come from the surface. The use of a speckle filter is not suggested when identifying small spatial features as it might delete this type of geospatial data. A median filter was considered when processing the Ghana images, substituting each signal value with the median of its neighbors, which helped with the "smoothing" process of the images.

*Terrain flattening:* radiometric variations of the images are reduced in this backscattering step, which is particularly important in embossed areas.

*Terrain correction:* the goal of terrain correction is to compensate for geometric distortions, the images were geocoded and matched to the real terrain as closely as possible.

*Convert to dB:* in this step, the GammaVV and GammaVH backscatter coefficient images are converted to decibels through the following logarithmic formula: 10*log10 (GammaX).




*Workflow 4.* ESA Sentinel-1 Toolchain.

After performing the preprocessing part of the analysis suggested by ESA, VV and VH paths were mosaicked and aligned separately at a 10m resolution using one of the Sentinel-2 bands as a master raster. The Sentinel-1 RGB composite of Ghana is shown in *Figure 6*.

Ghana covers an area of nearly 240000 km$^2$. Several tiles had to be downloaded in order to produce complete mosaics of the country, which requires high levels of computer processing power. Working with large satellite datasets could be cumbersome. Hence, a series of GDAL [69]commands including 'gdal_translate' [70]and 'gdalwarp' [71] were used for clipping, separating bands, resampling, and compressing raster images.

GDAL compressing command example:

*Sentinel-2:* `gdal_translate -co COMPRESS=DEFLATE -co PREDICTOR=2 -co BIGTIFF=YES -ot UInt16 B02_10m.tif B02_10m_compressed.tif`

*Sentinel-1:* `gdal_translate -co COMPRESS=DEFLATE -co PREDICTOR=3 -co BIGTIFF=YES -ot Float32 VV_10m.tif VV_10m_compressed.tif`

This function creates a BigTIFF file (file larger than 4GB), performs a deflate compression with predictor 3 (horizontal differencing), and creates an image with data type Float 32.

GDAL resampling command example:

*Sentinel-2:* `gdalwarp -tr 10 10 -r average B11_20m.tif B11_10m.tif`

This function creates a 10m by 10m .tif file, resampling the input image (20m resolution) with an averaging resampling method.

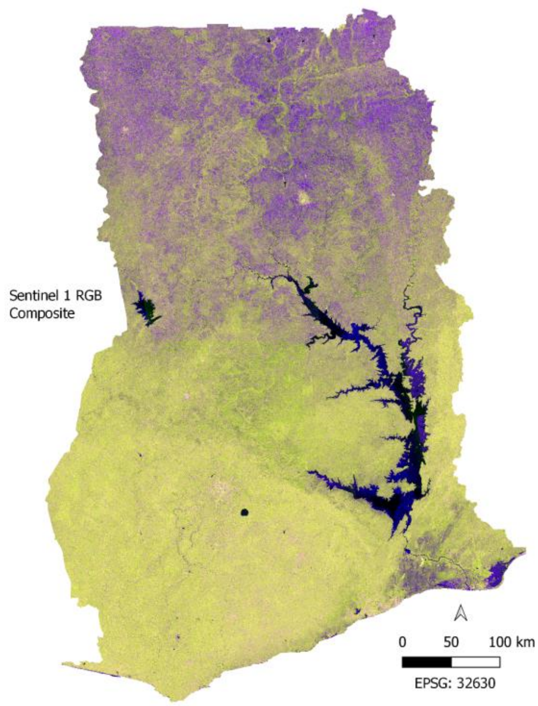



*Figure 6.* Sentinel-1 RGB composite of Ghana (December 6, 2021- December 20, 2021).

## Vegetation Indices

In remote sensing, vegetation indices have numerous benefits. Earth observation sensors play a key role when supporting vegetation research. In this study, two vegetation indices, NDVI [72]and NDMI[73] , were computed using GDAL and fed into the machine learning algorithm to better detect Galamsey hotspots. Vegetation indices are advantageous when identifying illegal mining activities since Galamseyers usually work in forest reserves where they can easily hide and be far from law enforcement authorities.

*Normalized Difference Vegetation Index (NDVI)*

NDVI is widely used in agriculture since it is a simple calculation for quantifying green vegetation through plants' active biomass. It is the most popular vegetation index and it was introduced in 1973 [72]. This index is based on chlorophyll absorption (Sentinel-2 red band 4 →B4) and dispersion of green leaves (Sentinel-2 band 8- NIR- near-infrared→ B8).

S-2 NDVI= NDVI (B8,B4)= (B8 – B4) / (B8 + B4)

NDVI values range between -1 and 1. Values close to -1 represent water, and values close to 0 represent barren lands such as sand, snow, or urban areas. Positive low values in between correspond to grassland and shrub, and positive high numbers close to 1 indicate dense green vegetation such as forests.

*Normalized Difference Moisture Index (NDMI)*

NDMI is used to estimate the water content of vegetation. This index is based on Sentinel-2 SWIR band 11 (B11) and Sentinel-2 band 8- NIR- near-infrared(B8).

S-2 NDMI=NDMI (B8, B11)= (B8 – B11) / (B8 + B11)

NDMI also ranges between -1 and 1. Low values correspond to low water content, and high values indicate high vegetation moisture.

## 3.3. Patch extraction and Normalisation

### 3.3.1. Patch Extraction

This step consisted of preparing the dataset to train the CNN model. The digitized vector mining areas (including the training and testing datasets) were masked and rasterized to a high resolution, labeled (so that the model can learn and then label new raster pixels as Galamsey spots in the prediction process), and then resampled to match the satellite imagery resolution (10m and 20m). The following step was to convert the raster data to NumPy [74]arrays. This process was done in Python, importing a ML function that extracted patches from satellite images. The 'Sys', 'os', and 'glob' modules were also required for the patch generation process.

```
folder= "C:/Users/MALT/Desktop/Ghana/"
m10 = glob(folder + "*10m*.tif")
m20 = glob(folder + "*20m*.tif")
out_path10= "C:/Users/MALT/Desktop/Ghana/training_data/out_patches_10m/"
out_path20= "C:/Users/MALT/Desktop/Ghana/training_data/out_patches_20m/"
mines_boundaries="C:/Users/MALT/Desktop/Ghana/training_data/mines_boundaries.gpkg"
mines= "C:/Users/MALT/Desktop/Ghana/training_data/mines.gpkg"

extract_patches(
```

```
            m20,
            out_path20,
            tile_size=16,
            zones=mines_boundaries,
            options=
                { "label_geom": mines },
        )
```

The patch extraction process is shown in *Workflow 5*. Studies have shown that deep-learning algorithms that use CNNs lead to processing times optimization and perform well when using patches as a training input [75]. The extraction of patches consists of breaking the original image into small segments. Patches allow not only target data information to be captured, but also neighboring local information to be detected, which makes it a better approach when comparing it with other methods such as semantic segmentation (that only relies on single pixels). It was assumed that the model would have trouble distinguishing objects with similar spectral signatures. Hence, the training data also contained patches with absence of Galamsey features, such as urban areas, roads, and water bodies that looked like tailing ponds. However, too many absence Galamsey sites could lead to an imbalanced dataset and therefore negatively affect the model's performance. As a result, a reasonable number of absence sites were included in the training data. The extracting patches process requires a raster list as an input, an output directory, the digitized features, and the boundaries of the digitized features. The tile size should also be specified. For the purposes of this project, 10m resolution images were assigned a tile size of 32, and the 20m raster list had a tile size of 16. The pixels that were spatially covering the extent of the Galamsey mining boundaries were given a float number that ranged from 0 to 100 *(Figure 7)*. Pixel values that were zero represented the absence of Galamsey activities, whereas the pixels that fully intersected the Galamsey locations were assigned a value of 100. Values in between account for the partial presence of Galamsey features, which gives some flexibility to the CNN predictions. This method follows a regression approach, which helps reduce spatial uncertainty and extracts the target pixels that are required for the training part of the process. This process was done for training and testing datasets.

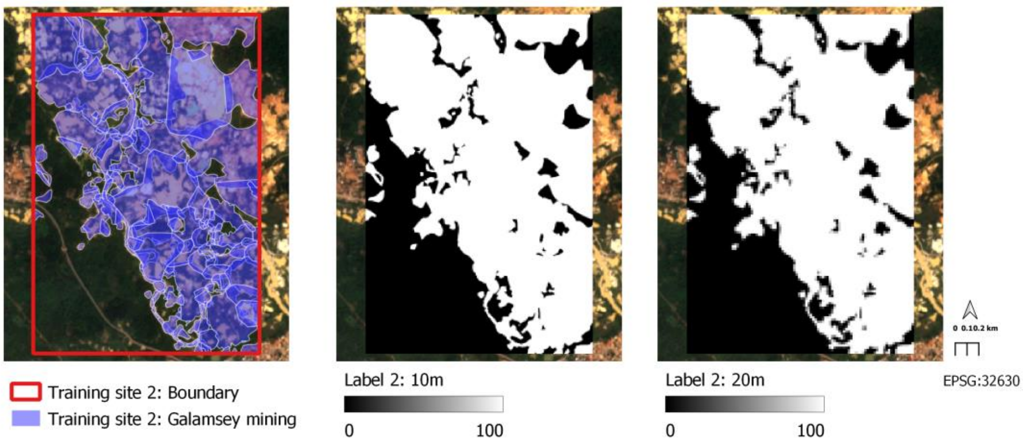
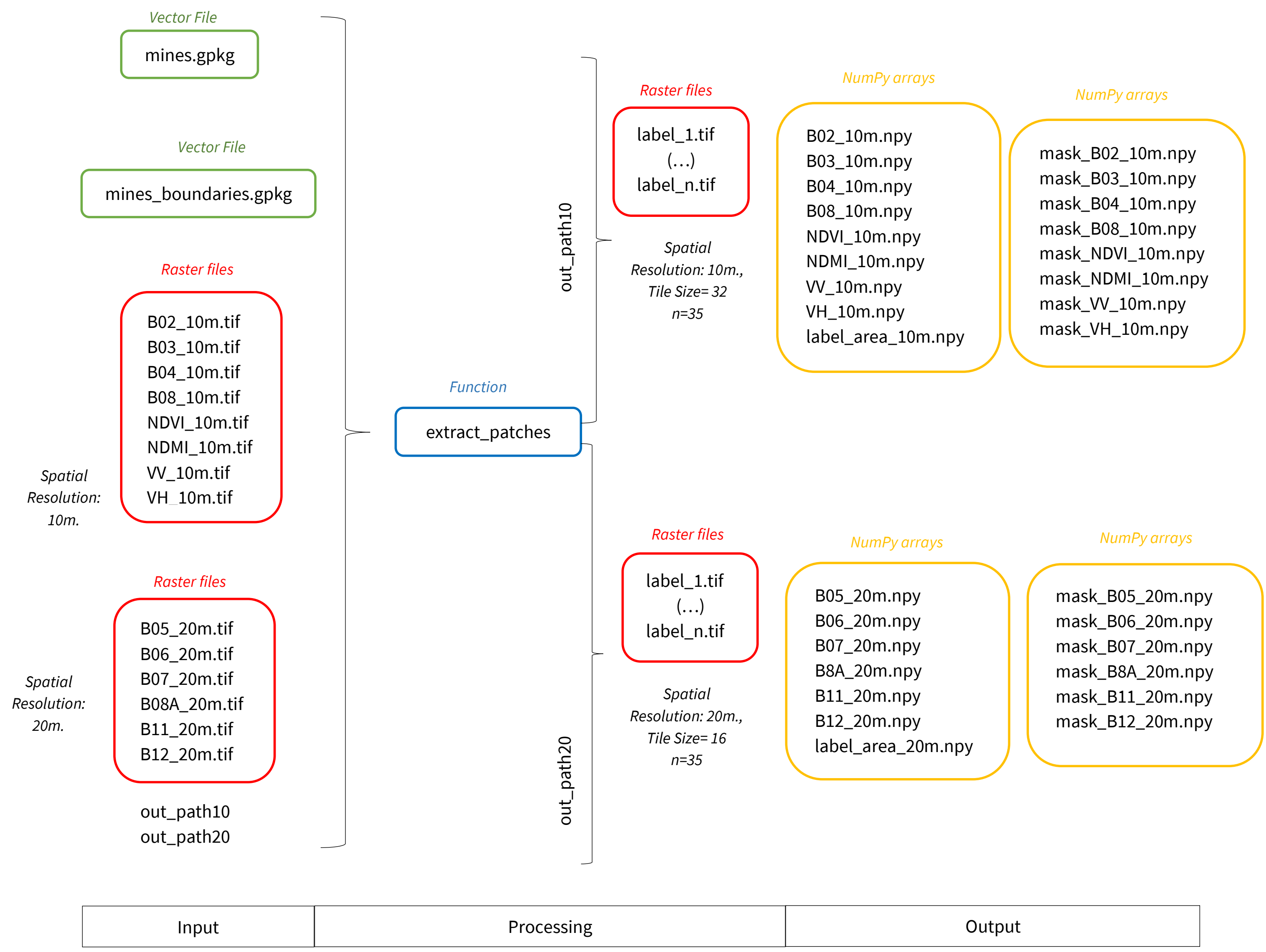



*Figure 7.* Example of the patch extraction output: a mining training site with its corresponding 10m and 20m .tif labels.




