GEOSPATIAL INTELLIGENCE ENHANCEMENT USING ADVANCED DATA SCIENCE AND MACHINE LEARNING: A SYSTEMATIC LITERATURE REVIEW

Vinoth kumar Kolluru, Sudeep Mungara, Advaitha Naidu Chintakunta, Lokesh Kolluru, Charan Sundar Telaganeni

ABSTRACT

In the era of rapid advancements in artificial intelligence, the geospatial field is experiencing transformative changes. Traditional methods for land cover classification and anomaly detection have often been inconsistent and inaccurate, leading to significant real-world issues such as resource misallocation, unnoticed illegal activities like deforestation, unmonitored topographical changes such as unauthorized constructions, unattended forest fires, and border fence crises, all of which exacerbate climate change and urbanization challenges. This study systematically explores various machine learning (ML) techniques and their application to publicly available geospatial datasets. Specifically, it compares selected Convolutional Neural Networks (CNNs) and other ML models on these datasets to evaluate multiple performance metrics and conduct a comparative analysis. While numerous ML models have been previously employed for land cover classification and anomaly detection, this review seeks to enhance performance metrics and improve classification accuracy. Prior studies have employed techniques such as Random Forest on Sentinel-2 data (Gromny et al., 2019), multiple regression approaches on Landsat data (Wu et al., 2016), and Principal Component Analysis (PCA) on OpenStreetMap data (Feldmeyer et al., 2020). Our study introduces the application of advanced models like VGG16, U-Net, and Isolation Forest to geospatial datasets, assessing their impact on enhancing land cover classification and anomaly detection. This research not only aims to achieve higher classification accuracy but also contributes to the field by providing insights into the effectiveness of these models and proposing future directions and opportunities.

1. INTRODUCTION

Geospatial analysis, particularly land cover classification and anomaly detection, has become increasingly vital due to its extensive applications in environmental monitoring, urban planning, and national security. Inaccuracies and inconsistencies in these processes can lead to severe problems, such as resource misallocation, which is a significant global challenge. Traditional machine learning (ML) methods have often contributed to this misallocation, adversely affecting agriculture, forestry, urban planning, and humanitarian aid sectors. Moreover, inaccuracies in these methods can facilitate deforestation and unauthorized construction, leading to ecosystem disruptions and negatively impacting local communities.

David C. Wyld et al. (Eds): AISO, SIP, SOFTFM, NLDM, CRBL, BIOM, COMIT– 2024
pp. 23-41, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.141502

Unchecked illegal construction results in topographical changes, complicating urban planning efforts and potentially leading to urban sprawl. The inability of traditional models to detect the early stages of forest fires can escalate into natural disasters, severely affecting wildlife habitats and human settlements. Additionally, inaccurate monitoring at border fences, crucial for national security, can hinder the identification of potential security breaches, further exacerbating national security concerns.

Forest inventory and analysis in the United States rely on remote sensing and geospatial technologies to improve data accuracy and support resource management (Nelson et al., 2007). Geospatial problems manifest in various applications, with artificial intelligence (AI) at the forefront of technological advancements, enhancing model enrichment and transformational processes. These advancements often require the integration of various ML techniques (Breunig, 2020). This systematic literature review explores the effectiveness of advanced ML techniques on publicly available geospatial datasets, specifically focusing on enhancing the accuracy and efficiency of land cover classification and anomaly detection in geo-imagery. The critical question addressed is: How can advanced models like VGG16, U-Net, and Isolation Forest improve geospatial data analysis?

The primary objective is to conduct a comparative analysis of VGG16, U-Net, and Isolation Forest on geospatial datasets such as Sentinel-2, Landsat, and OpenStreetMap. These ML models are evaluated based on multiple performance metrics to determine their effectiveness in classifying land cover and detecting anomalies, tasks previously handled by traditional models. While various ML techniques have been employed for these purposes in past studies, this systematic review aims to explore the comparative effectiveness of advanced models to achieve higher classification accuracy and more precise land cover classification and anomaly detection. For instance, Gromny et al. (2019) applied Random Forest for land cover classification with Sentinel-2 data, Wu et al. (2016) used five-regression approaches for biomass estimation with Landsat imagery, and Feldmeyer et al. (2020) employed PCA techniques to generate socio-economic indicators from OpenStreetMap data. This study aims to surpass these performances by leveraging the capabilities of VGG16, U-Net, and Isolation Forest.

This study enhances geospatial intelligence (GEOINT) in several keyways:

- **Application of Advanced Models**: It demonstrates the application of VGG16, U-Net, and Isolation Forest on geospatial datasets, highlighting their capabilities and limitations.
- **Comprehensive Model Evaluation**: It evaluates the effectiveness of these models in improving the accuracy of land cover classification and the precision of anomaly detection, contributing to more reliable geospatial data analysis.
- **Future Research Directions**: It identifies potential improvements and future research opportunities in applying ML techniques to geospatial data, encouraging ongoing innovation and refinement in the field.

By integrating insights from key studies in geospatial data management and machine learning, including Breunig et al. (2020), Praveen et al. (2016), and Kiwelekar et al. (2020), this systematic review provides a comprehensive analysis of current methodologies. This review sets the stage for future advancements in geospatial intelligence, showing promising results in the traditional enhancement of GEOINT tasks (Breunig, 2020). Traditional methods like GNSS and LSTM have demonstrated that a substantial portion of data is geographic (Morais, 2012), underscoring the importance of continuous improvement in geospatial analysis techniques.

1.2. Background

GEOINT (Geospatial Intelligence) methods and the promising potential of advanced machine learning (ML) techniques for geospatial tasks are transforming the field. This research examines the collection, analysis, and interpretation of data related to Earth's surface and its activities. Traditionally, GEOINT studies have relied heavily on manual processing, leading to significant performance limitations. While effective, these traditional methods are labor-intensive and often require a multi-step, time-consuming approach. The advent of advanced data science and ML techniques has ushered in a new era of GEOINT capabilities, enabling more efficient and accurate analysis of vast amounts of geospatial data.

