Forest/rural road network detection and condition monitoring based on satellite imagery and deep semantic segmentation

Dimitrios Kelesakis[1], Konstantinos Marthoglou[1], Eleni Tokmaktsi[2], Emmanouel Tsiros[2], Apostolos Karteris[2], Anastasia Stergiadou[3], George Kolkos[3], Petros Daras[1], and Nikos Grammalidis[1]

[1]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece – {dimkelesakis, koma, daras, ngramm}@iti.gr
[2]kartECO - Environmental and Energy Engineering Consultancy, Thessaloniki, Greece – {e.tokmaktsi, e.tsiros, a.karteris}@karteco.gr
[3]Institute of Forest Engineering and Topography, Aristotle University of Thessaloniki, Thessaloniki, Greece –{nanty, gnkolkos}@for.auth.gr

Keywords: Remote Sensing, artificial intelligence, forest and rural road network monitoring, gravel road detection and extraction, gravel road condition monitoring

Abstract

Sustainable forest and emergency management require comprehensive data on the forest road network and its condition. This paper presents the final framework of the INFOROAD project (https://inforoad.karteco.gr/), which integrates cutting-edge remote sensing and machine learning technologies for automated periodic extraction and monitoring of the forest road network. The framework includes gravel road extraction, road graph extraction, and gravel road condition monitoring, with a focus on the periurban forest in Thessaloniki. The road extraction employs the U-TAE network architecture, with a proposed modification using inverted residual blocks for improved accuracy. Road graph extraction involves creating a graph from road segmentation output or OSM data, enabling efficient road segment analysis. Gravel-road width calculation utilizes road segmentation results and a series of image processing steps, while road condition monitoring employs ML/AI classification algorithms. Worldview 3 high-resolution satellite images and various auxiliary data sources (e.g. DEM) are used as input, including field measurements for the training of classification algorithms. Results showcase the effectiveness of the proposed framework, with gravel road extraction accuracy improved by the modified U-TAE model. Regarding gravel road condition monitoring, algorithms achieving satisfactory results are identified, despite the challenges that arise, due to the significant surface and texture variations in forest and agricultural roads. A WebGIS platform facilitates information presentation, user interaction, and management of geospatial information, supporting various functionalities such as layer management and spatial data visualization. The INFOROAD project represents a significant advancement in leveraging technology for sustainable forest road management and emergency preparedness. Future steps may involve further enhancements and adaptations for improvement of results.

1. Introduction

Comprehensive data on the forest/rural road network and its condition is crucial for sustainable forest and emergency management. Timely spatial information covering the entire network, including issues like potholes, bumps, and erosion, is vital to the public and regulatory bodies for monitoring and maintenance tasks. This information also facilitates an access to the road network and contributes to the protection of valuable resources in case of emergencies. Consequently, there is a demand for the development of reliable tools and methodologies supporting quantitative measurements of specific indicators related to the forest and rural road network's state at a given time.

Various approaches have been proposed to collect information regarding forest/rural road-network, including remote sensing observations (via satellite systems and aerial drones) and field measurements. The automated analysis of satellite images by powerful Machine Learning/Artificial Intelligence (ML/AI) algorithms, can offer significant advantages, including increased spatial coverage as well as avoidance of labor-intensive field measurements. In this way, the time and cost associated with recording, mapping, and monitoring the forest and rural road network can be significantly reduced.

The primary contribution of the INFOROAD project (Kelesakis et al., 2023) lies in integrating cutting-edge remote sensing and machine learning technologies for the automated periodic extraction and monitoring of the forest road network. In this paper, the final framework designed and developed within the project is presented and evaluated for the Seich-Sou forest in Thessaloniki, Greece. More specifically, this framework includes: a) gravel road extraction approach based on deep semantic segmentation of high-resolution multispectral satellite sequences, b) road graph extraction combining road masks obtained by (a) as well as the available OSM data, and c) gravel road condition monitoring by using AI/ML models for classification based on the available high-resolution satellite data, which are trained by associated field measurements.

Furthermore, the presentation, dissemination, and management of this information by the different managing authorities and/or the general public are facilitated through an online Geographic Information System (WebGIS).

2. Related Work

In recent times, propelled by advancements in deep learning and artificial intelligence, substantial research has been dedicated to road detection from satellite imagery, particularly focusing on urban asphalt roads. A comprehensive review of automated road extraction from various aerial sources, including satellite images (optical or SAR) and 3-D Lidar point clouds, has been conducted (Ferraz et al., 2016). Most studies frame the problem as a two-class semantic segmentation task, aiming to distinguish

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W4-2024-81-2024

roads from the background. Leading semantic segmentation algorithms, such as DeepLab (Xia et al., 2018), U-NET (Zao & Shi, 2021), and others, leverage deep Convolutional Neural Networks for this purpose. In specific cases, researchers have combined Convolutional Neural Networks (CNN) with Graph Neural Networks (GNN) to enhance performance. Yan et al. (2022) integrated a CNN for extracting semantic features and a GNN for reasoning on these features to produce the final road map.