*Workflow 5.* 10m. and 20m raster imagery patch extraction process (for training and testing datasets)

### 3.3.2. Normalization

A major factor in training a deep-learning model is the scale of variables (input and output datasets). A learning process with unscaled values generally leads to slow and unstable model performance, and sometimes the ML algorithm fails to perform basic learning steps. The performance of CNNs can be dramatically improved if the input data is normalized by doing some feature scaling. The training and test datasets can have different types of data (e.g. floating or integer values), different ranges, and even different units. The data that was created in the patch extraction step was used as input for the normalization operation. NumPy arrays were normalized to values that ranged from 0 to 1, as it has been demonstrated that deep-learning models perform better with small weights. Training and testing datasets consisting of NumPy arrays of RESWIR, SAR, and RGBN-VEG were created *(Table 1)*, as well as a label area array with shape (20721, 32, 32, 1).

| Acronym | Satellite Mission | Spatial Resolution (m) | Bands | Array Shape (Train) | Array Shape (Test) |
|---|---|---|---|---|---|
| RGBN-VEG | Sentinel-2 | 10 | B02, B03, B04, B08, NDVI and NDMI | (20721, 32, 32, 6) | (803, 32, 32, 6) |
| RESWIR | Sentinel-2 | 20 | B05, B06, B07, B11 and B12 | (20721, 16, 16, 5) | (803, 16, 16, 5) |
| SAR | Sentinel-1 | 10 | VV and VH | (20721, 32, 32, 2) | (803, 32, 32, 2) |

*Table 1.* Satellite imagery information used to train and test the machine learning model

The final arrays were then converted into a .npz format [76], which is a compressed file containing the RGBN-VEG, RESWIR, SAR, and label arrays, for better data storing.

### 3.4. Deep learning: Training

The creation of the CNN machine learning model was done in Python language. Among other ML modules and libraries, TensorFlow and Keras [77]were imported to build the deep learning algorithm and then train it. CNNs require a lot of computer memory and in order to train the deep learning model of this project at least 24 GB of RAM was needed. Therefore, an external graphics processing unit (eGPU) was used to avoid memory limitations and allow faster processing times. In order to get the best outcome in a CNN model, it is important to optimize the hyperparameters according to the available training dataset and the characteristics of the target features[42]. The model was first trained on a pilot area *(Figure 8)* that partially covered the regions of Ashanti, Central, and Western Ghana.

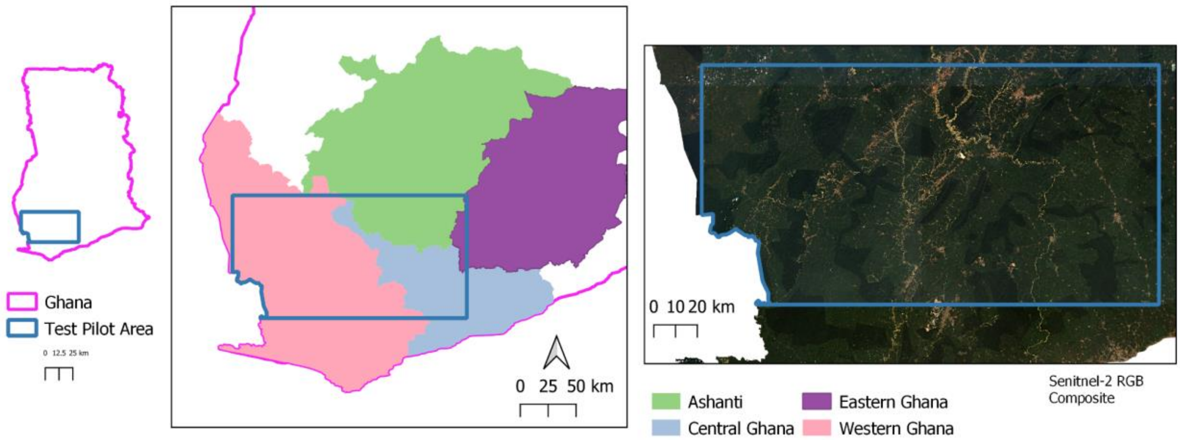



*Figure 8.* Pilot area selected for testing the model.

This area was assumed to have Galamsey activity, as illegal mining operations are often associated with the presence of licensed mining companies, which are located in the southwestern region of the country (a region which is rich in mineral deposits of gold)[40]. Several iterations were carried out, changing the dropout rate, the model size, and the number of inception blocks, among other hyperparameters. *Figure 9* shows an example of the different model versions that were generated.

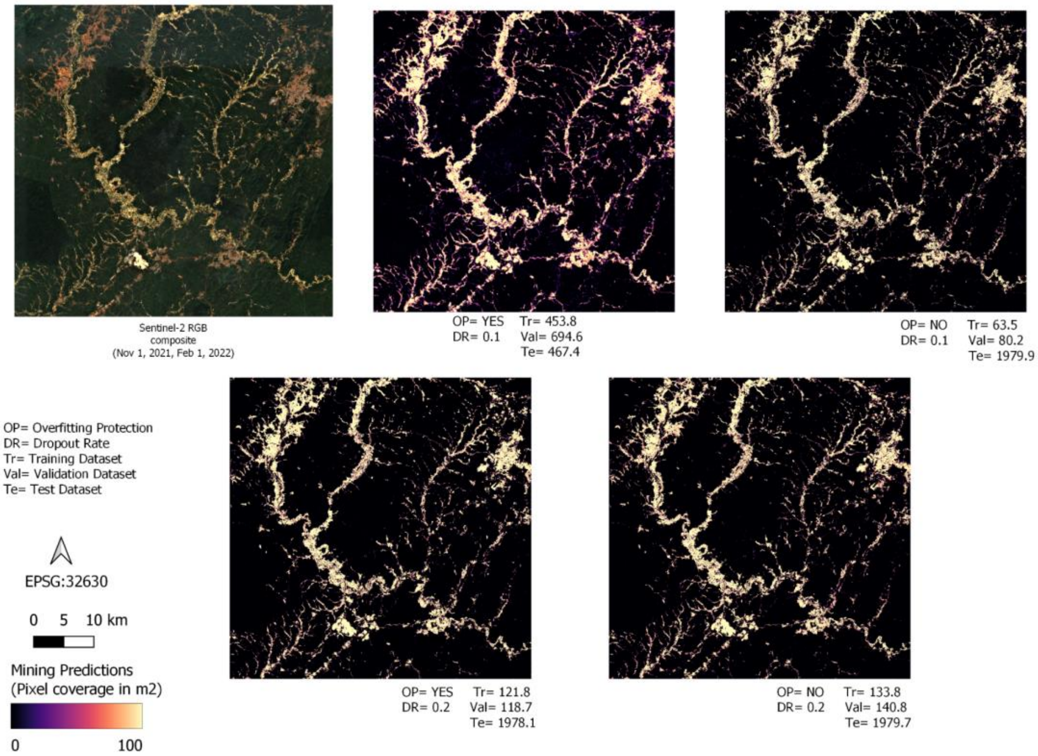
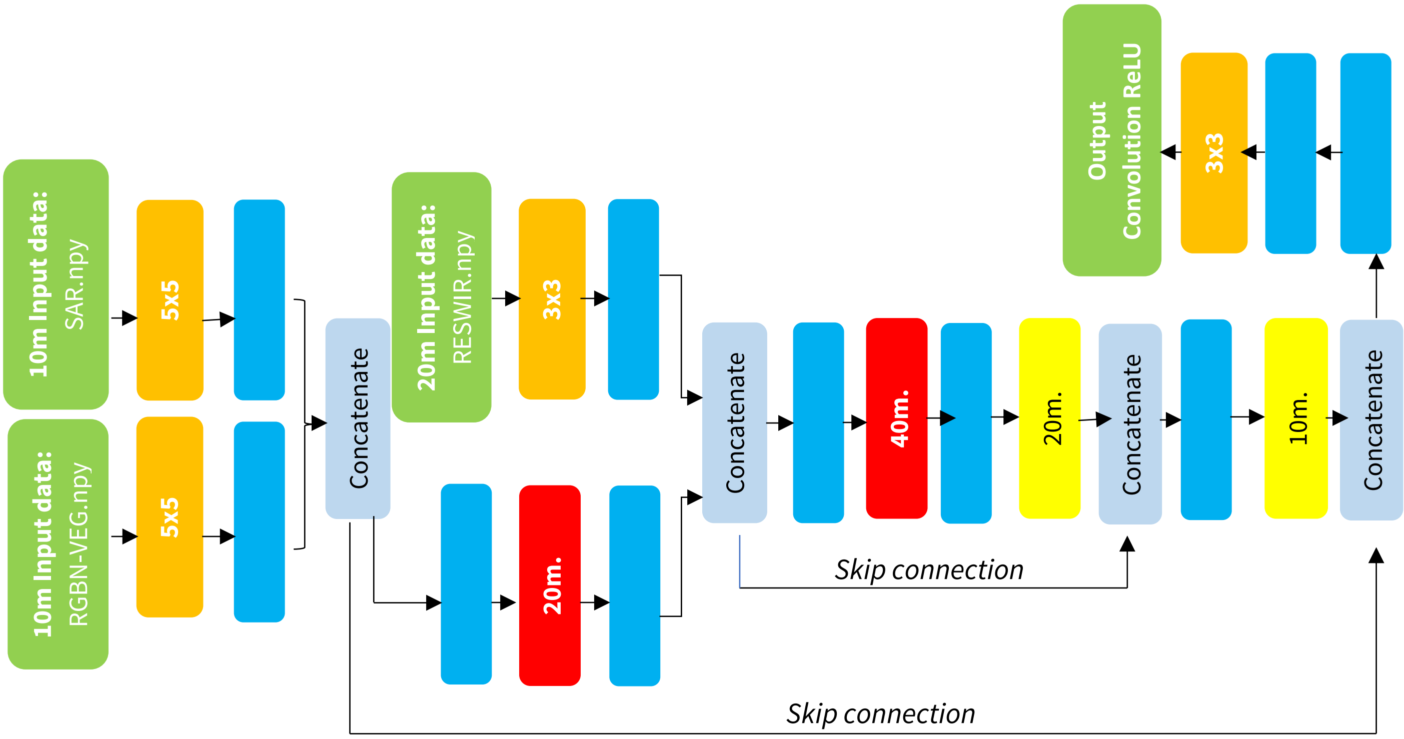



*Figure 9.* Galamsey predictions based on different model versions in a small segment of the pilot area.

In this case, the size of the training and test datasets, the model size, and the number of inception blocks remained constant whereas the overfitting protection option and the dropout rate values were changed in the different model versions. Although all models were overfitting and their metric values (such as the loss function) were not optimal, this initial training and predicting process in a test area helped adjust the hyperparameters when repeating the process of training and predicting for the entire country. As can be observed in *Figure 9*, urban areas and roads were identified as mining areas, thus when doing the model training at a larger scale, absence training sites (i.e. sites with no Galamsey mining areas) were incorporated to improve the overall performance. Visual inspection of the model predictions was key in determining the hyperparameters for building the final deep learning algorithm.