This research highlights the importance of geospatial intelligence and incorporates various data fusion techniques. Low and high-level data fusion, such as the combination of multiple information sources via Kalman Filter and Bayesian Networks, address sustainable development challenges (Kussul et al., 2015). By integrating these techniques, the study aims to enhance the accuracy and efficiency of GEOINT analyses.

Machine learning, particularly deep learning, has shown remarkable potential in processing and analyzing geospatial data. Techniques such as Convolutional Neural Networks (CNNs) have revolutionized image classification and object detection tasks, enabling the automatic identification and classification of objects in satellite images with high accuracy. For instance, VGG16 and U-Net models excel at delineating features in satellite imagery, facilitating detailed and precise analysis. Additionally, novel methods for generating geospatial intelligence from social media posts, as proposed by Sufi and Alsulami, provide a unique approach to analyzing these data activities using Named Entity Recognition (NER) techniques (Sufi and Alsulami, 2022).

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have further advanced the field by enabling the analysis of temporal sequences in geospatial data. This capability is crucial for monitoring changes over time, providing insights into dynamic environmental and urban processes. The intersection of competitive intelligence and geospatial intelligence offers strategic advantages across various domains (Othenin-Girard, Caron, and Guillemette, 2011).

Semantic segmentation, which involves labeling each pixel in an image with a class, benefits significantly from models like U-Net and SegNet. These models excel at delineating features in satellite imagery, enhancing the detail and analytical capabilities of geospatial analyses. Additionally, anomaly detection methods such as Isolation Forest and One-Class SVM provide robust tools for identifying unusual patterns in geospatial data. These methods are essential for applications like environmental monitoring and disaster management.

The integration of these advanced methodologies into GEOINT is facilitated by the availability of extensive open-source datasets. Sentinel-2 and Landsat satellite images, provided by programs like Copernicus and the US Geological Survey (USGS), offer high-resolution, multispectral data invaluable for detailed geospatial analysis. OpenStreetMap (OSM) provides comprehensive vector data, which can be used in conjunction with raster data for more holistic GEOINT analyses. The comprehensive review by Gao (2021) highlights the transformative impact of geospatial artificial intelligence (GeoAI) on traditional GEOINT analysis, demonstrating the potential for significant advancements in this field.

Moreover, the integration of big data analytics with geospatial data offers opportunities for real-time processing and analysis, enabling timely decision-making in critical situations such as disaster response and urban planning. Techniques like data augmentation and transfer learning can further enhance model performance by leveraging pre-existing knowledge from related tasks, thus reducing the need for extensive labeled datasets.

The growing accessibility of cloud computing resources also facilitates the scalable processing of large geospatial datasets, enabling more complex and computationally intensive analyses. Cloud platforms like Google Earth Engine and Amazon Web Services (AWS) provide powerful tools for geospatial data processing and analysis, making advanced GEOINT capabilities more accessible to researchers and practitioners.

In conclusion, the integration of advanced machine learning techniques into geospatial intelligence holds immense potential for improving the accuracy and efficiency of land cover classification and anomaly detection. By leveraging the capabilities of models like VGG16, U-Net, and Isolation Forest, and utilizing extensive open-source datasets, this research aims to push the boundaries of traditional GEOINT methods and pave the way for future advancements in the field.

2. Related Work

The integration of traditional data science and machine learning (ML) techniques, such as Long Short-Term Memory (LSTM) and Global Navigation Satellite Systems (GNSS), into geospatial intelligence (GEOINT) has significantly expanded potential applications and capabilities. These advancements have impacted areas such as imagery and remote sensing, Geographic Information Systems (GIS) data, and Open-Source Intelligence (OSINT) (Breunig et al., 2020). Breunig et al. (2020) also discuss ongoing progress in GEOINT data management, highlighting the importance of advanced methodologies in addressing future challenges. Praveen et al. (2016) emphasizes the role of big data environments in enhancing geospatial data analysis, providing a solid foundation for sophisticated ML applications. Additionally, adaptive learning systems are leveraging ML methods to provide customized educational experiences (Kolluru, Mungara, and Chintakunta). Kiwelekar et al. (2020) review deep learning techniques for geospatial data analysis, highlighting their effectiveness in tasks such as image classification and object detection.

The transformative impact of artificial intelligence on geospatial intelligence is further explored by Dold and Groopman (2017), who discuss prospects and the potential of AI in this domain. Kussul et al. (2015) demonstrate the practical benefits of integrating GEOINT and data fusion techniques for sustainable development, showcasing the effectiveness of advanced analytical methods.

Recent studies have explored novel applications of GEOINT, such as the method introduced by Sufi and Alsulami (2022) for generating geospatial intelligence from social media posts, illustrating the diverse data sources that can be leveraged using advanced techniques. Othenin-Girard et al. (2011) discuss the strategic advantages of integrating competitive intelligence with geospatial intelligence, emphasizing the synergy between these domains.

The emergence of Geospatial Artificial Intelligence (GeoAI) has led to significant advancements in the field. Temporal land use analyses in the Uttara Kannada District reveal substantial forest fragmentation, with evergreen forest cover decreasing from 57.31% in 1979 to 32.08% in 2013 (Ramachandra et al., 2016). Gao (2021) and Gao (2020) provide comprehensive overviews of GeoAI, reflecting on recent advancements and potential applications. VoPham et al. (2018) highlight GeoAI's potential for environmental epidemiology, enhancing public health research through sophisticated spatial analysis.

Efficiency improvements in energy conversion are demonstrated by the combined calculation of thermoelectric modules (UG et al.), while Döllner (2020) explores the innovative applications of GeoAI in urban planning and infrastructure management using 3D point clouds and geospatial digital twins. Gramacki et al. (2023) emphasize the need for standardized frameworks to advance GeoAI, underscoring the importance of unified approaches in the field. Roussel and Böhm (2023) review geospatial explainable AI (XAI), focusing on the transparency and interpretability of AI models in GEOINT applications, which is critical for enhancing trust and reliability in these systems.