In the case of forest road and rural road extraction, research faces significant challenges due to their heterogeneous appearance, due to diverse road materials, different illumination conditions, shadows, and occluding vegetation. The complexity of these environments leads to limited research outcomes on targeted export models and a lack of corresponding datasets. Current research in this domain is predominantly represented by the work of Kearney et al. (2020), who classified pixels of satellite images into forest/rural roads and non-roads using a CNN and training data obtained from vehicle monitoring in the area. Additionally, Jiang et al. (Jiang et al., 2022) developed the Roadformer, employing a pyramidal architecture of a deformable vision transformer for road network extraction from satellite images. The deformable vision transformer consistently attends to the most important semantic features, significantly improving performance. Zhang et al. (Zhang et al., 2022) contributed to the field by creating the "Road Datasets in Complex Mountain Environments (RDCME)" dataset, proposing the Light Roadformer model for forest road extraction, which utilizes a transformer based on self-attention units. Post-processing techniques are applied to reduce errors, resulting in a mIoU of 89.5% for RDCME.

Addressing the challenge of automatic extraction of road network graphs from satellite images, existing algorithms typically involve pixel-level image segmentation followed by vectorization. Bahl et al. (2022) introduced a method that directly outputs the final road graph in a single pass, significantly outperforming other algorithms in terms of speed. This method combines a Fully Convolutional Neural Network (FCN) for identifying key points, such as junctions, with a Graph Neural Network (GNN) predicting links between these points. Sat2Graph (He et al., 2020), based on Graph Tensor Encoding (GTE), represents another method that encodes the road graph into a tensor representation, allowing the generation of road graphs in a single step. This non-iterative approach captures global information more effectively and simplifies the training process.

The periodic monitoring of gravel-road conditions, which is essential for road maintenance, is an expensive and labour-intensive process, highlighting the need for the developing new, automated, and cost-effective methodologies. Evaluating the condition of each segment of forest or rural roads on a scale of 1 to 5 is valuable for providing users with accessibility information (Kolkos et all.2023). The International Roughness Index (IRI) is a globally utilized indicator for characterizing the longitudinal roughness of roads (Sayers, 1995). This specific index can be computed using a mobile phone appropriately fitted to a car, which traces the road's trajectory for assessment (Cadamuro et al., 2019). Various convolutional neural network (CNN) architectures, including AlexNet, SqueezeNet, and Visual Geometry Group (VGG), are compared to appraise road section conditions on a 1-5 scale based on their quality, utilizing satellite images. The initial labeling of the dataset, encompassing roads totaling 1,150 km in Kenya, involves leveraging georeferenced measurements of the IRI index from passing vehicles. A similar methodology is employed in another study (Kalpoma et al., 2022) using Vgg16 and ResNet50 for training, while a different study (Thegeya et al., 2022) utilizes IRI measurement data from the Philippines, Sentinel-2 satellite imagery, and a custom architecture, including ResNet-34 CNN, to categorize road quality into four classes. Additionally, a transfer learning approach is suggested in a different study (Brewer et al., 2021), where a CNN neural network is initially trained on a large road quality dataset from the United States and subsequently fine-tuned on an independent, smaller dataset collected from Nigeria.

3. Methodology

The project aims at exploiting satellite data both for forest/rural road network detection as well as road condition monitoring. Since, forest road width can be as small as 5m, it was decided to use Worldview 3 high-resolution satellite image with a resolution of 0.3m for the panchromatic image and a resolution of 1.2m for each of the eight multispectral channels. Auxiliary data from other sources were also used after any required pre-processing procedures, such as a) OpenStreetMap (OSM) data (OpenStreetMap contributors, 2017), b) a Digital Elevation Models (DEM), c) field measurements, and d) data from available forest maps, which can be used to extract additional information such as road width, roadside slope or aspect.

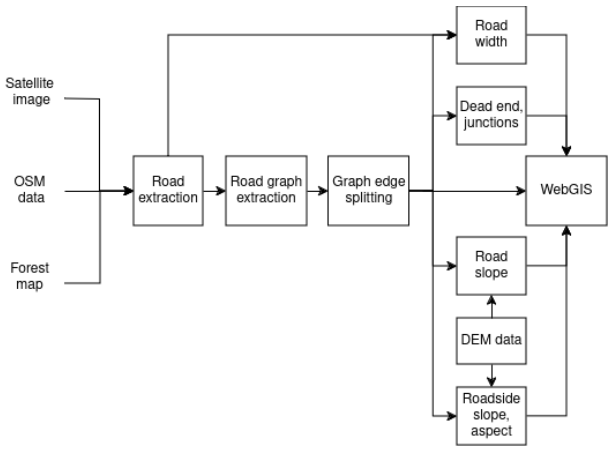


Figure 1: The proposed methodology for road network detection

The basic methodology for road network detection is shown in Figure 1. Initially, information from available forest maps is digitised to create binary road masks. If such forest maps are not available, OSM vector data can be used instead, taking into account specific assumptions regarding each road width. In both cases, these binary roads masks are used as ground truth to train a road extraction algorithm using the available satellite image. An efficient supervised deep semantic segmentation algorithm is used for this task, which may also identify additional (e.g. new) road segments that may not be present at the original forest/OSM maps. This road extraction step can be omitted, if the available annotation (e.g. from digitised forest maps) is accurate enough.