When increasing the number of layers in deep learning, the problem of vanishing/exploding gradient arises, resulting in a gradient that is either zero or too large, and increasing training and testing error rates. The ResNet (Residual Network) architecture provides a solution to this problem, skipping the model training in a few layers and directly connecting to the output [78], [79]. Large-scale convolution operations and very deep networks are computationally intensive and they tend to overfit. Instead of building a “deeper” model, having a “wider” (Inception) network is more beneficial as multiple filters can coexist and operate on the same level, providing the right kernel size. This also helps the deep learning process to be more computationally efficient and helps select the right kernel size depending on the information provided. After a period of thorough examination, the best results were yielded by an image recognition model that follows an Inception- ResNet -like architecture [80], which is an adapted version of the DL structure described in Szegedy et. al [81].




*Workflow 6.* General CNN model architecture. Legend:

Convolution | Inception Block | Reduction Block | Expansion Block

The idea behind this hybrid architecture *(Workflow 6)*, which is inspired by Fibaek et al. [80], is to make the network "more uniform" and remove any unnecessary complex structures, using inception and reduction blocks, and adding an average pooling layer.

The option of including the vegetation indices bands as only one input was considered, but this approach did not improve the overall performance of the model. As a result, the 10 m vegetation indices, NDVI and NDMI, were combined with the 10 m Sentinel-2 bands as a single input and then included into the deep learning architecture. Different training and testing sizes were evaluated when feeding the model, but the best results were obtained with a training dataset of 35 sites *(Figure 10)*.

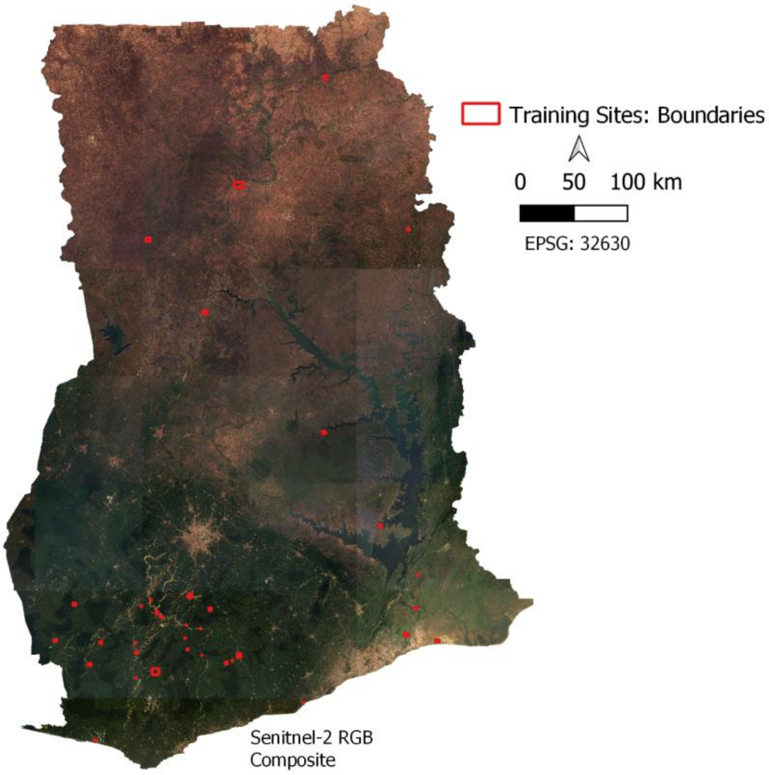



*Figure 10.* Galamsey training data to build the model.

In order to incorporate nonlinearity into the model, an output activation function was implemented in the algorithm: the Rectified Linear Unit (ReLU) function [45], which is widely used in machine learning. Essentially, this function returns the same value if the input is positive (this is the case for our three input datasets used as ground truth); otherwise, it will replace the input value with zero. Having this function is beneficial as it speeds up the training, making the algorithms easy to compute. A stochastic optimization algorithm, "Lookahead"[82], together with the Adam optimizer [83]were used, and the initial learning rate was 0.0001. In order to help the model converge to a local minimum and improve the overall performance by avoiding oscillations, a learning rate decay was implemented:

```
fits = [
    { "epochs": 10, "bs": 16, "lr": 0.0001 },
    { "epochs": 10, "bs": 32, "lr": 0.0001 },
    { "epochs": 10, "bs": 48, "lr": 0.0001 },
    { "epochs": 10, "bs": 64, "lr": 0.0001 },
    { "epochs": 10, "bs": 80, "lr": 0.0001 },
    { "epochs": 10, "bs": 96, "lr": 0.0001 },
    { "epochs": 20, "bs": 112, "lr": 0.0001 },
    { "epochs": 20, "bs": 112, "lr": 0.00001 },
    { "epochs": 20, "bs": 112, "lr": 0.000001 },
]
```

When performing the training of the model, the ground truth dataset was shuffled and split into a training dataset (75%), and a test dataset (25%). The 'Early Stopping' [84] function from Keras was included in the deep learning algorithm. With this Keras callback, the training stops if a previously defined monitored metric is not improving (having a baseline as a reference). The monitored metric was the validation loss, which is used to evaluate the performance of deep learning models. The model stops training based on the patience, which is an integer that was set to 5. This means that after 5 epochs with no val_loss improvement (mode: minimum, meaning when the val_loss is no longer decreasing) the model stopped training/learning. If stopped early, the algorithm restored the best weights.

An overfitting protection function was also included when creating the algorithm, to prevent the model from not performing accurately against unseen data. This function was set to allow 10% of overfitting and the model stopped the training after 3 epochs with no improvement. If not stopped by overfitting, a complete training of the model (120 epochs) took eight hours on average.

The Mean Squared Error (MSE) is defined as the average of the squared difference between the real target values and the predicted values of a model, and it was used as the "loss function". The MSE is one of the most common functions for deep learning regression techniques. Loss functions are very important in machine learning, as the parameters that are learned by the model are dependent on these functions, and they evaluate the performance of an algorithm in modeling the given dataset. In this study, the model was iterated several times and the loss function was constantly monitored. An ideal approach would have been to experiment with different loss functions at the beginning of the process, as loss functions are very much dependent on the type of training dataset and on the model structure [85]. However, the aim of this study is not to make a comparison between different loss functions in deep learning regression algorithms, and due to time limitations, the MSE was chosen as the best candidate as it generally gives good and stable learning results.

The training loss is used to evaluate how well the model is fitting the training data that is given, measuring the error of the training dataset. The validation loss evaluates how well the deep learning model is performing on the validation dataset.

As can be observed in *Figure 11*, the model is characterized by a good fit, as the training and validation loss curves decrease to a point where they reach stability, having a minimum difference between the two final loss function values *(Table 2)*. This is an overall good result, as it means that the model is neither underfitting (a learning curve where the model is not learning from the training dataset) nor overfitting (a learning curve where the model is matching the training data too well and it is less able to generalize patterns on new data).

The Mean Absolute Error (MAE) was also measured after each training epoch. This metric indicates the average of the absolute difference between actual ground truth values and the model estimated values. The smaller the MAE, the more efficient the model is. *Figure 12* shows that although the training and validation learning curves started training with a small gap, they converged over time, which is a good outcome.

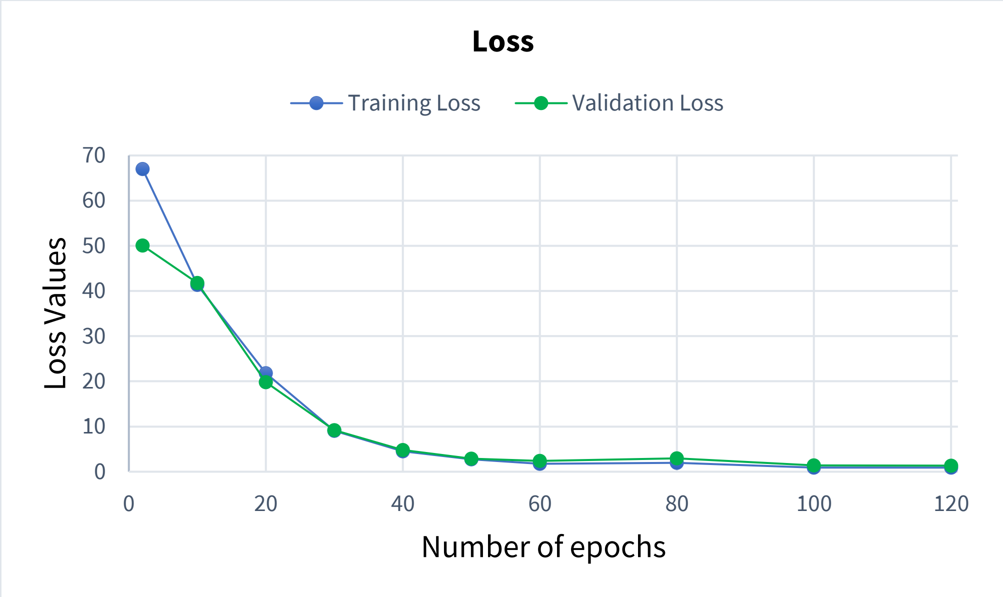



*Figure 11.* Training loss and validation loss against the number of epochs that the model was trained on.

| | Training | Validation | Test |
|---|---|---|---|
| MSE= loss | 0.9196 | 1.3493 | 40.611 |
| MAE | 0.0615 | 0.0623 | 0.6368 |
| TPE | 1.0013 | 0.9992 | 0.9491 |

*Table 2.* Deep learning model statistics for training, validation, and test sets.

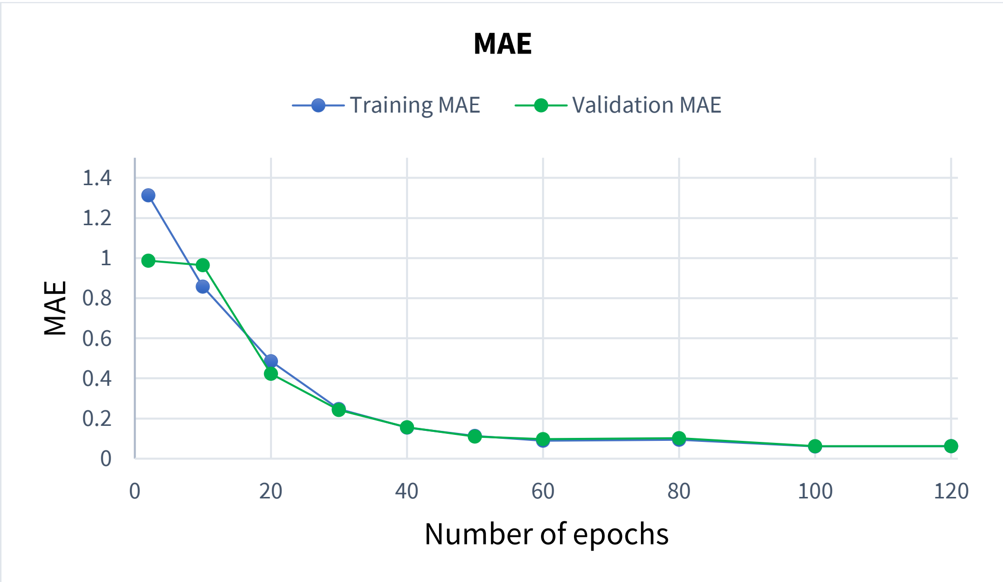



*Figure 12.* Training MAE and validation MAE against the number of epochs that the model was trained on.

Measuring the TPE (%) is also beneficial in deep learning models. If this metric is negative, the model is underestimating and if it is positive, it is overestimating the data. When the TPE is closer to zero, it means that the model is performing relatively well. When the model started learning on the training dataset, it was predicting much higher values than the actual target values *(Figure 13)*.