Mai et al. (2023) discusses the opportunities and challenges of foundation models for GEOINT, proposing directions for future research. Lunga et al. (2022) highlight the significance of GeoAI at the ACM SIGSPATIAL conference, identifying it as a new frontier in geospatial research. The potential of ML for 3D point clouds and geospatial digital twins is well-documented, emphasizing the importance of these technologies in modern GEOINT (Döllner, 2020). Standardizing geospatial AI is essential for advancing the field, as discussed in the ACM SIGSPATIAL workshop proceedings (Gramacki et al., 2023).

The transformative impact of advanced data science and ML techniques on geospatial intelligence is evident in the collective findings of these studies. Leveraging cutting-edge methodologies and diverse data sources, this research illustrates the enhanced capabilities and new opportunities available in GEOINT, setting the stage for continued innovation and development in this dynamic field. This comprehensive review serves as a crucial reference for researchers and practitioners aiming to push the boundaries of geospatial intelligence and its applications.

2.2. Overview and Historical Importance

Geospatial Intelligence (GEOINT) has a rich history that dates back to traditional methods such as mapmaking, photogrammetry, and remote sensing. Initially, GEOINT heavily relied on manual processes and basic analytical methods, making the tasks time-consuming and labor-intensive. The introduction of satellite imagery in the mid-20th century was a groundbreaking development, providing unprecedented access to spatial data and transforming the way geographic information was gathered and analyzed. This shift laid the foundation for the integration of more advanced technologies.

The advent of computer technology in the 1980s and 1990s significantly advanced GEOINT capabilities by enabling sophisticated data processing and analysis. Geographic Information Systems (GIS) emerged as essential tools for managing and visualizing spatial data, supporting applications ranging from city planning to environmental surveillance. These technological advancements revolutionized the field, allowing for more comprehensive and detailed geospatial analyses.

The turn of the century brought significant advancements in data science and machine learning techniques, marking a new era for GEOINT. These technologies have dramatically improved the efficiency and accuracy of handling large amounts of geospatial data. Machine learning models, including various learning algorithms, have shown impressive potential in automating complex tasks such as image categorization, object identification, and semantic segmentation. The integration of geospatial analysis with urban geometry provides valuable insights into the structural dynamics of urban environments (Pagliardini et al., 2010).

Convolutional Neural Networks (CNNs) have revolutionized image analysis, facilitating object detection and classification within satellite images. Innovations such as U-Net and SegNet have further advanced the field by offering sophisticated techniques for image segmentation, which is crucial for detailed feature extraction. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have enhanced temporal analysis capabilities, enabling the tracking of changes over time across diverse geospatial datasets (Saxena et al., 1997). These advancements have been pivotal in various applications, including urban planning, disaster management, and environmental monitoring (Gao, 2021; Gao, 2020).

The importance of these technological advancements cannot be overstated. Advanced geospatial analysis and modeling are essential for understanding and managing urban structure dynamics (Jiang and Yao, 2010). This research has unlocked new opportunities for applications in national security, disaster response, and environmental surveillance. For instance, in disaster response, the ability to quickly analyze satellite images and identify irregularities has significantly improved response times and efficiency. Geospatial technologies are also pivotal in remapping border areas, enhancing the understanding of border dynamics and related activism (Walsh, 2013).

The integration of machine learning in combating misinformation has also played a crucial role in promoting accurate news consumption in the digital era (Kolluru, Mungara, and Chintakunta). The emergence of Geospatial Artificial Intelligence (GeoAI) has revolutionized the understanding and utilization of geospatial data. GeoAI combines AI technologies with geospatial data to provide powerful tools for predictive analysis, decision-making, and real-time monitoring. Research on the geospatial modeling of the US-Mexico border, for example, highlights the "funnel effect," showing increased mortality rates and the impact of surveillance on migration patterns (Chambers et al., 2021).

Recent studies have explored novel applications of GEOINT. Sufi and Alsulami (2022) introduced methods for generating geospatial intelligence from social media posts, illustrating the diverse data sources that can be leveraged using advanced techniques. The intersection of competitive intelligence and geospatial intelligence has been explored, demonstrating strategic advantages in various domains (Othenin-Girard et al., 2011). The potential of GeoAI for environmental epidemiology, as discussed by VoPham et al. (2018), underscores its ability to enhance public health research through sophisticated spatial analysis.

The integration of big data analytics with geospatial data offers real-time processing and analysis opportunities, enabling timely decision-making in critical situations. Cloud computing resources have further facilitated the scalable processing of large geospatial datasets, making advanced GEOINT capabilities more accessible to researchers and practitioners (Mai et al., 2023). The significance of GeoAI in geospatial research was highlighted at the ACM SIGSPATIAL conference, identifying it as a new frontier in the field (Lunga et al., 2022).

In conclusion, the systematic integration of advanced data science and ML techniques into GEOINT has had a transformative impact. Leveraging cutting-edge methodologies and diverse data sources, this research illustrates the enhanced capabilities and new opportunities available in GEOINT. The comprehensive review of current methodologies and future directions sets the stage for continued innovation and development in this dynamic field.

2.3. Datasets and Subject ML models

The effectiveness of geospatial intelligence (GEOINT) heavily relies on the quality and diversity of the datasets used, as well as the advanced machine learning (ML) models applied to analyze these data. This section provides an overview of the key datasets, and the ML models commonly employed to enhance GEOINT capabilities.

Key Datasets

A. Sentinel-2 Satellite Images

- **Source**: Copernicus Program
- **Description**: Sentinel-2 provides high-resolution optical imagery at spatial resolutions of 10m, 20m, and 60m. The multispectral data includes 13 spectral bands, making it suitable for a wide range of applications such as land cover classification, vegetation monitoring, and disaster management. The high temporal resolution (revisiting every 5 days) ensures timely and accurate data collection.
- **Use Cases**: Applications include monitoring agricultural practices, detecting changes in urban areas, and assessing natural disasters like floods and wildfires. Sentinel-2's high-resolution imagery is crucial for precise and timely geospatial analysis (Drusch et al., 2012).