Following road extraction, a road graph extraction procedure based on skeletonization converts the binary segmentation output into a graph with nodes and edges (road segments). Then, long graph edges are further split into smaller edges to ensure that the maximum edge length is no more than 100

meters. In this way, big road segments are avoided, making road segment status monitoring more efficient. This final graph version is then used for extracting specific features, such as average road width, junctions, dead ends and junctions. Similarly, the road slope can be calculated based on digital elevation data from a Digital Elevation Model (DEM). A high resolution (5m) DEM provided to Aristotle University by the Hellenic Cadastre, namely "Digital Elevation Model - DEM - LSO(5m)", was used. Finally, the roadside aspect and slope can also be calculated based on the available digital elevation data, by defining appropriate (e.g. 10m width) “roadside” buffer areas at the two sides of each edge.

Regarding road condition monitoring, Figure 2 presents the proposed methodology based on the texture classification of each road segment texture using ML/AI algorithms. Specifically, the satellite image data within each buffered graph edge (defining the area of each road segment) are used as input to an ML/AI classification algorithm that is trained using specific field measurements (e.g. 1-5 rating scale surveys). Different classification algorithms are tested for this task, including optimised versions of Pixel-Set-Encoder (PSE, Garnot 2020), ResNet, CNN, kNN. After training, the algorithm is used to provide a classification output in a scale of 1 to 5 for any road segment. The outputs of the algorithms are finally saved in shapefile or geojson format and are presented through the WebGIS that was developed.

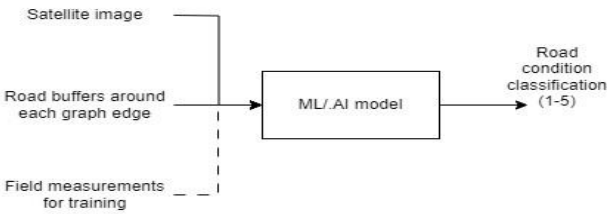


Figure 2: Road quality monitoring methodology

3.1 Gravel Road Extraction

For the road extraction task, the U-TAE (U-Net with Temporal Attention Encoder) network architecture (Garnot, 2021) is employed. This architecture has proven to be highly effective for semantic segmentation utilizing multi-temporal satellite data. The applied training procedure involves using the U-TAE model with a custom training dataset derived from the Seich-Sou forest in Greece, in the greater region of Thessaloniki. The image contains eight bands for a single temporal instance. The dataset was split into five folds, as in (Garnot, 2021), so as to be able to apply cross-validation. The ground truth masks are obtained by exploiting the OSM and forest map information.

In order to further improve the accuracy of the model, some modifications were introduced to the U-TAE model. Specifically, the architecture was modified by replacing the first and last convolutional layers with a series of expressive and powerful convolutional blocks, namely inverted residual blocks (Sandler et al., 2018), which are able to offer a less time consuming alternative to residual blocks (He et al., 2016).

More specifically, residual blocks are comprised of successive convolutional layers with a wide -> narrow -> wide structure, with respect to the number of their channels. This means that the input has a high number of channels, which are then compressed using a 1x1 convolution. In the end, the number of channels is increased again, in order to match the input and output. Inverted residual blocks have the opposite structure, i.e. narrow -> wide -> narrow. This means that the number of the channels is increased; therefore, the feature maps become wider using a 1x1 convolution. A 3x3 depth-wise convolution is then utilized, which decreases greatly the number of parameters. Finally, a 1x1 convolution decreases the number of channels in order to match input and output.

Figure 3 illustrates the structure of the respective blocks. Inverted residual blocks are used in order to replace in the structure of the architecture of the first and last convolutional layers in U-TAE. Inverted residual blocks can offer greater expressivity in comparison to the standard convolutional layers of the original U-TAE, while also avoiding the increase in parameters that would be incurred if residual blocks are used instead.

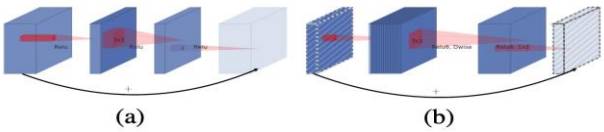


Figure 3: (a) Residual block, (b) Inverted residual block

3.2 Road Graph Extraction

This task is responsible for the creation of a graph representation of the road network. This graph can be obtained in two ways. The first method is the automatic creation of a graph using the output raster of the road segmentation process. It has the advantage of up-to-date road information extracted directly from the satellite image but it may face issues, such as low road connectivity, roads covered by vegetation, etc. that may require further post-processing. The second method entails creating the graph directly from OSM data, by converting OSM vector data into a standard graph format. This method has the advantage of better road connectivity, but data may not always be up-to-date, as OSM depends on voluntary contributions.

The first method includes two steps: first, an image skeletonization procedure is used to the extracted binary road mask and then a road graph, consisting of nodes and edges, is created based on the sknw python library[1], yielding graphs at the standard NetworkX format.

The second method involves converting OSM vector data into graph format, e.g. by using the osmnx python library[2]. As opposed to the first method, where graph edges are line segments, in this case graph edges may be polylines. As the length of these polylines may be arbitrary large and smaller segments are required for road quality monitoring, a graph editing procedure is introduced to limit the maximum edge length size below a threshold (e.g. 100m).