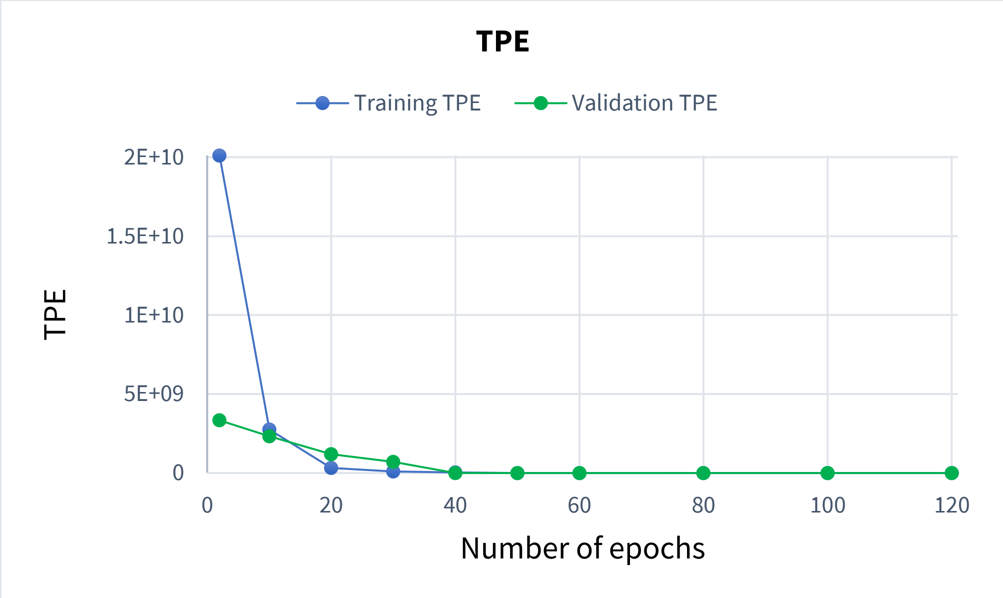



*Figure 13.* Training TPE and validation TPE against the number of epochs that the model was trained on.

However, as the number of iterations increased, the model started learning relatively fast and after the epoch number 40, the TPE learning curves stabilized. The final TPE values were closer to 1 *(Table 2)* for both the training and the validation sets, showing a relatively good performance.

Due to the absence of similar studies or previous research, these model statistics results could not be compared with deep learning reference values, but the model training accuracy is considered to be good given the spatial characteristics of the target features.

### 3.5. Deep learning: Predictions

Like in the training part of the process, the raster predictions were done using the eGPU to allow better processing times and increase memory capacity. The Galamsey mining predictions script was executed several times accounting for the different models that were developed. This allowed having multiple raster maps that were characterized by different hyperparameters and training data sizes. It took almost 48 hours for each prediction to complete. An equidistant grid covering Ghana was used for the prediction process, to avoid memory limitations. This vector layer created in QGIS consisted of 121 50km by 50km rectangles with 100m horizontal and vertical overlapping. A testing raster area was used and included in the script. RESWIR, SAR, and RGBN- VEG .tif raster files were clipped and stacked separately. Following this step, the prediction raster was generated by loading the stacked raster lists, as well as loading a specific model version, and using an output tile size of 32. The individual 121 raster grid predictions were then merged and clipped to the country's cutline. The output of this prediction process is a raster map with predicted Galamsey areas in the form of 10m by 10m squared pixels that range between zero and 100 *(Figure 14)*. Pixel values greater than zero represent the presence of the target features (e.g. if a pixel value is 56.5, then 56.5 m$^2$ of that pixel area is Galamsey, which represents the 56.5%).

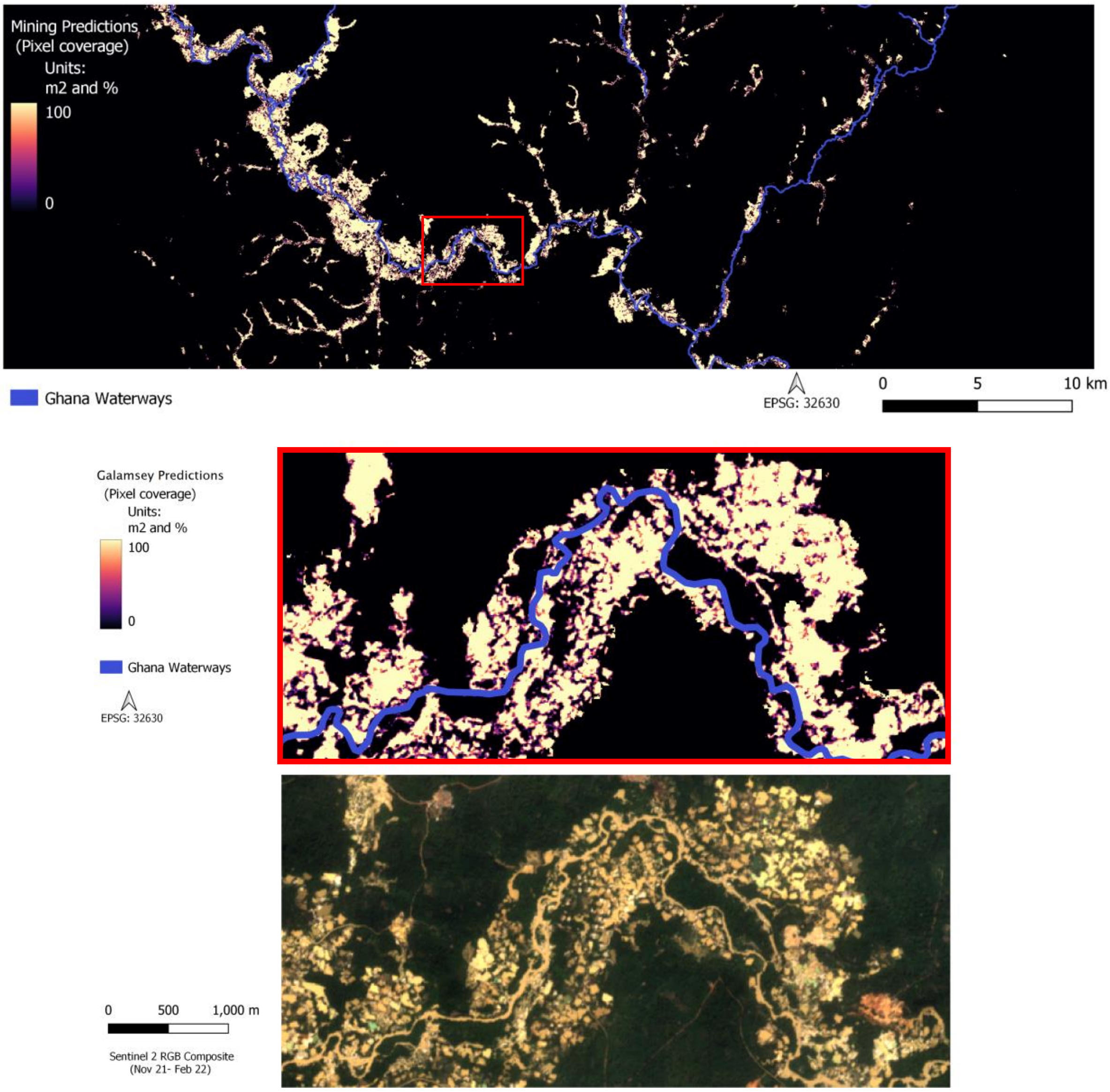



*Figure 14.* Galamsey predictions in Ashanti and Central Regions of Ghana, along the Ofin river.

### 3.6. Accuracy Assessments

When it comes to comparing the model's results, the lack of ground truth data presents a significant challenge. The Galamsey sites that were digitized in the early stages of the analysis are considered ground truth, as they were produced using field-based surveying literature as a reference [20]. Based on this assumption, three test sites not included in the training dataset were separated to assess the accuracy of the image recognition model. The three vector ground truth sites were rasterized at a high resolution using Python and then resampled to match the spatial resolution (10m by 10m) of the predictions. The prediction raster was clipped using the ground truth layer as a mask, and then all the raster data were aligned using the prediction output as a reference. When conducting the quantitative assessment analysis, several metrics were considered and a comparison of the true values and the predicted pixel values was made for each patch. Accuracy assessment results are shown in *Tables 3*, *4*, and *5*. The Mean Squared Error (MSE), Mean Absolute Error (MAE), and

Root Mean Square Error (RMSE) were used to quantify errors of the model, and the TPE values were computed for all testing sites. The TPE represents the difference between the ground truth values and the predicted values, over the true values, in percentage units. The closer to zero the values are, the better the model performs. When evaluating the results, the model had a TPE close to zero in all individual testing sites as well as in the overall approach *(Table 3)*. This is a good TPE result, as it means that the model is neither underestimating nor overestimating the true pixel values. *Tables 4* and *5* show the results of the model's performance using binary metrics, for different pixel sizes. Binary values are an indicator of the model's performance based on the absence and the presence of the target features. ACC is short for "accuracy", and in the 10m pixel size evaluation *(Table 4),* the model predicted Galamsey where there was true Galamsey in almost 90% of the cases. When resampling the raster datasets to 100m, the overall accuracy of the model improved *(Table 5)*, corresponding to a 96 % overall matching of pixels between the true and predicted values, which indicates a high prediction performance. BACC is short for "balanced accuracy", and it measures how good a binary classifier is. PREC is an abbreviation of "precision" and it represents the effectiveness of the model not to label as positive (presence), a sample that is negative (absence). REC is "recall", and it indicates how good is the model at detecting the presence of Galamsey. F1 score is the harmonic mean of REC and PREC values. As a general overview, the accuracy results for this regression-based method are not optimal but still good considering the scarcity of ground truth data in Ghana and the lack of previous similar research studies. Following a qualitative visual inspection *(Figure 15)*, the Galamsey predictions indicate an overall good performance of the model.

| **Area** | **Description** | **MAE** | **MSE** | **RMSE** | **TPE** |
|---|---|---|---|---|---|
| **Site 1** | Regression | 3.51 | 223.02 | 14.93 | -0.04 |
| | Regression Predicted | 0.07 | 0.07 | 0.26 | 0.07 |
| | Regression True | 0.08 | 0.08 | 0.29 | -0.08 |
| **Site 2** | Regression | 10.46 | 606.18 | 24.62 | 0.00 |
| | Regression Predicted | 0.13 | 0.13 | 0.37 | 0.15 |
| | Regression True | 0.06 | 0.06 | 0.24 | -0.06 |
| **Site 3** | Regression | 16.04 | 1061.10 | 32.57 | 0.04 |
| | Regression Predicted | 0.18 | 0.18 | 0.42 | 0.21 |
| | Regression True | 0.05 | 0.05 | 0.21 | -0.05 |
| **All** | Regression | 10.87 | 692.09 | 26.31 | 0.02 |
| | Regression Predicted | 0.15 | 0.15 | 0.39 | 0.18 |
| | Regression True | 0.05 | 0.05 | 0.23 | -0.05 |

*Table 3.* Regression accuracy assessment metrics. Pixel size: 10m.

| | ACC | BACC | PREC | REC | F1 |
|---|---|---|---|---|---|
| **Site 1** | 0.9638 | 0.948 | 0.9304 | 0.9177 | 0.924 |
| **Site 2** | 0.895 | 0.8928 | 0.866 | 0.9445 | 0.9036 |
| **Site 3** | 0.8492 | 0.8208 | 0.8244 | 0.9546 | 0.8847 |
| **All** | 0.895 | 0.8972 | 0.8506 | 0.9462 | 0.8958 |

*Table 4.* Accuracy assessments- Binary metrics. Pixel size: 10m.

| | ACC | BACC | PREC | REC | F1 |
|---|---|---|---|---|---|
| **Site 1** | 0.9615 | 0.9394 | 1 | 0.8788 | 0.9355 |
| **Site 2** | 0.9663 | 0.9601 | 0.9801 | 0.9737 | 0.9769 |
| **Site 3** | 0.9605 | 0.8571 | 0.9562 | 1 | 0.9776 |
| **All** | 0.9625 | 0.9562 | 0.9689 | 0.975 | 0.972 |

*Table 5.* Accuracy assessments- Binary metrics. Pixel size: 100m.