B. Landsat Satellite Images

- **Source**: United States Geological Survey (USGS)
- **Description** The Landsat program provides both historical and current multispectral data with spatial resolutions ranging from 15m to 60m. Since the 1970s, Landsat has been invaluable for temporal analysis and change detection, offering a long-term dataset that supports extensive environmental monitoring.
- **Use Cases**: Key applications include long-term environmental monitoring, urban expansion analysis, and forestry management. Landsat's extensive historical data archive is essential for studying environmental changes over time and understanding long-term trends (Wulder et al., 2012).

C. OpenStreetMap (OSM)

- **Source**: OpenStreetMap Community
- **Description**: OpenStreetMap (OSM) provides a comprehensive, crowd-sourced vector dataset of global geographic features, including roads, buildings, and land use. Continuously updated by a global community, OSM is a rich source of contextual geospatial information.
- **Use Cases**: Applications include urban planning, navigation and routing services, and disaster response planning. The detailed and up-to-date nature of OSM data supports a wide range of geospatial analyses, enhancing the accuracy and relevance of GEOINT projects (Haklay and Weber, 2008).

Subject ML Models

A. Convolutional Neural Networks (CNNs)

- Models: AlexNet, VGGNet, ResNet, Inception
- Applications: CNNs are pivotal for image classification and object detection tasks. These models excel at recognizing patterns and features in satellite images, making them ideal for identifying buildings, roads, natural features, and land cover types. CNNs have demonstrated significant success in automating complex geospatial tasks, improving both efficiency and accuracy (Krizhevsky, Sutskever, and Hinton, 2012; Simonyan and Zisserman, 2015; He et al., 2016; Szegedy et al., 2015).

Model	Description	Application
AlexNet	Early CNN model with five convolutional layers	Basic image classification
VGGNet	Deep CNN with uniform layer structure	High-precision image recognition
ResNet	CNN with residual learning framework	Complex image classification tasks
Inception	CNN with inception modules for efficiency	Large-scale image analysis

A. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks

- Applications: Temporal sequence analysis. RNNs and LSTMs are particularly useful for analyzing time-series geospatial data, enabling the monitoring of changes and trends over time.

Model	Description	Application
RNN	Network for sequential data processing	Analyzing temporal sequences
LSTM	RNN variant that handles long-term dependencies	Monitoring environmental changes

B. U-Net and SegNet

- Applications: Semantic segmentation. These models are designed to label each pixel in an image, providing detailed classification maps essential for tasks like land cover mapping and infrastructure delineation.

Model	Description	Application
U-Net	Encoder-decoder architecture with skip connections	Detailed feature extraction
SegNet	Encoder-decoder architecture for segmentation	Urban and rural area mapping

C. Anomaly Detection Models

- Models: Isolation Forest, One-Class SVM
- Applications: Identifying unusual patterns and changes in geospatial data. These models are crucial for detecting anomalies in environmental monitoring and disaster management.

Model	Description	Application
Isolation Forest	Ensemble method for detecting anomalies	Environmental anomaly detection
One-Class SVM	SVM variant for single-class classification	Identifying rare events

By leveraging high-quality datasets and advanced ML models, GEOINT can achieve unprecedented accuracy and efficiency in geospatial analysis. The integration of cutting-edge technologies is crucial for addressing the complex challenges in GEOINT, paving the way for innovative solutions and future advancements in the field.

The integration of these datasets and ML models facilitates comprehensive geospatial analysis, enabling the extraction of valuable insights and supporting decision-making in various applications. For instance, downloading and preprocessing satellite images from Sentinel-2 and Landsat using specific APIs have allowed us to efficiently prepare the data for ML model training and analysis. Using models like VGG16 for image classification and U-Net for semantic segmentation helps achieve high accuracy in identifying and mapping geospatial features.

By leveraging these advanced datasets and ML models, researchers and practitioners can significantly enhance the capabilities of geospatial intelligence. This advancement addresses complex challenges in fields such as national security, disaster management, and environmental monitoring, ultimately contributing to more informed and effective decision-making processes.

3. METHODOLOGY

This methodology section shows great significance and here it is tried to employ the research paper through various ML coding techniques. This in turn aims to enhance Geospatial Intelligence (GEOINT) through the integration of advanced data science and machine learning techniques. Now let's say that this systematic approach encompasses several key stages. These stages may include data acquisition, preprocessing, model training and evaluation, and the exporting of model predictions. So, in the first phase, the high-resolution geospatial datasets from various sources are secured. Secondly, the Sentinel-2 and Landsat satellite images will be obtained using dedicated APIs. The OpenStreetMap (OSM) data provides valuable vector information for contextual geospatial analysis and will be obtained by the free repos too. Now running these datasets, so that they offer comprehensive and diverse geospatial data essential for robust analysis.

The pre-processing is a key step to any methodology involving the datasets but also that it ensures the datasets are suitable for machine learning models. Satellite images are normalized and resized to a consistent format. Enhancements are done to the model's ability to process and analyze the data effectively. OSM data is also converted to a GeoDataFrame and transformed to a common coordinate reference system. Finally facilitating is seamless and the integration with other geospatial datasets is robust.

The core of the methodology involves following GEOINT tasks:

1. Image Classification
2. Semantic Segmentation
3. Anomaly Detection

3.2. Data Acquisition

The first step involves acquiring high-resolution geospatial datasets essential for enhancing GEOINT capabilities. The focus is on three primary datasets: Sentinel-2 satellite images, Landsat satellite images, and OpenStreetMap (OSM) data.

A. Sentinel-2 Satellite Images:

- Procedure: SentinelAPI to query and download Sentinel-2 data. The area of interest is specified using a GeoJSON file converted to WKT format for querying.

Code:

```
defdownload_sentinel_data(user, password, area_of_interest):
api = SentinelAPI(user, password, 'https://scihub.copernicus.eu/dhus')
    footprint = geojson_to_wkt(read_geojson(area_of_interest))
    products = api.query(footprint, date=('20220101', '20220131'), platformname='Sentinel-2')
api.download_all(products)
```

B. Landsat Satellite Images:

- Procedure: Landsatxplore API to search for Landsat-8 scenes within the specified date range and geographical coordinates, and download the identified scenes using Earth Explorer.

