

Article

Not peer-reviewed version

---

# Establishment of a Conceptual Framework to Detect Automatically the Urban Planning Offences Using Deep Learning and RGB orthophotos: Study Case of Morocco

---

[Mohamed Mellaki](#)\*, [Abderrazak El Harti](#), [Hassan Radoine](#), [Mohamed S. Chaabane](#), [Hassan J. Oulidi](#)

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0794.v1

Keywords: urban planning; offence; land management plan; orthophoto; deep learning; modelling; implementation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Establishment of a Conceptual Framework to Detect Automatically the Urban Planning Offences Using Deep Learning and RGB orthophotos: Study Case of Morocco

Mohamed Mellaki <sup>1,4,\*</sup>, Abderrazak El Harti <sup>1</sup>, Hassan Radoine <sup>2</sup>, Mohamed S. Chaabane <sup>3</sup> and Hassan J. Oulidi <sup>3</sup>

<sup>1</sup> Geomatics, Georesources and Environment Laboratory, Faculty of Sciences and Techniques (FST), Sultan Moulay Slimane University (USMS), Beni Mellal, Morocco

<sup>2</sup> UM6P-CITINNOV, Mohammed VI Polytechnic University, Ben Guerir, Morocco

<sup>3</sup> Hassania School of Public Works, Casablanca, Morocco

<sup>4</sup> Ministry of National Planning, Urban Planning, Housing and City Policy (MATUHPV), Rabat, Morocco

\* Correspondence: mellaki-mohamed@hotmail.com

## Abstract

Unregulated Housing (UrH) is a widespread urban phenomenon in Morocco, largely driven by rapid population growth and accelerated urbanization. It has expanded mainly on the outskirts of cities and within housing developments that already benefit from basic infrastructure and superstructure services. In response to this challenge, public authorities have adopted several urban planning instruments, particularly the Land Management Plan (LMP). According to Law No. 12-90 on urban planning, the LMP seeks to regulate urban expansion, improve the architectural and aesthetic quality of the built environment, and preserve the overall coherence of developed areas. As a legally binding planning document, the LMP establishes strict land-use regulations, and any breach of these rules constitutes an offence. Traditionally, detecting such violations requires on-site inspections by control officers, followed by the preparation of official reports submitted to the competent legal authorities. However, recent advances in aerial image acquisition and processing technologies provide powerful tools to improve and facilitate the monitoring of urban planning compliance. This paper proposes a conceptual framework that integrates artificial intelligence with urban planning regulations to enable the automatic detection of urban planning offences using RGB orthophotos covering areas subject to a Land Management Plan, relying on deep learning techniques.

**Keywords:** urban planning; offence; land management plan; orthophoto; deep learning; modelling; implementation

## 1. Introduction

Urban planning is a discipline that deals with the planning, development and management of urban areas in order to transform them into functional, sustainable and pretty spaces that take into consideration the present and future needs of their residents. It also aims to solve complex problems such as road congestion, population density, access to public services, preserving the natural and cultural heritage and promoting environmental sustainability [1].

To achieve this, territorial districts are equipped with urban planning documents, in particular the land management plan (LMP), which is an urban planning tool with legal force, since it is enforceable against administrations and third parties. It defines precise rules of land use [2]. It includes a graphic plan, a justifying note and planning regulations. It is therefore the legal basis for

controlling urban planning violations, due to the fact that it includes a distribution of zoning uses in a given urban area. Nowadays, inspections are carried out using the classic method, by control officers going out into the field or by visual interpretation of aerial images, with all the risk that human errors and weaknesses may occur during this operation, or that serious problems may arise with the population whose homes are subject to inspection and/or sanctions.

So, the question is, how can we get over the limitations of human control of urban planning and make it easier to detect offences? In other words, how can we overcome the weaknesses of the human factor in this process?

To answer this question, this work proposes the use of artificial intelligence, which offers a wide range of tools and possibilities, especially deep learning (DL) which, thanks to its performance and the great technological advances in computer hardware, has made progress in several areas, such as recognition and classification, extraction of objects from images [3], accurate agriculture [4], medical image processing, anticipation of time series, anomalies detection and weather classification [5]. Additionally, it allows the processing of a large amount of data with a very satisfactory cost/time ratio compared with human beings.