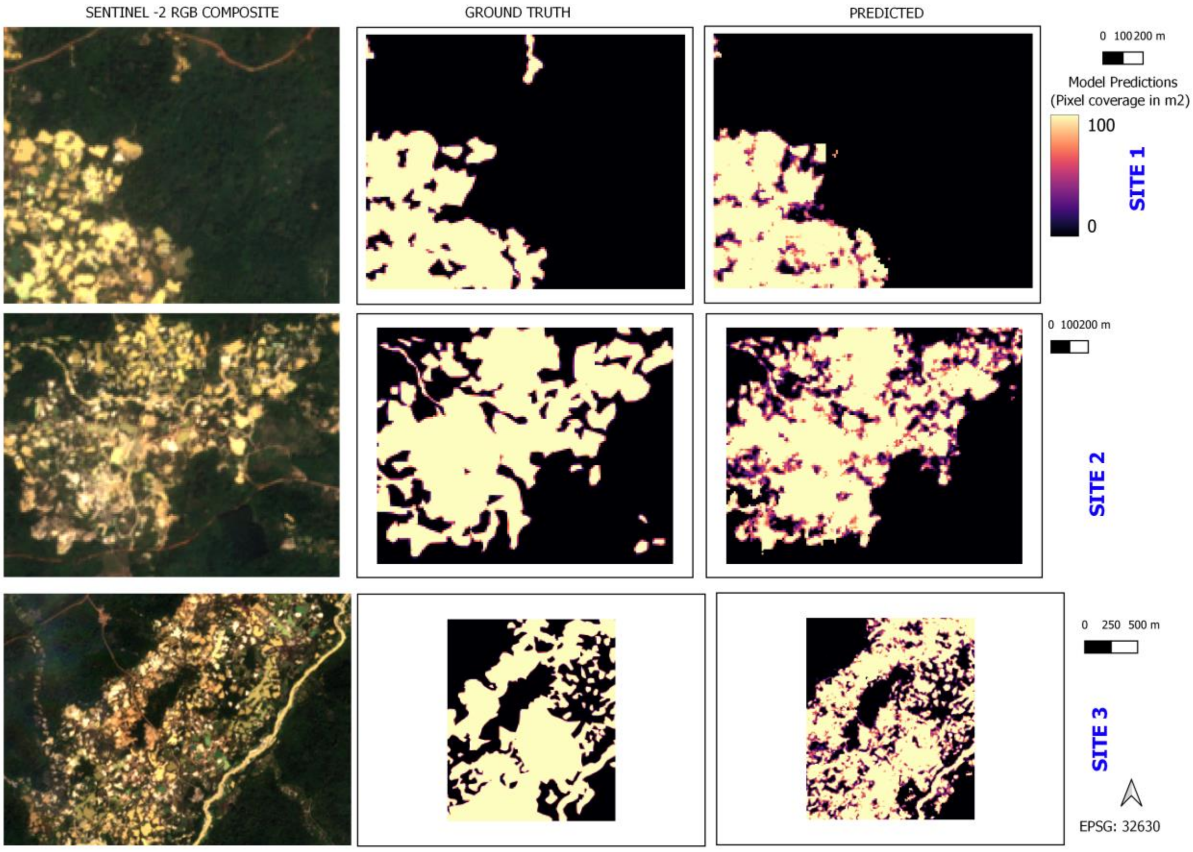
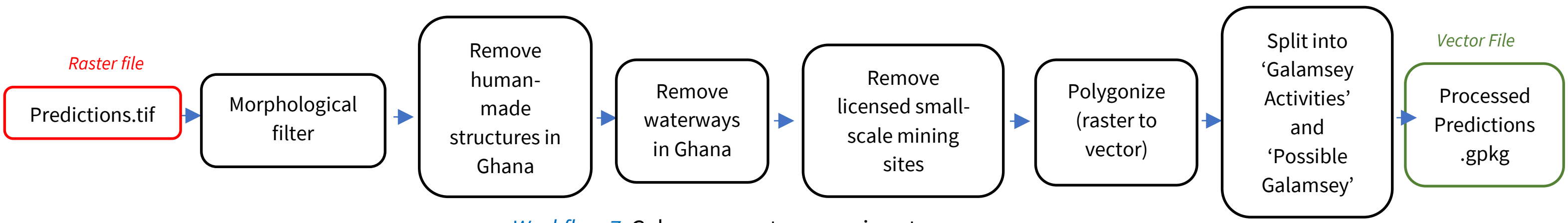



*Figure 15.* Designated ground truth sites and their corresponding raster predictions to assess the model's accuracy.

## 3.7. Data Postprocessing and Final Galamsey Estimations

Using a combination of QGIS, GDAL, and Python scripts, a post-processing procedure was conducted in order to improve the quality of the predictions *(Workflow 7)*.




*Workflow 7.* Galamsey post-processing steps.

A morphological filter was implemented, using Python, in order to sharpen the images and improve feature detection. A 3x3 median filter was applied to remove the unwanted noise of the images, followed by a 3x3 opening filter (eroding and then dilating) and a 5x5 closing filter (dilating and then eroding). The result of this first procedure was a much cleaner prediction map, as this allowed smoothing the boundary detection of features and getting rid of isolated pixels that did not represent the presence of Galamsey.

Because the model was having difficulties differentiating buildings and mining operations in the early stages of the analysis, the following step involved using a raster dataset that consisted of predicted human-made structures per pixel in Ghana (10m by 10m) [86]. This dataset contained the same spatial resolution and the same units as the Galamsey predictions (i.e. pixel values that ranged from 0 to 100, representing the target feature coverage per pixel). This building detection map in Ghana was reclassified into a binary map, assuming that pixel values greater than zero indicated the presence of structures (a value of zero was assigned) and pixels with zero values indicated the absence of structures (a value of 1 was assigned). This binary map was aligned and multiplied by the Galamsey detection map, to obtain a dataset with no human-made structures (i.e. removing the pixels where the model could have confused structures with Galamsey activities).

A 25 m buffer was applied (resulting in a 50m total width) to waterways in Ghana using QGIS, followed by a vector-to-raster conversion (at a 50cm resolution), and a final resampling operation (to match the 10m resolution of the Galamsey predictions) using Python language. After completing this step, a binary raster map was generated, reclassifying values to zero (presence of rivers) and 1 (absence of rivers). This waterways raster map was then multiplied by the Galamsey with no structures raster map previously created, to obtain a prediction map with no rivers.

The Minerals Commission of Ghana [4] made available a dataset with all mining licenses in the country. This dataset was exported as a .kml file and then converted to a vector polygon layer to allow good visualization of the licensed mining sites. Only small and large scale active licenses were considered, and licenses that started after the date on which the satellite imagery was retrieved, were filtered out (i.e. mining licenses that were granted after February 1, 2022). All minerals with the exception of gold were filtered out, selecting only active gold extraction licenses. Since Galamsey operations are intrinsically considered artisanal or small-scale mining, the image recognition model can easily predict illegal mining in areas where there are legal authorization permits, as the spectral signature of these features is the same *(Figure 16)*. For this reason, it was assumed that the pixels that spatially intersected the small-scale licensed polygons were not Galamsey spots, and they were assigned a value of zero (this step was carried out in a similar way to that of rasterizing and resampling the waterways and multiplying the structures in Ghana).

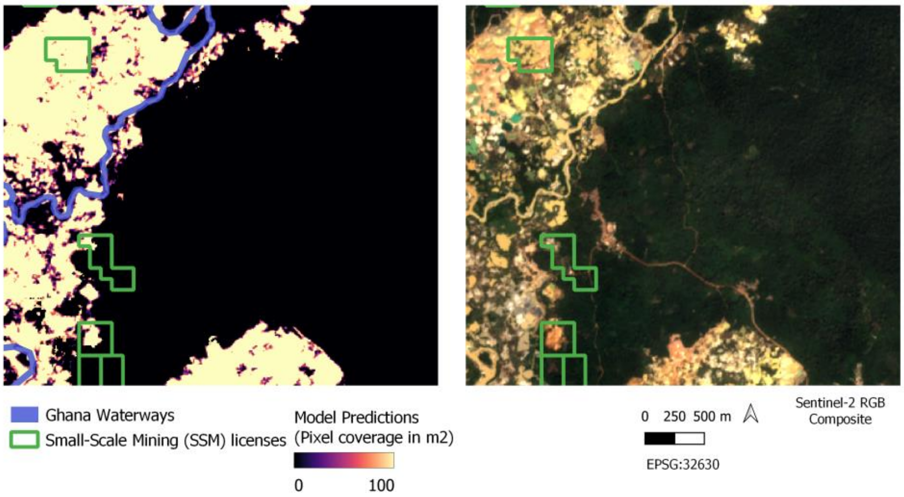



*Figure 16.* Small-scale licensed sites (polygons) and model predictions (10m. by 10m. raster map).

In order to provide a better visual representation of the prediction results, the final post-processing step consisted of vectorizing the raster dataset, setting a threshold value of 1. Pixel values between 1 and 100 were converted into vector format as they indicated the presence of Galamsey activities. Studies have shown that Galamsey miners can clandestinely intrude licensed large-scale mining sites and illegally extract gold [21]. For this reason, large-scale licensed sites could not be ruled out like the small-scale licensed sites when creating the final raster dataset. Following this approach, the Galamsey model predictions that spatially overlapped the legal large-scale mining polygons were considered 'Possible Galamsey in licensed LSM sites', and the rest of the model predictions were classified as 'Galamsey Activities' with a high degree of certainty *(Figures 17 and 18)*.

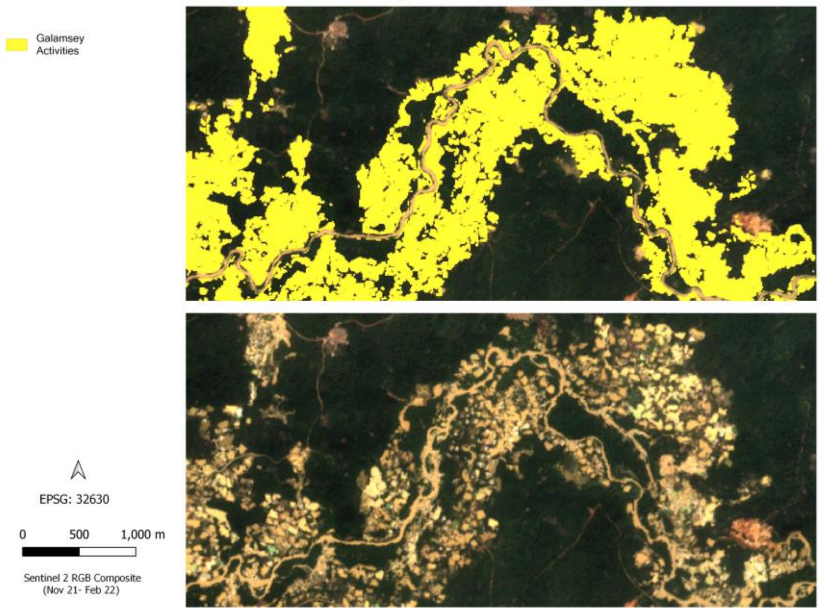



*Figure 17.* Galamsey predictions in Ashanti and Central Regions of Ghana, along the Ofin river (vector format of Figure 14).

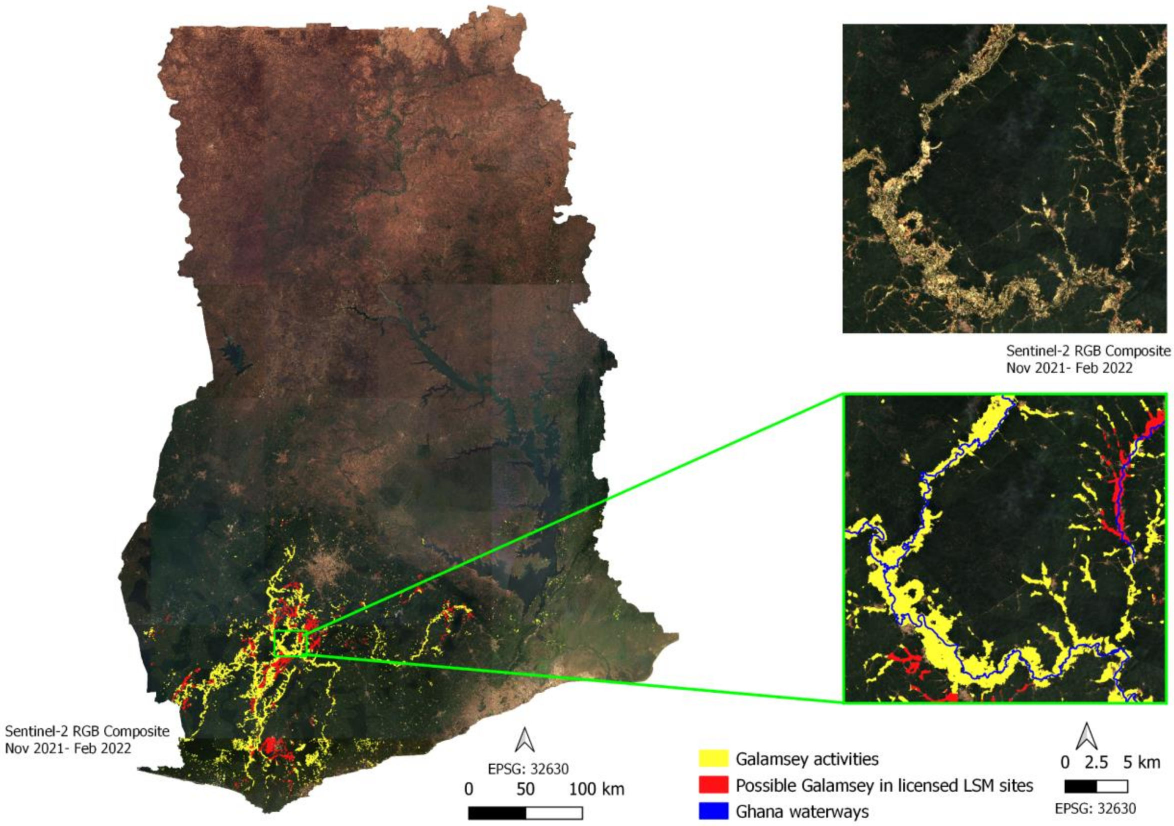



*Figure 18.* November 2021 to February 2022 model predictions: 'Galamsey Activities' and 'Possible Galamsey in licensed LSM sites'.