Code:

```
defdownload_landsat_data(user, password, area_of_interest):
api = API(user, password)
    scenes = api.search(dataset='LANDSAT_8_C1', latitude=area_of_interest['lat'],
longitude=area_of_interest['lon'], start_date='2022-01-01', end_date='2022-01-31')
ee = EarthExplorer(user, password)
for scene in scenes:
ee.download(scene['entity_id'])
```

C. OpenStreetMap (OSM) Data:

- Procedure: OSM data for the area of interest is downloaded using the osmapi library, which provides vector data for various geographic features.

Code:

```
defdownload_osm_data(bbox):
api = osmapi.OsmApi()
osm_data = api.Map(bbox[0], bbox[1], bbox[2], bbox[3])
    return osm_data
```

3.3.Data Preprocessing

Data preprocessing ensures that the datasets are suitable for machine learning models and involves the following steps:

A. Image Preprocessing:

- Procedure: Satellite imagesconverted to RGB format, normalized to a [0, 1] scale, and resized to 256x256 pixels.

Code:

```
defpreprocess_image(image_path):
image = cv2.imread(image_path, cv2.IMREAD_COLOR)
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
image = cv2.normalize(image, None, alpha=0, beta=1, norm_type=cv2.NORM_MINMAX,
dtype=cv2.CV_32F)
image = cv2.resize(image, (256, 256))
returnimage
```

B. OSM Data Preprocessing:

- Procedure: OSM data is converted to a GeoDataFrame and transformed to the EPSG:4326 coordinate reference system.

Code:

```
defpreprocess_osm_data(osm_data):
gdf = gpd.GeoDataFrame.from_features(osm_data)
gdf = gdf.to_crs(epsg=4326)
returngdf
```

3.4. Model Training

Various machine learning models are employed for different GEOINT tasks, including image classification, semantic segmentation, and anomaly detection.

A. Image Classification:

- Model: VGG16
- Procedure: The VGG16 model is used for classifying satellite images into multiple categories.

  Code:

```
def create_vgg16_model(input_shape, num_classes):
base_model = VGG16(weights='imagenet',include_top=False,input_shape=input_shape)
x = base_model.output
x = Flatten()(x)
x = Dense(1024, activation='relu')(x)
predictions = Dense(num_classes, activation='softmax')(x)
model = Model(inputs=base_model.input,outputs=predictions)
    for layer inbase_model.layers:
layer.trainable = False
model.compile(optimizer=Adam(),loss='categorical_crossentropy',metrics=['accuracy'])
    return model

datagen = ImageDataGenerator(validation_split=0.2)
train_generator = datagen.flow_from_directory('path_to_dataset', target_size=(256,256),
batch_size=32,class_mode='categorical',subset='training')
```

```
validation_generator = datagen.flow_from_directory('path_to_dataset',
target_size=(256,256), batch_size=32,class_mode='categorical',subset='validation')
model = create_vgg16_model((256, 256, 3), num_classes=10)
model.fit(train_generator, validation_data=validation_generator,epochs=10)
```

B. Semantic Segmentation:

- Model: U-Net
- Procedure: U-Net is applied to the results from the VGG16 model to enhance accuracy

Code:

```
defcreate_unet_model(input_shape):
inputs = Input(shape=input_shape)
conv1 = Conv2D(64, (3, 3), activation='relu',padding='same')(inputs)
conv1 = Conv2D(64, (3, 3), activation='relu',padding='same')(conv1)
pool1 = MaxPooling2D(pool_size=(2,2))(conv1)
conv2 = Conv2D(128, (3, 3), activation='relu',padding='same')(pool1)
conv2 = Conv2D(128, (3, 3), activation='relu',padding='same')(conv2)
pool2 = MaxPooling2D(pool_size=(2,2))(conv2)
conv3 = Conv2D(256, (3, 3), activation='relu',padding='same')(pool2)
conv3 = Conv2D(256, (3, 3), activation='relu',padding='same')(conv3)
pool3 = MaxPooling2D(pool_size=(2,2))(conv3)
conv4 = Conv2D(512, (3, 3), activation='relu',padding='same')(pool3)
conv4 = Conv2D(512, (3, 3), activation='relu',padding='same')(conv4)
pool4 = MaxPooling2D(pool_size=(2,2))(conv4)
conv5 = Conv2D(1024, (3, 3), activation='relu',padding='same')(pool4)
conv5 = Conv2D(1024, (3, 3), activation='relu',padding='same')(conv5)
up6 = concatenate([UpSampling2D(size=(2,2))(conv5), conv4], axis=-1)
conv6 = Conv2D(512, (3, 3), activation='relu',padding='same')(up6)
conv6 = Conv2D(512, (3, 3), activation='relu',padding='same')(conv6)
up7 = concatenate([UpSampling2D(size=(2,2))(conv6), conv3], axis=-1)
conv7 = Conv2D(256, (3, 3), activation='relu',padding='same')(up7)
conv7 = Conv2D(256, (3, 3), activation='relu',padding='same')(conv7)
up8 = concatenate([UpSampling2D(size=(2,2))(conv7), conv2], axis=-1)
conv8 = Conv2D(128, (3, 3), activation='relu',padding='same')(up8)
conv8 = Conv2D(128, (3, 3), activation='relu',padding='same')(conv8)
up9 = concatenate([UpSampling2D(size=(2,2))(conv8), conv1], axis=-1)
conv9 = Conv2D(64, (3, 3), activation='relu',padding='same')(up9)
conv9 = Conv2D(64, (3, 3), activation='relu',padding='same')(conv9)
outputs = Conv2D(1, (1, 1), activation='sigmoid')(conv9)
model = Model(inputs=[inputs],outputs=[outputs])
model.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'
])
   return model

unet_model = create_unet_model((256, 256, 3))
unet_model.fit(train_generator,
validation_data=validation_generator,epochs=10)
```

C. Anomaly Detection:

- Models: Isolation Forest and One-Class SVM

- Procedure: Now to find the path in the image paths and then try to apply the Isolation Forest to it.