However, most existing research focuses on the tasks of detecting, classifying and extracting objects from orthophotos, such as buildings, without going as far as checking their legal conformity with urban planning regulations. For this reason, our subject will be inspired by research work related to the exploitation of artificial intelligence techniques, and in particular deep learning, to resolve issues relating to the estimation of building heights [6] and forest trees (canopy) [7], the buildings extraction from aerial images [3], extracting building footprints from orthoimages using deep learning [8, 9] and semantic segmentation of urban scene images from remote sensing using **transformers** [10]. [3] worked on the automatic extraction of land cover characteristics by applying deep learning to different spatial-temporal series of aerial orthophotos. The results obtained showed that the quality of the extraction improved as the spatial-temporal series of aerial orthophotos used increased. However, extraction accuracy varies depending on the type of construction, and acquiring time series orthophotos for model training is costly and must be done during the same season. [6] have carried out work on height prediction and refinement from aerial images with semantic and geometric guidance. They proposed a two-stage approach: the first involves the multi-task learning used to predict heights from an individual RGB image, thus going beyond conventional stereoscopic processes. The second step is to refine the results of the first, producing a height map with more accuracy. However, prediction errors are often concentrated at the building's edges because of rapid changes of brightness and colour, and trees where shadows create a considerable quantity of colour noise. [7] proposed an approach for classifying the vertical structure of the forest based on RGB orthophotos and Lidar data using ANN artificial neural networks, in order to overcome the issues of the traditional method based on inventories and field investigations. The test area was Gongju province in South Korea, which contains forest structures with one, two and three layers. After making the comparison between the results and data acquired by field survey, the overall accuracy obtained was approximately 70%. However, it should be noted that:

- There is a strong influence of the dominant species on the classification by ANN;
- Picture taking conditions must be the same for the area of the same image;
- It is an expensive method in terms of covering large study areas with dense LIDAR data.

[11] focused on extracting building footprints using deep learning techniques applied to drone-based orthoimages. The area of interest was an urban region along the Euphrates River in Iraq, where 474 aerial images were acquired. These images underwent several processing stages to generate digital surface models (DSMs) and orthophotos. Initially, building footprints were extracted using existing deep learning models; however, their performance was limited due to being trained on datasets from environments different from the study area. To address this limitation, the authors iteratively trained a Mask R-CNN model using orthophotos at spatial resolutions of 1.5, 10, and 20 cm per pixel. The newly trained models demonstrated promising performance during testing. The results confirmed the effectiveness and relevance of the proposed approach for building extraction from drone-derived orthophotos. Nevertheless, accurately delineating individual buildings remains a significant challenge in high-density urban areas, particularly where buildings are closely spaced

or physically adjacent. [9] worked on a study area located in Hsinchu, Taiwan, and used real aerial orthoimages with a resolution of 25 cm to detect built-up areas using the improved semantic segmentation method: **TransUNet**; this is a hybrid DL model between U-Net and vision transformers (ViT). An additional area covered by 7 km<sup>2</sup> of orthoimages was used as an independent validation area. The precision and recall values for building detection were 90% and 91% respectively. Once the boundaries had been regularized, the building footprints were successfully extracted from the identified areas. This proposed approach demonstrates the effectiveness of the automatic building footprint extraction by combining aerial orthoimagery with DL technology, greatly facilitating the production of accurate topographic maps. However, for this approach, irregular building boundaries are a problem commonly encountered in DL semantic segmentation of aerial images, in addition to the difficulties of identifying individual buildings, particularly in high-density urban areas. For their part, [10] worked on the effective semantic segmentation of images of urban scenes obtained from remote sensing using a ViT. They proposed a transformer-based **decoder** and developed a UNet-type transformer, which they called '**UNetformer**,' for real-time urban scene segmentation. To achieve an effective segmentation, they adopted in the **encoder** side the lightweight ResNet18 architecture and developed an effective **mechanism of global-local attention** to model global and local information in the **decoder** side.

Extensive experiments have shown that this approach is both faster and more accurate than current state-of-the-art lightweight models. In further research, the proposed decoder—based on a Transformer architecture combined with a Swin Transformer encoder—achieved the highest performance on the Vaihingen dataset. While the use of a Transformer-based encoder is justified by its capacity to capture rich semantic information, it substantially increases computational complexity, resulting in slower processing speeds that limit its applicability to large-scale urban scenes.

This study therefore addresses a significant gap in the current state of the art, which primarily focuses on detecting and extracting objects—especially buildings—from orthophotos, without considering their legal compliance. By integrating the automatic verification of building conformity with existing urban planning regulations, the proposed approach reduces dependence on labour-intensive field inspections by public authorities. Moreover, it allows control officers to perform assessments under safer and less stressful conditions, thereby supporting more objective, accurate, and informed decision-making.