Results show that illegal mining activities are concentrated in mainly four hotspot regions south of Ghana *(Figure 19)*. 'Possible Galamsey in licensed LSM sites' represent a total area of 72,3 km$^2$, while 'Galamsey Activities' account for 310,4 km$^2$. Western Ghana was the hotspot of illicit operations, comprising 44.2% of the total 'Galamsey Activities' predictions. This is supported by previous studies, as they have confirmed the presence of this unregulated activity in this administrative region [20]. Ashanti is ranked second, accounting for 40.2% of the total area of Galamsey estimations, which represents a spatial coverage of 124.8 km$^2$. Previous findings have also documented the prevalence of this uncontrolled activity in this region, degrading a very important shelterbelt forest reserve protected at a national level [87]. Central Ghana and Eastern Ghana come in second and third place, accounting for 40.2 km$^2$ and 8.1 km$^2$ respectively.

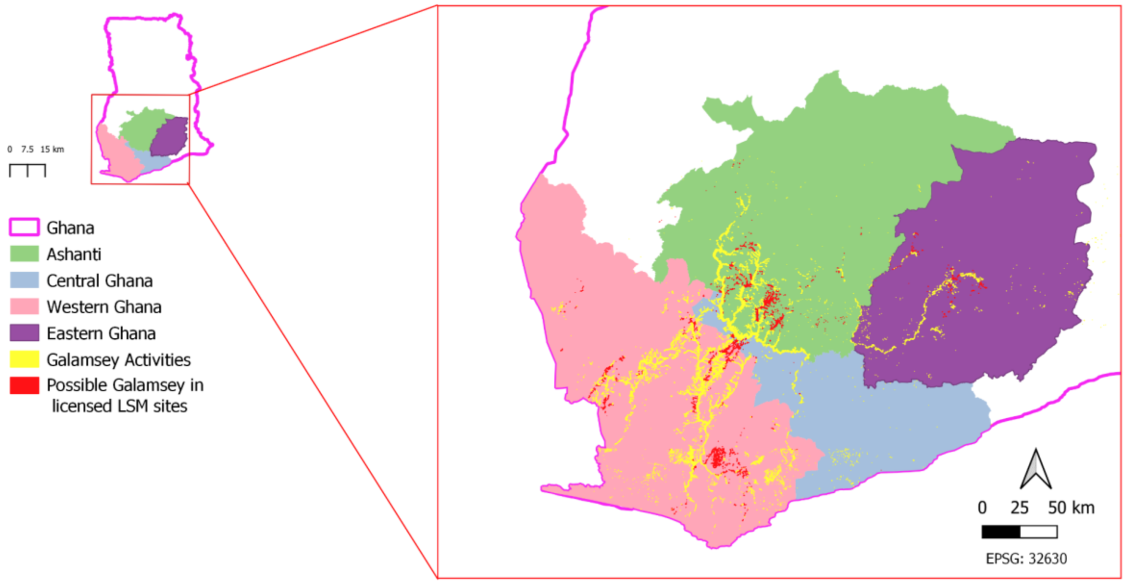



*Figure 19.* Galamsey activities in four districts of the country: Ashanti, Central, Western and Eastern Ghana.

An analysis of the presence of illicit mining operations in national and international protected areas was also conducted, using the WDPA open-source dataset [50]. The presence of Galamsey in national forest reserves is supported by literature [87] and confirmed by the model predictions *(Figure 20)*.

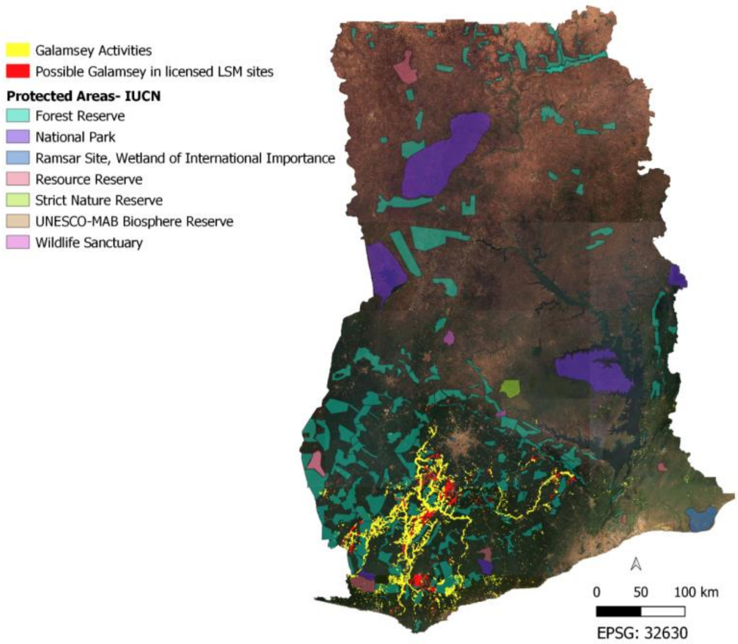



*Figure 20.* Presence of Galamsey in protected forest reserves in Ghana.

The Galamsey estimations show the presence of unlicensed small-scale miners in forest reserves that are under protection, as observed in *Figure 21*. Local online newspapers have corroborated the invasion of Galamseyers in Oda and Fure River forest reserves and photographs documented the presence of unregulated gold extraction activities [88][89]

[90].Galamsey has become a significant concern for Ghanaians: many species of animals live within these protected areas, and they are in danger of losing their habitat, as Galamseyers pose a real threat when extracting gold in this environmentally sensitive region.

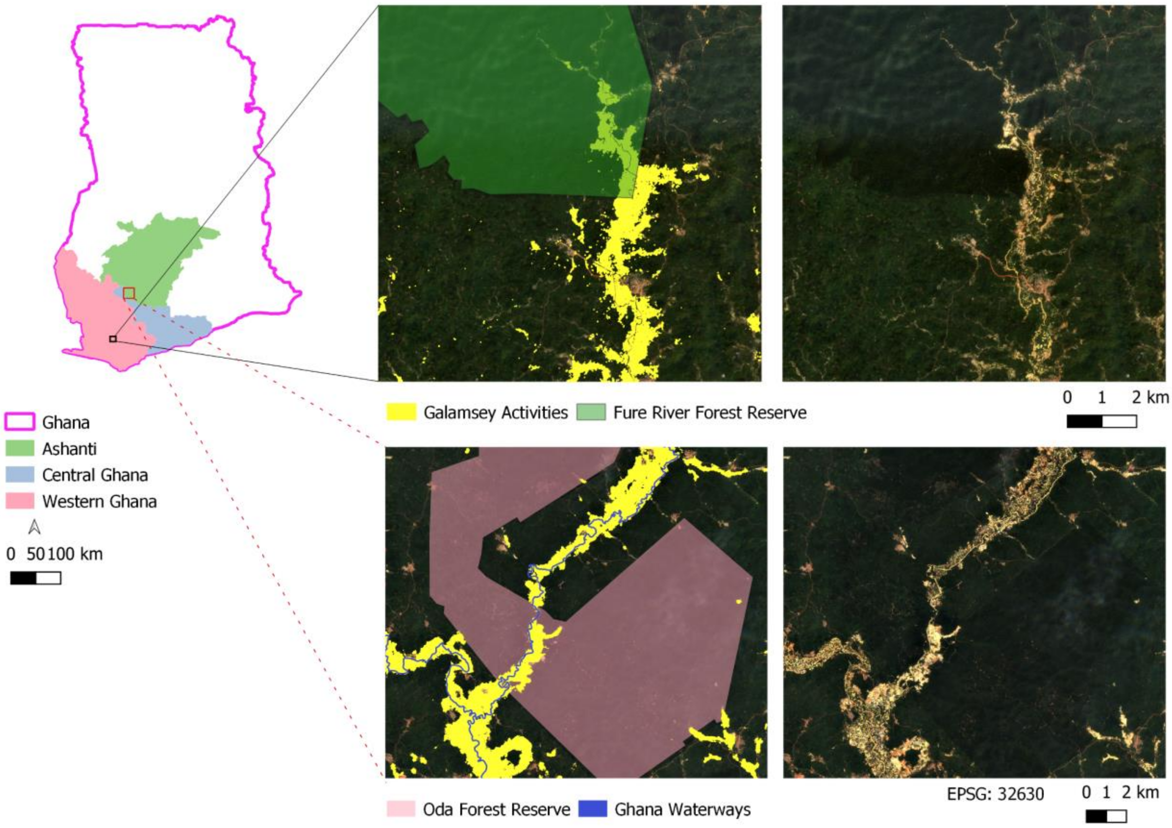



*Figure 21.* 'Galamsey Activities' in two forest reserves that are under national protection in Ghana.

## 4. Discussion

According to the obtained results, it can be concluded with a high degree of certainty that Sentinel 1 and 2 open-source imagery can be used to extract illegal mining features in Ghana. Although the model performed very well in certain hotspot areas, it had some difficulties in the ramification of river networks, as can be observed in *Figure 22*. The images show that the model is underestimating pixel values in those areas, thus not detecting Galamsey features along the stream ramifications. This limitation is probably due to the lack of ground truth data in those areas. Because of the intrinsic nature of these illicit activities, there are no reliable data sources that can be utilized as ground truth, and the creation of training data posed a time constraint when digitizing target features at the early stages of this research.

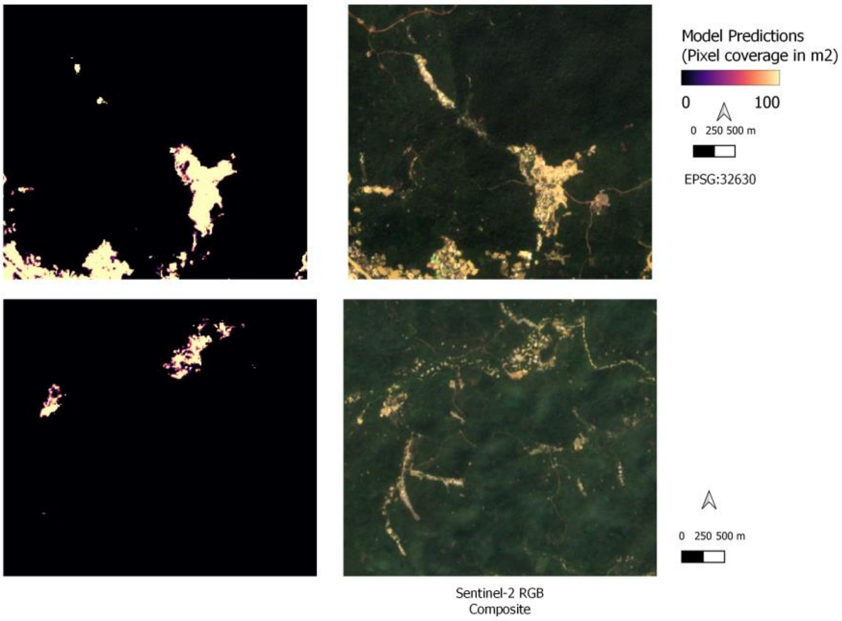



*Figure 22.* Model limitation: Galamsey predictions along different river ramifications, in southern Ghana.