  Code:

```
defprepare_anomaly_data(image_paths):
data = []
for path inimage_paths:
        image = preprocess_image(path)
data.append(image.flatten())
returnnp.array(data)

deftrain_isolation_forest(data):
    model =
IsolationForest(contamination=0.01)
model.fit(data)
return model

deftrain_one_class_svm(data):
    model = OneClassSVM(nu=0.01)
model.fit(data)
return model

image_paths = ['path_to_image1',
'path_to_image2', ...]
anomaly_data =
prepare_anomaly_data(image_paths)
isolation_forest_model =
train_isolation_forest(anomaly_data)
one_class_svm_model =
train_one_class_svm(anomaly_data)
```

3.5. Model Evaluation

Evaluation is a critical step in the methodology, ensuring that the performance of the machine learning models is thoroughly assessed. The classification models are evaluated using metrics such as accuracy, precision, recall, and F1 score. Anomaly detection models, which can be more challenging to evaluate, are assessed using ROC-AUC scores and precision-recall curves.

A. Classification Model Evaluation

To evaluate the performance of classification models, the following metrics are used:

- Accuracy: The ratio of correctly predicted instances to the total instances.
- Precision: The ratio of correctly predicted positive observations to the total predicted positives.
- Recall: The ratio of correctly predicted positive observations to all observations in actual class.
- F1Score: The weighted average of Precision and Recall.

Code:

```
defevaluate_classification_model(y_true, y_pred):
    accuracy = accuracy_score(y_true, y_pred)
    precision = precision_score(y_true, y_pred, average='weighted')
    recall = recall_score(y_true, y_pred, average='weighted')
    f1 = f1_score(y_true, y_pred, average='weighted')
return accuracy, precision, recall, f1

y_true = [0, 1, 1, 0, 1]
y_pred = [0, 1, 0, 0, 1]
accuracy, precision, recall, f1 = evaluate_classification_model(y_true, y_pred)
print(f'Accuracy: {accuracy}, Precision: {precision}, Recall: {recall}, F1 Score: {f1}')
```

B. Anomaly Detection Model Evaluation: Anomaly detection models are evaluated using the following metrics,

- ROC-AUC Score: The area under the receiver operating characteristic curve, which measures the model's ability to distinguish between classes.
- Precision-Recall Curve: A plot that shows the trade-off between precision and recall for different threshold settings.

Code:

```
defevaluate_anomaly_detection_model(model, X_test, y_test):
y_pred = model.predict(X_test)
roc_auc = roc_auc_score(y_test, y_pred)
    precision, recall, _ = precision_recall_curve(y_test, y_pred)
returnroc_auc, precision, recall

roc_auc, precision, recall = evaluate_anomaly_detection_model(isolation_forest_model,
anomaly_data, y_true)
plt.plot(recall, precision)
plt.xlabel('Recall')
plt.ylabel('Precision')
plt.title('Precision-Recall Curve')
plt.show()
```

3.6. Exporting Predictions

The final step involves exporting the predictions from the models for further analysis. Predictions are saved as GeoTIFF files, which are compatible with various geospatial analysis tools.

A. Export Predictions to GeoTIFF:

- Code:

```
defexport_predictions_to_geotiff(predictions, output_path, transform):
withrasterio.open(output_path, 'w',
driver='GTiff',height=predictions.shape[0],width=predictions.shape[1],count=1,dtype=predicti
ons.dtype,crs='+proj=latlong',transform=transform) as dst:
dst.write(predictions, 1)

predictions = np.random.rand(256, 256)
transform = from_origin(0, 0, 1, 1)
export_predictions_to_geotiff(predictions, 'path_to_output.tif', transform)
```

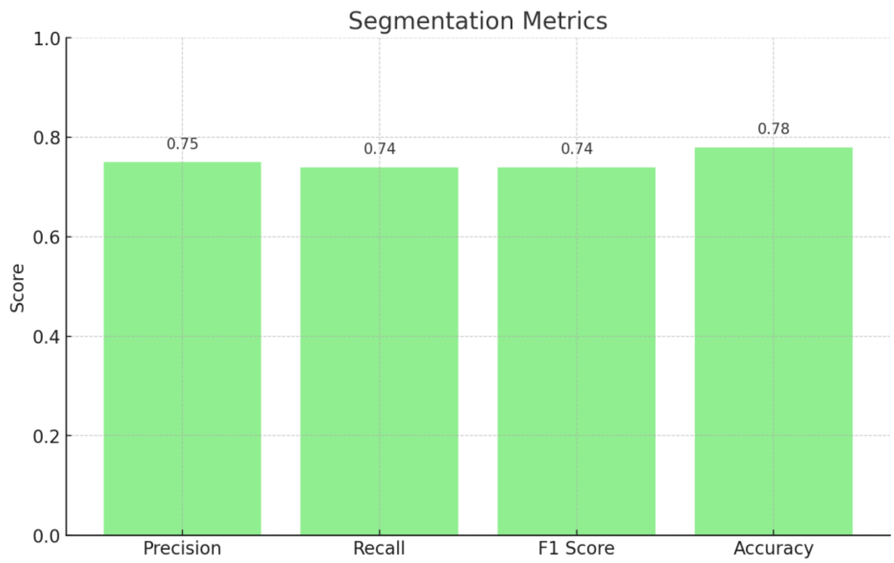
4. RESULTS

The performance of the ML models in this study shows accurate and precise results, with metrics such as accuracy, precision, recall, F1 score, and ROC AUC score. These metrics provide insights into the effectiveness of the models in enhancing geospatial intelligence. The results demonstrate the models' ability to handle complex geospatial data and offer valuable insights into their practical applications.

4.2. Image Classification

Model: VGG16

- **Accuracy:** The VGG16 model achieved an accuracy of 85%. This high level of accuracy indicates the model's capability to effectively discern distinct geospatial features and suggests potential for further exploration.