The following is a comparative table of state-of-the-art studies (Table 1):

**Table 1.** Comparative table of state-of-the-art studies.

Study/Use case	Reference	Task targeted	Method/model	Used data	Limitations in the offences's detection
Application of a Deep Learning Method on Aerial Orthophotos to Extract Land Categories	Won et al., 2020.	Extraction Classification	Smaller VGG-Net	Spatial-temporal series of aerial orthophotos and cadastral maps	Extraction accuracy varies depending on the type of construction. Acquiring time series orthophotos for model training is costly, and they must be taken during the same season
Height Prediction and Refinement from Aerial Images with	Elhousni et al., 2021.	Prediction	DenseNet 121	2018 Data Fusion Contest (DFC) ISPRS Vaihingen	Prediction errors are often concentrated at the building's edges because of rapid changes of

Semantic and Geometric Guidance				(true orthophotos) Digital Surface Model DSM Semantic labels Surface normals maps	brightness and colour, and trees where shadows create a considerable quantity of colour noise
Classification of Forest Vertical Structure in South Korea from Aerial Orthophoto and Lidar Data Using an Artificial Neural Network	Kwon et al., 2017.	Classification	Artificial Neural Network ANN	RGB Orthophotos Lidar point cloud	Classification by ANN is strongly influenced by the dominant species or category Picture-taking conditions must be the same for the area covered by the same image Expensive method in terms of covering large study areas with dense LIDAR data Irregular building boundaries are a problem commonly encountered in DL semantic segmentation of aerial images Difficulty to identify individual buildings, especially in high-density urban areas
Building footprint extraction from aerial imagery through semantic segmentation techniques	Teo et al., 2024.	Extraction Semantic segmentation	TransUNet	True aerial orthoimages Digital Surface Model DSM	The decision to use the transformer as an encoder, although justified in order to obtain accurate semantic information, slows down the processing speed of the segmentation network, which affects the extraction of
UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery	Wang et al., 2022.	Extraction Semantic segmentation	UNetformer	UAVid and LoveDA datasets ISPRS Vaihingen dataset (true orthophotos) Potsdam dataset (true orthophotos)	

---

information on  
large urban scenes.

---

## 2. Offences of Urban Planning in Morocco

In a highly dynamic context marked by the launch and implementation of several strategic initiatives—such as advanced regionalisation, the national *Cities Without Slums* programme, and preparations for hosting the 2030 FIFA World Cup jointly with Spain and Portugal—addressing the phenomenon of unregulated housing, particularly urban planning violations, has become a critical priority. Tackling these issues is essential for improving the urban landscape, enhancing land supply, and strengthening the competitiveness of Moroccan cities. However, breaches of urban planning regulations and their adverse socio-economic impacts are not unique to Morocco; they are observed worldwide. This reality underscores the importance of reviewing and learning from relevant experiences at both African and international levels.

### 2.1. African and International Approaches of Managing or Detecting Urban Planning Offences

The issue of urban space management and the fight against unsanitary conditions and urban challenges is not specific to Morocco. Indeed, [12] analyse the challenges of urban planning in Africa, including violations due to inappropriate top-down planning, and suggest strategies that incorporate resilience and remote sensing to better manage uncontrolled urban expansion. They focus on 107 African cities and highlight the need for local policies to reverse environmental degradation such as the densification, the inner-city redevelopment, satellite or secondary cities development, and the creation of green cities.

Furthermore, according to [13], the occupation of public spaces, particularly by street vendors, poses a real challenge for authorities in southern Tanzanian cities, given their significant socio-economic contribution, accounting for 48% of the country's informal economy. Despite this contribution to the local economy of Tanzanian cities, especially Arusha, street vendors do not attract enough attention in urban planning practices and from the authorities responsible, which leaves them exposed to constant displacement and eviction from one place to another. [13] proposes the implementation of regulations and laws in favour of street vending in order to regulate street sales activities, take street vendors into account in urban planning, and overcome the problem of illegal occupation of public space.

According to [14], the city of Bhubaneswar (India), known for attracting millions of tourists each year, faces the same problem. One of the main challenges is keeping the city's main roads and public spaces free of vendors. The city has launched an innovative approach aimed at improving informal trade and better managing public spaces. This initiative was unique in that it developed a partnership model between public bodies, private business owners, and community stakeholders. In this model, permanent kiosks were allocated to street vendors in designated sales areas. Vendors cannot claim ownership of the land on which they operate. The Bhubaneswar Municipal Corporation (BMC) charges them an annual fee of 500 rupees per shop.