This procedure includes three steps for each edge: a) the total length of the edge (polyline), i.e. sum of its line segments is computed, b) if this length is higher than the threshold, the total number of segments is calculated by dividing each length with the threshold and selecting the next greater integer, c) the desirable edge length is calculated by dividing the edge length with this number of segments and d) the original edge is split according to this desirable edge length and additional nodes are

[1] https://github.com/Image-Py/sknw/
[2] https://github.com/gboeing/osmnx

added to the graph. An example result of this procedure is shown in Figure 4,

The road graph representation of the road network facilitates road segment analysis, such as identification of dead ends and junctions. Specifically, the degree D of each node, i.e. the number of edges (lines) connecting to the node is computed, indicating whether the node is a junction (if D > 2) or dead end[3] (if D = 1).

Figure 4: OSM graph before graph editing (left) and after graph editing (right) (© OpenStreetMap Contributors)

3.3 Gravel Road Width Calculation

For each graph edge, the average road width is also estimated, taking into account the road segmentation results. First, the output of the road extraction module is processed to be used for road width estimation. A chamfer distance transform is applied, where the Euclidian distance of each road pixel from the nearest non-road pixel is calculated and stored to create a *chamfer image*. Thus, each pixel in this chamfer image that belongs to a road contains a value equal to the distance from the nearest edge of the road. In the center of the road, i.e. the central axis of the road, the specific values are maximized. These distances are typically multiples of pixel size, but can also be converted to meters by multiplying with the pixel size. By multiplying the chamfer image with the skeletonized mask of the extracted road network (see Section 3.2), distances of road centerline pixels from the roadside are extracted.

Then, a rectangular buffer area is constructed around each edge of the graph, with a buffer width that is larger than the real road width, e.g. 20 meters. In this way, edges of the graph are transformed to polygon masks, labelled with the respective edge id. Then, for each graph edge, the corresponding rectangular buffer area is used to mask the road centreline. Finally, we can estimate the average distance ***d*** between the centreline and the roadside for the specific edge by calculating the mean of the corresponding non-zero values from the road centerline distance mask. The final width estimate (from one roadside to the corresponding opposite roadside) will then be *2*****d***.

In order to assess the accuracy of the width estimation process, any available forest map data can be utilized. Road width from forestry map is calculated by using the same distance transform and then a similar mask is employed for each edge. In this way, the calculated road widths from the forestry map are matched to their corresponding edge ids.

3.4 Gravel Road Condition Monitoring

Gravel road condition monitoring is crucial to ensure safe transportation via forest and rural roads. For this reason, an ML/AI model was constructed to estimate the road condition from satellite images. The input of the model consists of selected channels from a multispectral Worldview-3 satellite image (including a panchromatic channel), appropriate masks of the road areas and the corresponding road quality annotation that is used for training the model.

OSM road data are retrieved to create appropriate road masks. Street data are either in geojson or shapefile format, with streets represented as lines and including additional attributes like street name and road type. A similar buffering process is implemented with a buffer width size of 10 metres and the corresponding rectangle buffers are created. Then, vector data are rasterized with the same resolution and extent as the satellite image.

Annotation data, provided by specialists from the Forestry school, was used as input for the training and evaluation of the model. Specifically, annotation data was collected during two field visits at two main roads of the Seich Sou forest in Thessaloniki at dates close to the date that the satellite image was captured. Parts of the two main roads, namely A1T1 (10 km) and A2T2 (12 km), were first divided into 100 meters long segments. Then, each of the resulting 220 segments was evaluated concerning road surface damage, taking into account the frequency, intensity and location of the damage. The final annotation contains an evaluation of the quality of each road segment at a scale of 1 (best quality) to 5 (worst quality), as well as notes/description of specific damages at that segment that can serve as a justification of this evaluation.

3.4.1 Training models: Different ML and deep learning classification methods were compared during the training process, including CNNs, ResNet50 architecture, PSE-MLP (Pixel Set Encoder-MultiLayer Perceptron) architecture and k-Nearest Neighbors (KNN). As an alternative approach, a linear regression method was also evaluated.

Convolutional neural networks (CNNs) are a class of deep neural networks that can recognize specific features from images with high accuracy and are widely used for image analysis and classification (Lecun Yann et al., 1989).

ResNet50 (Kaiming He et al., 2016) is another convolutional neural network (CNN)-based model, which consists of 50 layers, including convolutional layers as well as a “Max Pooling” level and an “Average Pooling” level. This residual network framework allows the incorporation of shortcut connections and the use of residual functions, thus allowing the stacked layers of deep neural networks to reduce their training errors.

Garnot (Garnot et al., 2020) suggested that convolution functions may not be suitable for detecting different classes from images with high spectral variations over time. CNNs were also observed to consume large amounts of memory. To overcome the above issues the PSE-TAE architecture was proposed, consisting of a pixel-set encoder (used to sample a fixed-size vector of data from arbitrary sized regions), a temporal attention encoder and a classifier. In this work, we used PSE to obtain data from each graph edge and then use a set of linear layers (multilayer perceptron) to obtain the final classification result.