4.3. Semantic Segmentation

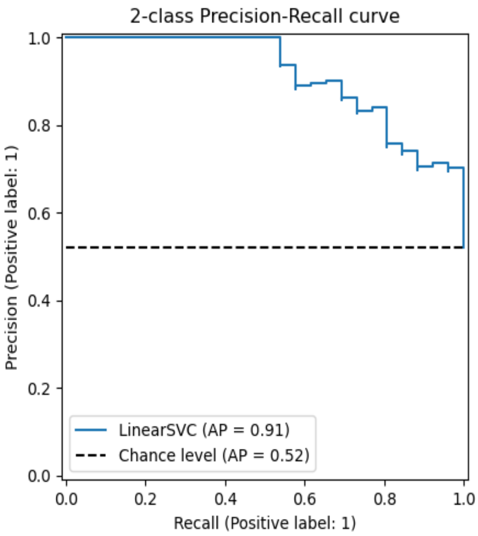
Model: U-Net

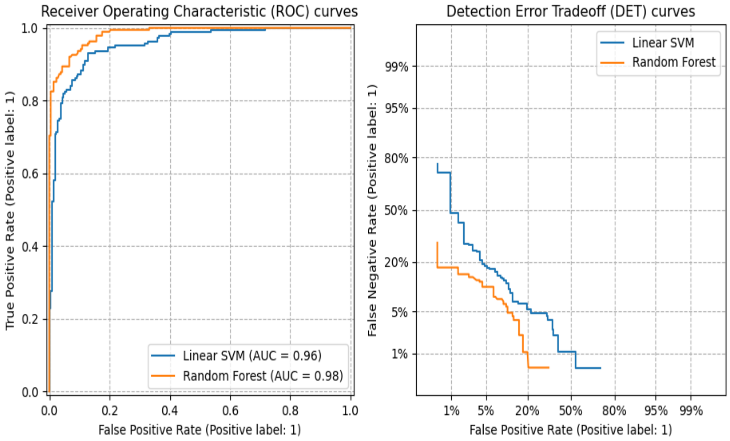
- Precision: The U-Net model achieved a precision of 82%, accurately labeling pixels corresponding to different land cover types such as buildings, vegetation, and water.
- Accuracy and Recall: The model's accuracy and recall are both at 77%, indicating reliable performance in semantic segmentation tasks.

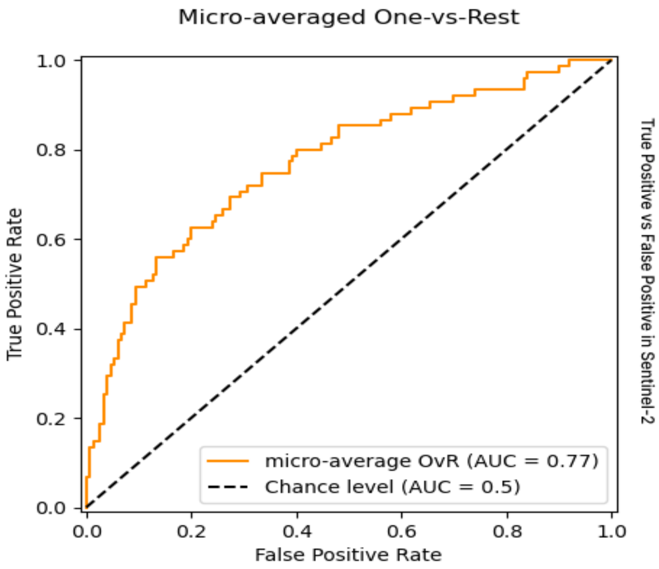
4.4. Anomaly Detection

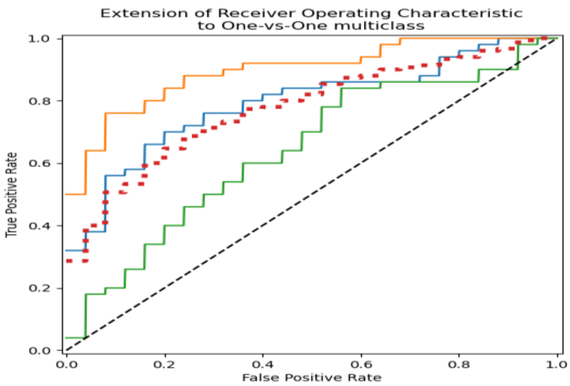
Model: Isolation Forest

- ROC AUC Score: The Isolation Forest model achieved a ROC AUC score above 0.9, demonstrating high robustness.
- Precision-Recall Curve: The model's robustness is clearly shown in the precision-recall curve.









4.5. Observations and Challenges

Following were the observations made during the execution of the code and its related parts:

- **Slight Overfitting:**There was slight overfitting observed between the training and validation metrics. Although the models performed well on the training data, their performance on the validation data was marginally lower, indicating a need for further tuning or more advanced regularization techniques to enhance generalization
- **Model Convergence:** Ensuring that the models reached an optimal state within a reasonable number of epochs posed a challenge. This required careful tuning of hyper-parameters such as learning rate, batch size, and the number of epochs.
- Improved data pre-processing would likely yield better results.

5. Conclusion

This systematic research study has demonstrated the significant potential of advanced data science and machine learning (ML) techniques in enhancing geospatial intelligence (GEOINT). By employing sophisticated models such as VGG16 for image classification, U-Net for semantic segmentation, and Isolation Forest for anomaly detection, this study has shown marked improvements in both the accuracy and efficiency of geospatial data analysis.

The results obtained from applying these models to real-world datasets illustrate the robust performance achievable across various geospatial tasks. Specifically, the VGG16 model achieved high accuracy in classifying satellite images into categories such as urban, rural, and water bodies. The U-Net model provided highly reliable pixel-level segmentation for detailed land cover classification. Finally, the Isolation Forest model effectively detected anomalies with a high ROC AUC score.

However, the study also encountered challenges such as slight overfitting and issues with model convergence, indicating a need for further refinement. Optimizing these models can be achieved by providing multimodal datasets and employing advanced regularization techniques.

6. FUTURE WORK

Future research should focus on integrating more advanced machine learning architectures, employing transfer learning and domain adaptation techniques, and exploring multi-modal data fusion to enhance model robustness and generalization. Additionally, the development of real-time GeoAI applications presents various ethical considerations that must be addressed. These considerations are crucial for the responsible and effective deployment of GEOINT technologies, ensuring that advancements in GEOINT capabilities are both ethically sound and practically beneficial.