According to [15], in Algeria, particularly in cities such as Oran, building permits and construction practices illustrate widespread non-compliance with urban planning regulations. In practice, building permits—intended to function as preventive control instruments—often do not precede construction but instead are issued retrospectively to accommodate works that have already been completed. As a result, they are frequently reduced to a formal administrative approval with limited regulatory effectiveness. Although interest in building permits has increased in recent years due to their requirement by banks for granting construction loans, numerous alterations are commonly introduced during the construction phase. These include changes in building height and footprint, expansions in built-up area, and the conversion of residential spaces—particularly ground floors—into commercial premises. The subsequent application for a modification permit, typically required in cases of sale or inheritance, often enables the systematic regularization of these initially non-compliant constructions.

[16] emphasizes that slums remain a persistent challenge in France today. According to the Interministerial Delegation for Housing and Access to Housing (DIHAL), France still has more than

300 slums, hosting nearly 15,000 people (mostly European) in difficult living conditions. Around 20 departments are affected. To better equip those working in the field, DIHAL has produced a practical guide aimed to provide them with concrete benchmarks and inspiring examples. The aim is to give local authorities and partners the tools to develop local, tailored and sustainable solutions. The ultimate goal is not only to evacuate the sites, but to achieve a global slum clearance based on support and access to rights. The guide emphasizes the need to accurately map the sites using DIHAL's dedicated national platform.

In Turkey, [17] worked on detecting offences relating to exceeding regulatory heights in the study area of Istanbul, and particularly in the district of Guzeltepe, by using a couple of data: stereo satellite images KOMPSAT-3 and development plans. To get normalized digital elevation models (nDEM) and then to extract building heights, they create digital surface models (DSM) and digital elevation models (DEM) from this couple of data and make the difference between them. These extracted heights were then compared to the regulatory limits. According to the development plan for the neighbourhood, which had 2,141 buildings, the maximum permitted height was 12.5 m, or approximately four to five floors. The heights of building eaves are provided very precisely with LIDAR data and the average heights of the blocks of buildings are given by the nDSM. Due to differences in resolution, the results of the two datasets are different. When both datasets were compared to the development plan, only 1% of buildings went over the regulatory heights with the KOMPSAT model, while 16% exceeded the limits when using LIDAR data.

### 2.1. Offences Inventory in Morocco

The following is a list of offenses encountered on Moroccan territory:

#### 2.1.1. Depending on the Legal Status of the Project

- In authorised projects: building activity without respecting the authorised plan.
- In unauthorised projects: building activity without permission to build, illegal housing estates and illegal land division.

#### 2.1.2. Depending on the Occupied Sites

- Encroachment on third-party land: This includes occupying land owned by private individuals or legal entities. Observed cases often involve private land, state-owned land, habous land, forest land, maritime land, collective land, and military land.

- Non-compliance with urban planning regulations: This occurs when construction violates the designated urban planning zones outlined in regulatory documents such as the Land Management Plan (LMP). Examples include non-aedificandi zones, non-Altus Tolendi zones, strategic reserves, forestation areas, regulatory easements and setback zones (e.g., high- and medium-voltage power lines), public hydraulic domains (e.g., canals, dams, lakes, Chaabas), cemeteries, and classified public road areas.

- Occupation of protected or sensitive sites: This includes areas with high agricultural potential, riverbanks, protected historic sites (e.g., Volubilis in Meknes), Sites of Biological and Ecological Interest, nature reserves, and Ramsar wetlands.

- Occupation of high-risk areas: Buildings or structures located in areas prone to natural hazards—such as landslides, earthquakes, floods, or rockfalls—based on reference documents or specialized studies, including the Urbanization Ability Map (UAM).