Moreover, the k-Nearest Neighbors (KNN) algorithm is a popular and one the simplest traditional Machine Learning

[3] After a dead end is detected, it is extended with additional edges. until the first junction is found.

algorithms. It uses a non-parametric, supervised learning classifier, that classifies each individual data point based on proximity.

Finally, the problem was also treated as a regression problem in order to exploit the continuous nature of road condition values. Specifically, a linear regression model was employed and its output was rounded to the nearest road condition class.

3.4.2 Training optimization: To optimize results of the training process certain techniques were considered and evaluated to optimize the classification accuracy.

The Normalized Difference Vegetation Index (NDVI) is a widely used index to quantify vegetation density using sensor data. NDVI (Normalized Difference Vegetation Index) masks were used, in order to mask out areas of the roads covered by dense vegetation during the road quality evaluation process and to take into account only the rest of the road area. It is calculated from spectrometric data in two specific bands: red (RED) and near-infrared (NIR), and is characterized by high accuracy correlation with the actual state of vegetation. The NDVI index is calculated by the following formula:

$$NDVI = \frac{NIR-RED}{NIR+RED}, \quad (1)$$

From the definition of NDVI, it follows that an area containing dense vegetation will be characterized by positive values, mainly in the interval [0.3, 0.8] (Carlson et al., 1997). Pixels with NDVI values equal or greater than 0.3 are considered to picture plants and not forest road so they were discarded.

The pan-sharpening process was used in order to combine the high spatial resolution panchromatic image data (0.3m) from the Worldview 3 satellite with the low resolution multispectral image (1.2m) and obtain a high resolution multispectral image. The benefit resulting from this particular process is that the pan-sharpened image has high spectral and spatial resolution allowing a more detailed analysis and visualization of the image. This particular image has more detail and color information, making it easier to identify objects and features. However, care is necessary because the pansharpening process does not preserve the spectral integrity of the data. For example, it is not recommended to perform NDVI analysis on a pansharpened image. Three pan-sharpening methods were tested. The nearest neighbor method, bilinear interpolation and cubic convolution. The specified methods determine how the pixel values of the output raster are determined after performing a geometric operation. The nearest neighbor method is best suited for categorical data such as land use classification. The values entered in the output grid remain exactly the same as the original grid and no new values are displayed, in contrast with the other two methods. The value of each pixel of the output grid is determined by the nearest pixel center in the input grid.
Also, extra features, including the detected dead-end and junction areas were used as additional features during the training process and were seen to lead to improved results. Finally, augmentation (random horizontal and vertical flips) was used for the training of CNN and ResNet architectures to avoid overfitting.

3.5 WebGIS

In order to support the display and management of the information by relevant stakeholders (forest authorities, public, etc.) an online Geographical Information System (WebGIS) supporting different information layers was developed. Great emphasis was put into the usability, security, interoperability, and scalability of the open-source WebGIS infrastructure, integrating all existing and produced information, enabling users to manage and process spatial data and to effectively visualize the results of the AI models. The WebGIS enables access from any computer through a browser, without requiring installation.

3.5.1 WebGIS Architecture: The WebGIS architecture is depicted in Figure 5, providing a top-level overview. The system comprises of two primary components: the front-end (client-side) and the back-end (server-side). The front-end constitutes the portion of the platform with which end-users interact. It was designed with a clean, intuitive, and straightforward front-end to ensure a user-friendly interface (UI) for both experienced and inexperienced users of geospatial data. Furthermore, it was developed to be compatible with various devices, ranging from large screens to smaller ones. Open-source languages, frameworks, and libraries were utilized for both the front- and back-end components of the application. Specifically, HTML5, CSS3, and JavaScript were employed for the UI, along with JavaScript libraries such as Cesium, Axios, and React. For the back-end, we utilized a GeoServer (WebGIS) server, allowing sharing, processing and editing of geospatial data files and a PostgreSQL / PostGIS database for storing additional spatial data. Libraries such as Node.js, Express.js were used and the Bcrypt algorithm by (Provos and Mazieres, 1999), which is a cryptographic hash function created using the Blowfish Algorithm (Schneier, 1993) and designed for password hashing and secure storage.

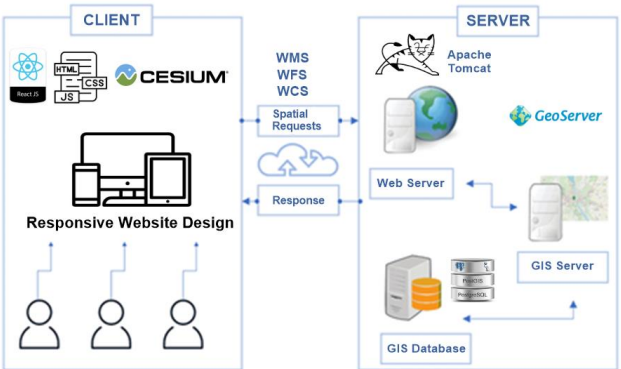


Figure 5: The WebGIS architecture (Server and Client side)

3.5.2 Main Functionalities: As depicted in Figure 6, the application comprises three main Sections accessible to users of any level. Section 1 features the map centred on the Seich-Sou peri-urban forest of Thessaloniki (i.e., the pilot area).