Due to the inherent nature of some SSM clandestine activities, some operations like Underground Galamsey were excluded from the model training and, therefore, could not be detected using remote sensing technology. The creation of target features for the model training was partially guided by field-based surveys (which consisted of point coordinates) that were carried out in Western Ghana [20]. Even though they served as an invaluable guide for visualizing illicit mining activities, this on-site research was limited to the most active gold mining region in Ghana, and the creation of Galamsey features in other regions of the country was based on general spatial characteristics of these unlicensed activities *(Section 3.2.1)*. The model predictions confirm Western Ghana as being the most attractive destination for clandestine SSM operators at the time of this research.

Because the CNN model requires long computational times to be executed, it could not be fully optimized for Galamsey identification. Data augmentation should be considered to reduce overfitting and improve the overall model performance. Given the time constraints and the computer power needed, a thorough testing of the model's hyperparameters and deep learning architecture was difficult. If this method is applied to other image recognition similar tasks, it would be worth training the deep learning algorithm with different neural network architectures to improve the quality of the images and lessen the need for post-processing. Furthermore, further experimentation should include training the model with ground truth data from other critical areas that are threatened by this illicit activity, such as the Amazon rainforest in Brazil and Peru [91][92][93][94].This would make the model more robust and ease the development of a more generic tool that could be applied on a global scale.

While the model initially struggled to distinguish between buildings and mining activities, it greatly improved after more training data was included and some hyperparameters were adjusted, as seen in *Figure 23*. The post-processing procedure of combining a map of Ghanaian structures with the Galamsey predictions was just for the purpose of ensuring that no buildings with similar spectral signatures were included in the final unlicensed small scale identification map, but the predictions remained accurate even without this step.

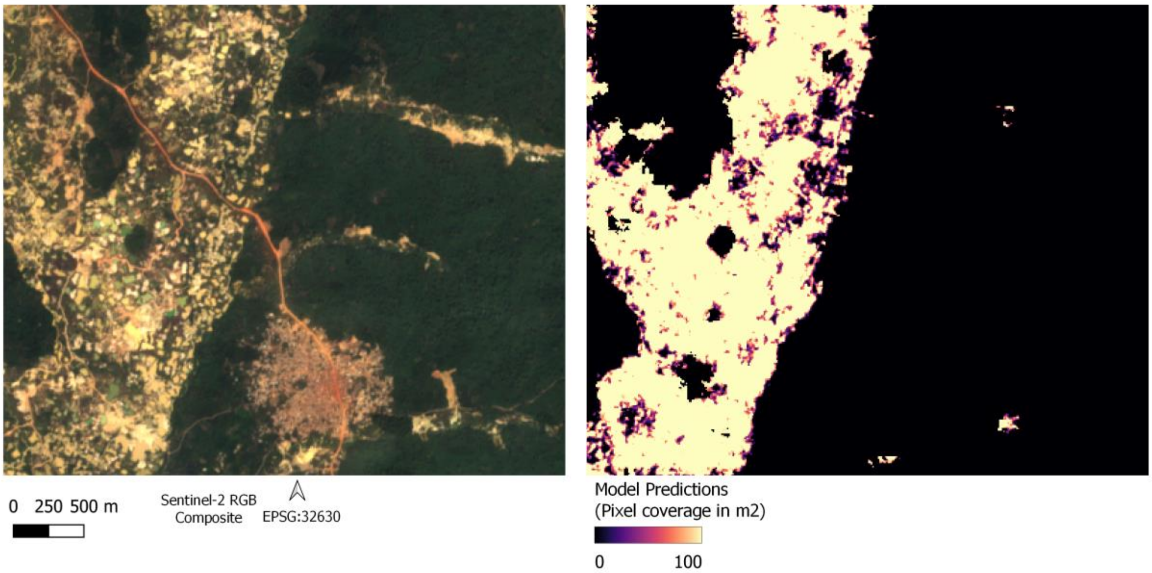



*Figure 23.* Galamsey detection model improvement (it does not recognize structures as target features) in a village south of Kumasi, Ashanti Region, Ghana.

## 5. Conclusion

The use of deep learning methods, more specifically Convolutional Neural Networks on freely available Sentinel images yielded good results in identifying pixels where Galamsey activities are present. This pixel-based regression model suggested a binary accuracy of nearly 90% when computing 10m resolution pixels in the designated validation areas, reaching the highest level of performance when resampling the pixels to 100m (96%).

Model predictions show that Galamsey activities are concentrated in four main administrative regions in Ghana, being Western Ghana the highest producer of gold in the country and the most popular area for illicit artisanal mining operations *(Figure 19)*[20][40][23].This comes with no surprise, as out of the fourteen large-scale gold extraction companies, half of them are located in the Western Region *(Figure 1)*. The machine learning estimations suggest that Galamsey spatial distribution is mainly characterized by clusters along the river ramifications of the country, which is corroborated with previous field-based studies [20]. Predictions also confirm that these unlicensed gold mining operations are taking place in national forest reserves, thereby posing a threat to the country's ecosystem *(Figure 21)*.

Galamsey has become one of the most discussed environmental issues in Ghana and is probably one of the most significant social problems that the country is facing in the 21$^{st}$ century. Galamsey not only degrades the environment at an alarming rate but also poses a major threat to human lives. Illegal artisanal mining is not only limited to countries in Africa as this activity has also spread and been documented in other continents such as South America. Previous studies attempted to locate and analyze the distribution patterns of Galamsey over the past decade[20] [23] [24], but they were limited to a small-scale study area in Ghana or used medium-resolution satellite images. A Galamsey detection high-resolution map (i.e. of at least 10m spatial resolution) in the entire country derived from open-source satellite data has not been undertaken until now. Although there is still room for improvement in the future, this study sheds some light on detecting these dangerous practices through automation, implementing a method that is timely and cost-efficient, and avoids labor-intensive and high-priced work.

Recent advancements in deep learning and neural network technology, as well as the increasing accessibility of open-source satellite imagery, have allowed the development of new approaches in image recognition techniques for the extraction of geographical features. Considering that geodata in developing countries is scarce or nonexistent, this study highlights the importance of machine learning methods to address social and environmental issues that help reach the SDGs.

The use of a machine learning tool to detect Galamsey is therefore of great importance for government authorities and stakeholders involved in enforcing legislation against these illegitimate activities.

The public GitHub repository can be found at: [https://github.com/marltrill/buteo-Galamsey](https://github.com/marltrill/buteo-Galamsey)

## 6. References


[1] Casper Samsø Fibæk, " Buteo: Facilitating EO-Driven Decision Support Systems", https://github.com/casperfibaek/buteo(accessed May 20, 2022).

[2] World Industry & Corporate, "World Gold Council" *Global mine production*, Jun. 16, 2021. https://www.gold.org/ (accessed Feb. 07, 2022).

[3] United States Geological Survey (USGS), "Gold Statistics and Information" https://www.usgs.gov/centers/national-minerals-information-center/gold-statistics-and-information%20 (accessed Feb. 07, 2022).

[4] "Minerals Commission of Ghana", *Government agency of Ghana*, https://www.mincom.gov.gh/ (accessed Jun. 02, 2022).

[5] Richard A. Kuffour, B. M. Tiimub, I. Manu, and W. Owusu, "The Effect of Illegal Mining Activities on Vegetation: A Case Study of Bontefufuo Area in the Amansie West District of Ghana" *East African Scholars Journal of Agriculture and Life Sciences*, vol. 3, no. 11, pp. 353–359, Nov. 2020, doi: 10.36349/easjals.2020.v03i11.002.

[6] A. Aubynn, "Sustainable solution or a marriage of inconvenience? The coexistence of large-scale mining and artisanal and small-scale mining on the Abosso Goldfields concession in Western Ghana" *Resources Policy*, vol. 34, no. 1–2, pp. 64–70, Mar. 2009, doi: 10.1016/j.resourpol.2008.04.002.

[7] "Water Resources Commission", *Government Agency of Ghana*, https://www.wrc-gh.org/ (accessed Feb. 07, 2022).

[8] "Environmental Protection Agency Ghana", *Government Agency of Ghana*, http://www.epa.gov.gh/epa/ (accessed Feb. 07, 2022).

[9] "Forestry Commission", *Government Agency of Ghana*, https://fcghana.org/ (accessed Feb. 07, 2022).

[10] K. D. Kessey and B. Arko, "Small Scale Gold Mining and Environmental Degradation, in Ghana: Issues of Mining Policy Implementation and Challenges" 2013. [Online]. Available: https://www.researchgate.net/publication/282912603

[11] Ghana Talks Business, "Illegal Mining Killing Cocoa Farms" Oct. 08, 2014. https://ghanatalksbusiness.com/2014/10/illegal-mining-killing-cocoa-farms/ (accessed Mar. 02, 2022).

[12] D. Gilbert and O.-B. Albert, "Illegal Small-Scale Gold Mining in Ghana: A Threat to Food Security" *Journal of Food Security*, vol. 4, no. 5, pp. 112–119, 2016, doi: 10.12691/jfs-4-5-2.

[13] D. Ackah, "The Negative Effect of Legal Mining (Gather Them and Sell) in Ghana" *Project Management Scientific Journal*, vol. 1, no. 4, pp. 19–25, 2019, doi: 10.15373/22501991.

[14] I. A. Alhassan, "Galamsey and the Making of a Deep State in Ghana: Implications for National Security and Development" Online, 2014. [Online]. Available: www.iiste.org

[15] A. Donkor, V. Nartey, J. Bonzongo, and D. Adotey, “Artisanal mining of gold with mercury in Ghana” *West African Journal of Applied Ecology*, vol. 9, no. 1, Sep. 2009, doi: 10.4314/wajae.v9i1.45666.

[16] Ghana Web, “Residents worried over abandoned galamsey pits” Sep. 05, 2016. https://www.ghanaweb.com/GhanaHomePage/regional/Residents-worried-over-abandoned-galamsey-pits-467469 (accessed Feb. 07, 2022).

[17] Centre de Ressources sur les Entreprises et les Droits de l’Homme, “Ghana: Abandoned mining pit collapses, killing two suspected illegal miners”, Accessed: Feb. 07, 2022. [Online]. Available: https://www.business-humanrights.org/fr/derni%C3%A8res-actualit%C3%A9s/ghana-abandoned-mining-pit-collapses-killing-two-suspected-illegal-miners/

[18] Ghana Web, “6 dead as galamsey pit caves in” Aug. 30, 2019. https://www.ghanaweb.com/GhanaHomePage/NewsArchive/6-dead-as-galamsey-pit-caves-in-776940 (accessed Mar. 14, 2022).

[19] A. Barenblitt *et al.*, “The large footprint of small-scale artisanal gold mining in Ghana” *Science of the Total Environment*, vol. 781, Aug. 2021, doi: 10.1016/j.scitotenv.2021.146644.

[20] F. Owusu-Nimo, J. Mantey, K. B. Nyarko, E. Appiah-Effah, and A. Aubynn, “Spatial distribution patterns of illegal artisanal small scale gold mining (Galamsey) operations in Ghana: A focus on the Western Region” *Western Region. Heliyon*, vol. 4, p. 534, 2018, doi: 10.1016/j.heliyon.2018.

[21] J. Mantey, F. Owusu-Nimo, K. B. Nyarko, and A. Aubynn, “Operational dynamics of ‘Galamsey’ within eleven selected districts of western region of Ghana” *Journal of Mining & Environment*, vol. 8, no. 1, pp. 11–34, 2017, doi: 10.22044/jme.2016.627.

[22] United Nations, “Sustainable Development Goals (SDGs).” https://sdgs.un.org/goals#goals (accessed Apr. 05, 2022).

[23] G. Forkuor, T. Ullmann, and M. Griesbeck, “Mapping and monitoring small-scale mining activities in ghana using sentinel-1 time series (2015-2019)” *Remote Sensing*, vol. 12, no. 6, Mar. 2020, doi: 10.3390/rs12060911.

[24] NASA AST team, AGRHYMET/SERVIR-WA, A Rocha, and CERSGIS, “Small Scale Mining Portal” http://ssmportal.cersgis.org/map/ (accessed Apr. 29, 2022).

[25] “Copernicus”, *European Union’s Earth observation programme*, https://www.copernicus.eu/en (accessed May 02, 2022).

[26] “NIRAS”, *International Engineering Consultancy Company*, https://www.niras.com/ (accessed May 15, 2022).

[27] “QGIS”, *Quantum Geographic Information Systems*. https://www.qgis.org/ (accessed Dec. 10, 2021).