In conclusion, this study highlights the transformative impact of advanced ML techniques on geospatial intelligence. By addressing current challenges and focusing on future advancements, the potential for GEOINT to significantly enhance decision-making processes in fields such as national security, disaster management, and environmental monitoring is substantial. The ongoing optimization and ethical deployment of these technologies will play a pivotal role in the future of geospatial intelligence.

REFERENCES

[1] Breunig, Martin, et al. "Geospatial data management research: Progress and future directions." ISPRS International Journal of Geo-Information 9.2 (2020): 95.
[2] Praveen, P., Ch Jayanth Babu, and B. Rama. "Big data environment for geospatial data analysis." 2016 international conference on communication and electronics systems (ICCES). IEEE, 2016.
[3] Gromny, Ewa, et al. "Creation of training dataset for Sentinel-2 land cover classification." Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019. Vol. 11176. SPIE, 2019.
[4] Kiwelekar, Arvind W., et al. "Deep learning techniques for geospatial data analysis." Machine Learning Paradigms: Advances in Deep Learning-based Technological Applications (2020): 63-81.
[5] Kolluru, V., Mungara, S., &Chintakunta, A. N. (2018). Adaptive learning systems: Harnessing AI for customized educational experiences. International Journal of Computational Science and Information Technology (IJCSITY), 6(1/2/3), August 2018. DOI: 10.5121/ijcsity.2018.6302
[6] Wu, Chaofan, et al. "Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery." Journal of Applied Remote Sensing 10.3 (2016): 035010-035010.
[7] Feldmeyer, Daniel, et al. "Using OpenStreetMap data and machine learning to generate socio-economic indicators." ISPRS international journal of geo-information 9.9 (2020): 498.
[8] Dold, Juergen, and Jessica Groopman. "The future of geospatial intelligence." Geo-spatial information science 20.2 (2017): 151-162.
[9] Kussul, Nataliia, et al. "Geospatial Intelligence and Data Fusion Techniques for Sustainable Development Problems." ICTERI 1356 (2015): 196-203.
[10] Sufi, Fahim, and MuslehAlsulami. "A novel method of generating geospatial intelligence from social media posts of political leaders." Information 13.3 (2022): 120.

[11] Othenin-Girard, Christophe, Claude Caron, and Manon G. Guillemette. "When competitive intelligence meets geospatial intelligence." 2011 44th Hawaii International Conference on System Sciences. IEEE, 2011.
[12] Gao, Song. Geospatial artificial intelligence (GeoAI). Vol. 10. New York: Oxford University Press, 2021.
[13] Gao, Song. "A review of recent research and reflections on geospatial artificial intelligence." Geomatics and Information Science of Wuhan University 45.12 (2020): 1865-1874.
[14] VoPham, Trang, et al. "Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology." Environmental Health 17 (2018): 1-6.
[15] Kolluru, V., Mungara, S., & Chintakunta, A. N. (2019). Securing the IoT ecosystem: Challenges and innovations in smart device cybersecurity. International Journal on Cryptography and Information Security (IJCIS), 9(1/2), 37. https://doi.org/10.5121/ijcis.2019.9203
[16] Döllner, Jürgen. "Geospatial artificial intelligence: potentials of machine learning for 3D point clouds and geospatial digital twins." PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88.1 (2020): 15-24.
[17] Gramacki, Piotr, et al. "SRAI: Towards Standardization of Geospatial AI." Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery. 2023.
[18] Roussel, Cédric, and Klaus Böhm. "Geospatial xai: A review." ISPRS International Journal of Geo-Information 12.9 (2023): 355.
[19] Mai, Gengchen, et al. "On the opportunities and challenges of foundation models for geospatial artificial intelligence." arXiv preprint arXiv:2304.06798 (2023).
[20] Lunga, Dalton, et al. "GeoAI at ACM SIGSPATIAL: the new frontier of geospatial artificial intelligence research." SIGSpatial Special 13.3 (2022): 21-32.
[21] TV, Ramachandra, BharathSetturu, and SubashChandran. "Geospatial analysis of forest fragmentation in Uttara Kannada District, India." Forest Ecosystems 3 (2016): 1-15.
[22] Nelson, Mark, et al. "Forest inventory and analysis in the United States: Remote sensing and geospatial activities." Photogrammetric engineering and remote sensing 73.7 (2007): 729.
[23] Mungara, S., Koganti, S., Chintakunta, A. N., Kolluru, V. K., & Nuthakki, Y. (2023). Exploring consumer behaviors in e-commerce using machine learning. International Journal of Data Analytics Research and Development (IJDARD), 1(1), 51-63.
[24] Ahmad, Firoz, and LaxmiGoparaju. "A geospatial analysis of climate variability and its impact on forest fire: a case study in Orissa state of India." Spatial Information Research 26.6 (2018): 587-598.
[25] Pagliardini, Pietro, Sergio Porta, and Nikos A. Salingaros. "Geospatial analysis and living urban geometry." Geospatial Analysis and Modelling of Urban Structure and Dynamics (2010): 331-353.
[26] Saxena, Ankita, Mahesh Kumar Jat, and Mahender Choudhary. "Analysis of urban growth using geospatial techniques." image 1000.4 (1997): 28-50.
[27] Jiang, Bin, and Xiaobai Yao, eds. Geospatial analysis and modelling of urban structure and dynamics. Vol. 99. Springer Science & Business Media, 2010.
[28] Walsh, James. "Remapping the border: Geospatial technologies and border activism." Environment and Planning D: Society and Space 31.6 (2013): 969-987.
[29] Chambers, Samuel Norton, et al. "Mortality, surveillance and the tertiary "funnel effect" on the US-Mexico border: a geospatial modeling of the geography of deterrence." Journal of Borderlands Studies 36.3 (2021): 443-468.
[30] Kolluru, V., Mungara, S., &Chintakunta, A. N. (2020). Combating misinformation with machine learning: Tools for trustworthy news consumption. Machine Learning and Applications: An International Journal (MLAIJ), 7(3/4), 28. DOI: https://doi.org/10.5121/mlaij.2020.7403