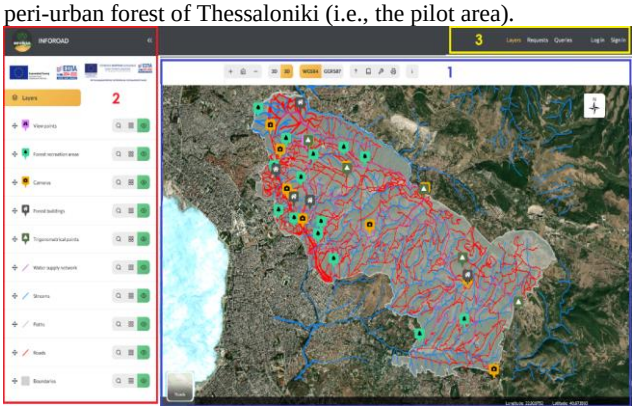


Figure 6: The project webpage and the three main sections of the application.

This section also includes all the necessary tools for interacting with the mapping application and guiding users. Section 2 serves as the repository for available geospatial information (Layers) and consists of functions for layer management. Section 3 hosts the application's main menu, which includes not only the main page (i.e., Layers) but also the pages 'Requests,' 'Queries,' 'Log in,' and 'Sign in'. Currently, the application is available only in Greek, but translation into English is being considered as a future step. The main functionalities of WebGIS can be categorized into the following broader areas:

Map interactivity: The WebGIS employs various front-end and back-end technologies to achieve map interactivity. These functionalities include zooming in and out, restoring the original (initial) map view, transitioning from a 2D to a 3D projection, switching base maps from satellite to roads, changing the geodetic system (from WGS84 to GGRS87), providing user instructions for map navigation on both computers (mouse/cursor) and mobile/tablet devices (touch), printing the current map view as an image with legend, and viewing and managing layers. Users can control layer appearance and hierarchy within the application.

User management system (UM): The WebGIS features a comprehensive user management system to handle activities related to individual access and functionalities within the application. Four levels of accessibility—User, Registered User, Administrator, and Super Administrator—have been constructed to address varying user needs. PostgreSQL facilitates access restriction to system data based on the user's role. Each level of accessibility inherits the capabilities of the previous one while offering additional functions. The Bcrypt algorithm ensures password security in the backend of the application. Users can upgrade to the registered status through the Sign-in page, while administrators have the authority to create registered users and manage user accounts through the administrator panel.

Layer management and spatial data visualization: The ability of all the application users to view and modify layer appearances, including colour, transparency, and line weight when applicable. Registered users and administrators can save their preferences for future sessions. Additionally, users can adjust the hierarchy of layers as needed. Administrators have additional privileges such as viewing, editing, and downloading geospatial information.

Requests: This feature serves as a communication channel between the public and local authorities. Through the 'Requests' page, any user can pinpoint a location on the map and describe a problem or offer a suggestion related to that point. Users also have the option to upload photos related to the point of interest. Administrators receive these requests and can manage them through the administrator panel.

Queries: Accessible to all users, the 'Queries' page offers essential tools for data filtering to obtain subsets. Users can access descriptive statistics for each available dataset and utilize various data classification methods for categorization. Queries are limited to individual road characteristics such as width, slope, accessibility, etc.

Future steps may include implementing changes to simplify user experience or further modernize the user interface to meet technological advancements.

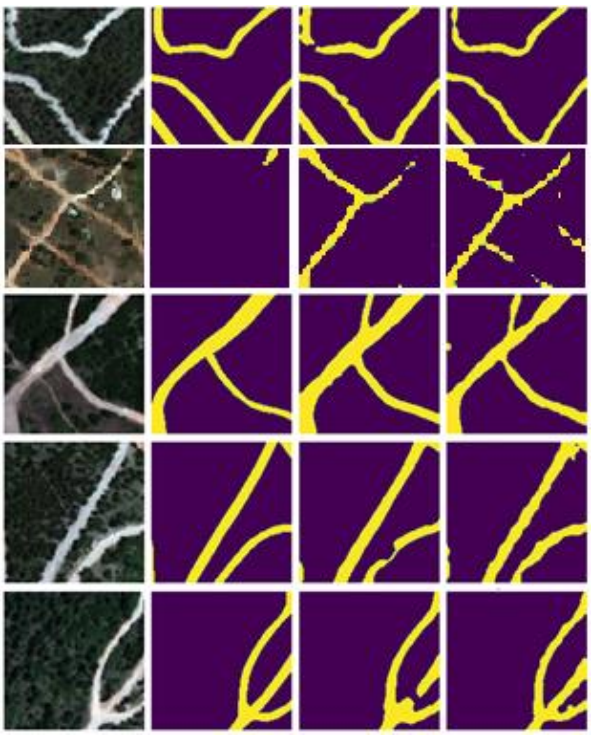
4. Results

Results from the road extraction and road condition monitoring modules are presented below.

4.1 Road Extraction Results

We compare results of the original U-TAE model and the modified alternative proposed here, using examples from the Seich-Sou forest as validation.

In , from left to right, we can see firstly the satellite image for a specific region of the forest, next the road segment from OSM, the results from the original U-TAE model, and the results from the proposed version, which uses inverted residual blocks in the input and output of the architecture.