[28] “HCMGIS- Plugin”, QGIS, Accessed: Feb. 05, 2022. [Online]. Available: https://plugins.qgis.org/plugins/HCMGIS/

[29] "GitHub", *Open Source Hosting Service for Software Development*, Accessed: Jun. 02, 2022. [Online]. Available: https://github.com/

[30] "Python", *Open-Source Programming Language*, Accessed: May 15, 2022. [Online]. Available: https://www.python.org/

[31] "Tensorflow", *Open Source Platform for Machine Learning*, Accessed: May 17, 2022. [Online]. Available: https://www.tensorflow.org/

[32] "Google" https://about.google/ (accessed Feb. 09, 2022).

[33] "Visual Studio Code (VSC)", *Open Source Text Editor for Programming*, Accessed: May 31, 2022. [Online]. Available: https://code.visualstudio.com/

[34] "Microsoft", *Computer Software Company*, https://www.microsoft.com/ (accessed Feb. 22, 2022).

[35] "OpenStreetMap", *Open Source Collaborative Worldwide Geodatabase*, https://www.openstreetmap.org/ (accessed Feb. 21, 2022).

[36] "QuickOSM- Plugin", QGIS, Accessed: Apr. 04, 2022. [Online]. Available: https://plugins.qgis.org/plugins/QuickOSM/

[37] "Geofabrik Download Server", OpenStreetMap (OSM), https://download.geofabrik.de/ (accessed Mar. 15, 2022).

[38] "Google Maps", Google, https://www.google.com/maps (accessed Feb. 20, 2022).

[39] Minerals Commission of Ghana, "Small Scale and Community Mining- Operational Manual" Sep. 2021. Available: https://www.mincom.gov.gh/wp-content/uploads/2021/11/Small-Scale-and-Community-Mining-Operational-Manual-Sep.-2021-1.pdf

[40] Minerals Commission of Ghana, "List of Operating Mines" https://www.mincom.gov.gh/operating-mines/ (accessed May 28, 2022).

[41] Minerals Commission of Ghana, "Ghana Mining Repository- Mining Licenses" https://ghana.revenuedev.org/map (accessed May 03, 2022).

[42] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning" *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015. doi: 10.1038/nature14539.

[43] Cassandra Handan-Nadera, Daniel E. Hoa, and Larry Y. Liuf, *Deep learning with satellite imagery to enhance environmental enforcement*, vol. 10. Elsevier- Data Science Applied to Sustainability Analysis, 2021. doi: 10.1016/C2018-0-02415-9.

[44] C. Robinson, F. Hohman, and B. Dilkina, "A deep learning approach for population estimation from satellite imagery" in *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities, GeoHumanities 2017*, Nov. 2017, pp. 47–54. doi: 10.1145/3149858.3149863.

[45] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network" May 2015, [Online]. Available: http://arxiv.org/abs/1505.00853

[46] J. Chen *et al.*, "Convolution Neural Network Architecture Learning for Remote Sensing Scene Classification" Jan. 2020, [Online]. Available: http://arxiv.org/abs/2001.09614

[47] M. I. Sameen, B. Pradhan, and O. S. Aziz, "Classification of very high resolution aerial photos using spectral-spatial convolutional neural networks" *Journal of Sensors*, vol. 2018, 2018, doi: 10.1155/2018/7195432.

[48] H. Song, Y. Kim, and Y. Kim, "A patch-based light convolutional neural network for land-cover mapping using Landsat-8 images" *Remote Sensing*, vol. 11, no. 2, Jan. 2019, doi: 10.3390/rs11020114.

[49] C. Samsø Fibæk, H. Laufer, C. Keßler, and J. Jokar Arsanjani, "Geodata-driven approaches to financial inclusion – Addressing the challenge of proximity" *International Journal of Applied Earth Observation and Geoinformation*, vol. 99, Jul. 2021, doi: 10.1016/j.jag.2021.102325.

[50] UNEP-WCMC and IUCN, "Protected Planet: The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM)" www.protectedplanet.net (accessed Apr. 29, 2022).

[51] "European Commission", European Union (EU), https://ec.europa.eu/ (accessed Mar. 16, 2022).

[52] "European Space Agency (ESA)", European Union (EU), https://www.esa.int/ (accessed Mar. 16, 2022).

[53] "Sentinel 1", European Space Agency (ESA), https://sentinel.esa.int/web/sentinel/missions/sentinel-1 (accessed Mar. 16, 2022).

[54] "Sentinel 2", European Space Agency (ESA), https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed Mar. 16, 2022).

[55] M. Darvishi *et al.*, "Sentinel-1 and ground-based sensors for continuous monitoring of the Corvara Landslide (South Tyrol, Italy)" *Remote Sensing*, vol. 10, no. 11, Nov. 2018, doi: 10.3390/rs10111781.

[56] I. Elkhrachy *et al.*, "Sentinel-1 remote sensing data and Hydrologic Engineering Centres River Analysis System two-dimensional integration for flash flood detection and modelling in New Cairo City, Egypt" *Journal of Flood Risk Management*, vol. 14, no. 2, Jun. 2021, doi: 10.1111/jfr3.12692.

[57] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review" *Remote Sensing*, vol. 12, no. 14. MDPI AG, Jul. 01, 2020. doi: 10.3390/rs12142291.

[58] S. de Petris, R. Berretti, E. Guiot, F. Giannetti, R. Motta, and E. Borgogno-Mondino, "Detection and characterization of forest harvesting in piedmont through sentinel-2 imagery: A methodological proposal" *Annals of Silvicultural Research*, vol. 45, no. 1, pp. 92–98, 2020, doi: 10.12899/asr-2018.

[59] "Sys", Python Module, [Online]. Available: https://docs.python.org/3/library/sys.html (accessed: Feb. 15, 2022)

[60] "Os", Python Module, [Online]. Available: https://docs.python.org/3/library/os.html (accessed: Feb. 15, 2022)

[61] "Glob", Python Module, [Online]. Available: https://docs.python.org/3/library/glob.html (accessed: Feb. 15, 2022)

[62] "Shutil", Python Module, [Online]. Available: https://docs.python.org/3/library/shutil.html#module-shutil (accessed: May 15, 2022)

[63] Copernicus and ESA, "ONDA" https://www.onda-dias.eu/cms/ (accessed Feb. 15, 2022).

[64] "Scihub", *Copernicus Open Access Hub*. https://scihub.copernicus.eu/dhus/#/home (accessed Dec. 15, 2021).

[65] Copernicus and ESA, "SciHub API" https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/APIHubDescription (accessed Feb. 15, 2022).

[66] "STEP ESA" *Science Toolbox Exploitation Platform*. http://step.esa.int/main/doc/tutorials/ (accessed Dec. 20, 2021).

[67] European Space Agency (ESA), "SNAP" *Sentinel Application Platform*. https://step.esa.int/main/download/snap-download/ (accessed Oct. 20, 2021).

[68] "SkyWatch" https://www.skywatch.com/ (accessed Mar. 16, 2022).

[69] The Open Source Geospatial Fundation, "GDAL" https://www.osgeo.org/ (accessed May 16, 2022).

[70] "gdal_translate", GDAL Program, [Online]. Available: https://gdal.org/programs/gdal_translate.html (accessed: Feb. 17, 2022)

[71] "gdalwarp", GDAL Program, [Online]. Available: https://gdal.org/programs/gdalwarp.html (accessed: Feb. 17, 2022)

[72] R. W. H. Rouse, J. A. W. Haas, and D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS" *NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Sect. A. Geophysics*, vol. 1, Jan. 1974.

[73] "Normalized Difference Moisture Index (NDMI)", Sentinel-2, [Online]. Available: https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndmi/ (accessed: May 05, 2022)

[74] "NumPy" *User Guide*. https://numpy.org/doc/stable/user/index.html (accessed Oct. 15, 2021).

[75] D. Misra, C. Crispim-Junior, and L. Tougne, "Patch-based CNN evaluation for bark classification Patch-based CNN evaluation for bark" [Online]. Available: https://hal.archives-ouvertes.fr/hal-02969811v2

[76] "numpy.savez", NumPy- Python Library, [Online]. Available: https://numpy.org/doc/stable/reference/generated/numpy.savez.html (accessed: Feb. 19, 2022)

[77] "Keras", *Python deep learning API*, [Online]. Available: https://keras.io/ (accessed: May 05, 2022)

[78] Microsoft, "Microsoft researchers win ImageNet computer vision challenge" Dec. 10, 2015. https://blogs.microsoft.com/ai/microsoft-researchers-win-imagenet-computer-vision-challenge/ (accessed Apr. 11, 2022).

[79] Microsoft, "Microsoft Research" https://www.microsoft.com/en-us/research/ (accessed May 05, 2022).

[80] C. S. Fibæk, C. Keßler, and J. J. Arsanjani, "A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning" *International Journal of Applied Earth Observation and Geoinformation*, vol. 105, Dec. 2021, doi: 10.1016/j.jag.2021.102628.

[81] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning " *31st AAAI Conference on Artificial Intelligence*, pp. 4278–84, 2017. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806

[82] M. R. Zhang, J. Lucas, G. Hinton, and J. Ba, "Lookahead Optimizer: k steps forward, 1 step back" *Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver, Canada.*, Jul. 2019, [Online]. Available: https://github.com/michaelrzhang/lookahead.

[83] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization" *3rd International Conference on Learning Representations (ICLR) 2015 Conference Paper*, Dec. 2015, [Online]. Available: http://arxiv.org/abs/1412.6980

[84] "EarlyStopping-Keras", Python Library, [Online]. Available: https://keras.io/api/callbacks/early_stopping/ (accessed: Apr. 16, 2022)

[85] S. Lathuiliere, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A Comprehensive Analysis of Deep Regression" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2065–2081, Sep. 2020, doi: 10.1109/TPAMI.2019.2910523.

[86] C. S. Fibæk, C. Keßler, J. J. Arsanjani, and M. L. Trillo, "Creating Globally Applicable Population Estimates from Sentinel Data (Accepted- not yet published)" *Transactions in GIS*, no. Methods, Models, and Resources for Geospatial Knowledge Graphs and GeoAI.Special Issue, 2022.

[87] S. Boadi, C. A. Nsor, O. O. Antobre, and E. Acquah, "An analysis of illegal mining on the Offin shelterbelt forest reserve, Ghana: Implications on community livelihood" *Journal of Sustainable Mining*, vol. 15, no. 3. Central Mining Institute in Katowice, pp. 115–119, Jan. 01, 2016. doi: 10.1016/j.jsm.2016.12.001.

[88] Graphic Online, "Galamsey operators invade Oda River Forest Reserve" Jan. 18, 2020. https://www.graphic.com.gh/news/general-news/galamsey-operators-invade-oda-river-forest-reserve.html (accessed May 29, 2022).

[89] Daily Guide Network, "Oda River Forest Reserve Under Siege" Jan. 17, 2020. https://dailyguidenetwork.com/oda-river-forest-reserve-under-siege/ (accessed May 29, 2022).

[90] Citi Newsroom (CNR), “Lands Minister blames destruction of Fure Forest Reserve on lack of guards” Feb. 27, 2020. https://citinewsroom.com/2020/02/lands-minister-blames-destruction-of-fure-forest-reserve-on-lack-of-guards/ (accessed May 29, 2022).

[91] Nature, “Illegal mining in the Amazon hits record high amid Indigenous protests” Sep. 30, 2021. https://www.nature.com/articles/d41586-021-02644-x (accessed May 29, 2022).

[92] Mining Technology, “Illegal gold in the Brazilian Amazon” Mar. 10, 2022. https://www.mining-technology.com/analysis/amazon-illegal-gold-mining/ (accessed May 29, 2022).

[93] National Aeronautics and Space Administration (NASA), “Teaming up to Track Illegal Amazon Gold Mines in Peru” Jun. 16, 2021. https://www.nasa.gov/feature/teaming-up-to-track-illegal-amazon-gold-mines-in-peru (accessed May 29, 2022).

[94] Panoramas Scholarly Platform, “Saving the Amazon: Peru Intervenes with Illegal Mining” Mar. 28, 2019. https://www.panoramas.pitt.edu/saving-amazon-peru-intervenes-illegal-mining#:~:text=The%20southeast%20region%20of%20Madre,Amazon%20Conservation%20Association%2C%202019 (accessed May 29, 2022).