(a) (b) (c) (d)

Figure 7: (a) RGB channels of satellite image, (b) Ground truth, (c) U-TAE prediction, (d) modified U-TAE prediction using inverted residual blocks

As seen in the second and third row, the predictions provided by the models may often detect new road segments not accounted for in the OSM road network, which may correspond to valid road segments that have not yet included in the ground truth. In Table 1, evaluation metrics for the two methods used for road extraction are presented. By jointly examining these metrics as well as the visual examples, one can argue that due to these roads missing from the ground truth, some “correct” road predictions may be erroneously treated by the metrics as false positives. This renders evaluation metrics less meaningful, given that they can only assess the ability of the network to mimic the –often inaccurate- ground truth, ignoring this important ability of detecting new unknown road segments. For this reason, visual assessment of the performance of the models is always very important. As seen, the evaluation metrics in Table 1 favour the original U-TAE architecture, while the visual examples favour the proposed method. Therefore, a more fair comparison of these models in the future will require re-training

after thoroughly validating the existence of any new roads and updating accordingly the ground truth.

Architecture	Accuracy	Precision	Recall	F1	IoU
U-TAE	97%	87%	84%	85%	77%
Proposed	96%	82%	80%	81%	71%

Table 1. Testing accuracy, precision and recall of different architectures for road quality classification

As regards to the connectivity of the road network, we can see in that the road connectivity provided by the proposed model is better, while also preserving the road width. The U-TAE model and the proposed modification can correctly extract road segments, even those not provided through the ground truth OSM. However, the proposed algorithm has the overall advantage of predicting more accurate road width and of increased connectivity between road segments.

4.2 Forest/rural road Condition Monitoring Results

The best results for each of the four architectures that were tested are presented in this section. For the CNN, the best results were achieved with the concatenation of the multispectral and the panchromatic image, random horizontal or vertical flip with a 40% probability, the use of NDVI masks with a threshold value of 0.3 and using dead-end and junctions as extra features during training. Image size used was 224x224, batch size was equal to 32 and learning rate was equal to 1e-3. With the above specifications, the model presented accuracy of 56%, precision of 71% and recall of 50%. In the case of ResNet50 architecture optimal results were accomplished with a similar setting as in the CNN case. The only differences are the use of the 3 RGB channels from the multispectral image instead of the concatenation of the multispectral and the panchromatic image and the use of a ReLU unit with batch normalization after Resnet. The results were as follows: accuracy 63%, precision 42% and recall 53%.

In the case of the PSE-MLP algorithm, the best results were achieved with the concatenation of the multispectral and the panchromatic image, the use of NDVI masks with a threshold value of 0.3 and the use of dead-ends and junctions as additional features during training. The batch size was set to 32 and the number of sample pixels for each edge was 128. This architecture achieved accuracy of 53%, precision of 56% and recall of 50%. In the case of kNN, with the concatenation of the multispectral and the panchromatic image and using k=4, we obtained the following results: accuracy of 53%, precision of 36% and recall of 45%. The above results are summarized in .

Architecture	Accuracy	Precision	Recall
CNN	56%	71%	50%
ResNet50	63%	42%	53%
PSE-MLP	53%	56%	50%
kNN	53%	36%	45%

Table 2. Testing accuracy, precision and recall of different architectures for road quality classification

Finally, when treating the problem as a regression task, a 53% accuracy was achieved by the linear regression model. Although this result is similar to the results obtained by the classification methods, an advantage of this approach is that a higher percentage of predictions are either correct or correspond to the neighbouring class.

5. Conclusions and Future Work

This article presented the final framework that was developed by the INFOROAD project, which integrates state-of-the-art technologies in remote sensing and machine learning for the automated and periodic extraction and condition monitoring of the forest road network. The methodology applies Worldview 3 high-resolution satellite images and diverse auxiliary data sources. The framework includes road extraction, road graph extraction, and road condition monitoring modules and was evaluated using data from the Seich-Sou forest in Thessaloniki. The road extraction module utilizes the U-TAE network architecture, incorporating a proposed adjustment with inverted residual blocks to enhance accuracy. Road graph extraction entails the creation of a graph from road segmentation output or OpenStreetMap (OSM) data, enabling efficient analysis of road segments. Road width calculation relies on road segmentation results, while road condition monitoring utilizes machine learning and artificial intelligence classification algorithms.

Once the proposed method is trained for road extraction and/or road condition monitoring, it can be applied to other forest areas using just a high-res satellite image as input. OpenStreetMap data and field measurements are needed only for the training process. In addition, Digital Elevation data can be used, as an auxiliary source, to estimate additional road features (e.g. slope, aspect).

Results highlight the efficacy of the proposed framework, demonstrating enhanced accuracy in road extraction through the modified U-TAE model. The proposed model was also able to detect even new roads that did not exist in the available forest maps. Furthermore, road condition monitoring results were encouraging, accomplishing a peak accuracy of 63% with the ResNet50 architecture for a 5-class classification task. This task had to deal with significant challenges, including the lack of related datasets, issues with data integration of data from different sources, outdated or inaccurate maps, the significant surface and texture variations of forest and agricultural roads – as opposed to asphalt roads as well as vegetation cover obstructing road visibility.

We believe that the project made a notable advancement in leveraging modern ML/AI technology towards sustainable forest road management. Future work will mainly involve further refinements and adjustments of the road condition monitoring module by addressing in a greater degree the aforementioned challenges in order to make it more robust and accurate.

Acknowledgements

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds under the framework of the Action ``Investment Plans of Innovation'' of the Operational Program ``Central Macedonia 2014-2020'' as part of the project ``INFOROAD – Development of an innovative online tool for mapping and monitoring the forest and rural road network'' Project code: KMP6-0079153. Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org was used.

References

Bahl, G., Bahri, M., & Lafarge, F., 2022. Single-shot end-to-end road graph extraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1403-1412.

Brewer, E., Lin, J., Kemper, P., Hennin, J., & Runfola, D., 2021. Predicting road quality using high resolution satellite imagery: A transfer learning approach. *Plos one*, *16*(7), e0253370.

Cadamuro, G., Muhebwa, A., & Taneja, J., 2019, July. Street smarts: measuring intercity road quality using deep learning on satellite imagery. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, 145-154.

Carlson, T. N., & Ripley, D. A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, *62*(3), 241-252.

Ferraz, A., Mallet, C., & Chehata, N., 2016. Large-scale road detection in forested mountainous areas using airborne topographic lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *112*, 23-36.

Garnot, V. S. F., Landrieu, L., Giordano, S., & Chehata, N., 2020. Satellite image time series classification with pixel-set encoders and temporal self-attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,12325-12334.

Garnot, V. S. F., & Landrieu, L., 2021. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4872-4881.

He, K., Zhang, X., Ren, S., & Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

He, S., Bastani, F., Jagwani, S., Alizadeh, M., Balakrishnan, H., Chawla, S., ... & Sadeghi, M. A. (2020). Sat2graph: Road graph extraction through graph-tensor encoding. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, 51-67. Springer International Publishing.

Jiang, X., Li, Y., Jiang, T., Xie, J., Wu, Y., Cai, Q., ... & Zhang, H., 2022. RoadFormer: Pyramidal deformable vision transformers for road network extraction with remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, *113*, 102987.

Kalpoma, K. A., Robin, G. R. K., Ferdaus, J., Mitul, M. M. R., & Rahman, A., 2022, July. Satellite Image Database Creation for Road Quality Measurement of National Highways of Bangladesh. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, 3047-3050. IEEE.

Kearney, S. P., Coops, N. C., Sethi, S., & Stenhouse, G. B., 2020. Maintaining accurate, current, rural road network data: An extraction and updating routine using RapidEye, participatory GIS and deep learning. *International Journal of Applied Earth Observation and Geoinformation*,*87*, 102031.

Kelesakis, D., Marthoglou, K., Grammalidis, N., Daras, P., Tsiros, E., Karteris, A., & Stergiadou, A. 2023, September. A holistic framework for forestry and rural road detection based on satellite imagery and deep semantic segmentation. *Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023)*, Vol. 12786, 126-134. SPIE.

Kolkos, G., Stergiadou, A., Kantartzis, A., Tampekis, S., & Arabatzis, G., 2023. Effects of forest roads and an assessment of their disturbance of the natural environment based on GIS spatial multi-criteria analysis: Case-study of the University Forest of Taxiarchis, Chalkidiki, Greece. *Euro-Mediterranean Journal for Environmental Integration*, *8(2), 425-440.*

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, *2*.

OpenStreetMap contributors. 2017. "Planet dump retrieved from https://planet.osm.org ." https://www.openstreetmap.org.

Provos, N., & Mazieres, D. (1999). Bcrypt algorithm. In *USENIX*.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).*

Sayers, M. W. (1995). On the calculation of international roughness index from longitudinal road profile. *Transportation Research Record*, (1501).

Schneier, B. (1993, December). Description of a new variable-length key, 64-bit block cipher (Blowfish). In *International Workshop on Fast Software Encryption*, 191-204. Berlin, Heidelberg: Springer Berlin Heidelberg.

Thegeya, A., Mitterling, T., Martinez Jr, A., Bulan, J., Durante, R. L., & Mag-atas, J., 2022. Application of Machine Learning Algorithms on Satellite Imagery for Road Quality Monitoring: An Alternative Approach to Road Quality Surveys. *Asian Development Bank Economics Working Paper Series*, 675.

Xia, W., Zhang, Y. Z., Liu, J., Luo, L., & Yang, K., 2018, March. Road extraction from high resolution image with deep convolution network—A case study of GF-2 image. *Proceedings*, Vol. 2, No. 7, 325. MDPI.

Yan, J., Ji, S., & Wei, Y., 2022. A combination of convolutional and graph neural networks for regularized road surface extraction. *IEEE transactions on geoscience and remote sensing*, *60*, 1-13.

Zao, Y., & Shi, Z., 2021. Richer U-Net: Learning more details for road detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1-5.

Zhang, X., Jiang, Y., Wang, L., Han, W., Feng, R., Fan, R., & Wang, S., 2022. Complex Mountain Road Extraction in High-Resolution Remote Sensing Images via a Light Roadformer and a New Benchmark. *Remote Sensing*, *14*(19), 4729.