#### 2.1.3. Other Types of Urban Planning Offences

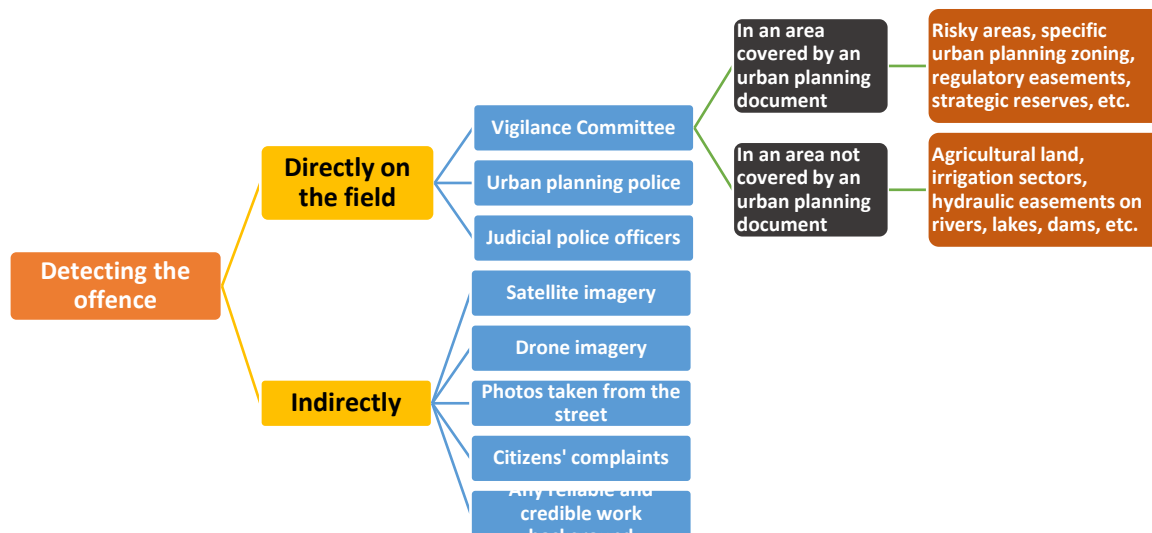
- Delivering of building permits by certain territorial districts without the approval of the Urban Agencies: This type of infringement puts the Urban Agencies in a conflict situation against the local councils, instead of working together in collaboration and cohesion and jointly dealing with Unregulated Housing;

- Non-respect of site management formalities: for example, site installations, working site barriers and all kinds of storage facilities that should warn the public of any risks, look good and be maintained in good condition;

-Building without earthquake-proofing measures in tectonically active areas.

#### 2.1.4. Business Process for Monitoring Urban Planning Offences

Collecting and acquiring offences data through



**Figure 1.** Job processes for detecting and monitoring planning offences (Source: Authors' construction, 2025).

Verification Example of an Indirect Finding of an Offence Concerning Non-Conformity with the Authorised Plan

- Sources of data to use

**Table 2.** Sources of data to use.

Source of data	Description
Drone images	in digital format or on paper, of the location of the study area including the building subject of the law-violation.
Land management plan (LMP) Urbanization ability map (UAM)	in usable digital format (vector) or on paper. with a zoom on the area where the law-violation is located.

- Georeferencing of the aerial image on CAD software: with the same projection system used in the LMP: Lambert Conformal Conic projection;
- Superimposing the image, the LMP and the UAM and zooming in on the site of the infringement;
- Assessment of building compliance: In general, a building is considered compliant if it adheres to the regulations of the Land Management Plan (LMP) and the Urbanization Ability Map (UAM). Conversely, any building that does not conform to the authorized plan constitutes an urban planning offence, as it violates the rules specified in the LMP. Furthermore, because the LMP is developed in accordance with the recommendations of the UAM, both the building site and the structure itself must align with these guidelines. This includes ensuring the building is located within an urbanable zone, verifying that the foundations are appropriately reinforced, confirming the suitability of the land for construction, and assessing potential risks from natural hazards such as landslides, earthquakes, or flooding.

### 3. Deep Learning of Aerial Photos

Techniques of Deep Learning have been mainly used in the field of the computer vision (CV), especially for tasks related to detection, classification and extraction. In the majority of researches,

DL models have been utilized in CV-based tasks, with a focus on visual data such as satellite and street images [18].

In the urban planning context, [19] have employed convolutional neural networks (CNNs) to classify urban land use by combining satellite imagery with social detection data. Likewise, [20] also demonstrated the adaptability of the deep learning using a geosystem-based model that enhanced the accuracy of classifying processing in environments where data is rare.

In the context of detection tasks related to infrastructures, [21] applied a custom DL model to Google Street View images to recognize and geolocate road signs. For feature extraction, [22] proposed the GL-Dense-U-Net model, enabling effective extraction of road networks from high-resolution imagery while overcoming challenges related to occlusion.

In all these applications, CNNs, U-Net, and attention-based Dense Nets are the most widely used DL models, mainly thanks to their performance to process complex, high-dimensional spatial data [18].

Furthermore, **Transformer**, initially developed for **natural language processing** (NLP), is a type of deep neural network primarily built on the **mechanism of self-attention**. Due to its powerful ability to represent information, researchers are studying ways to apply Transformer to computer vision (CV) tasks [23]. In vision applications, convolutional neural networks are considered the basic compounding element [24, 25], but today, transformers are emerging as a potential alternative to CNNs. [26] trained a sequence transformer to make autoregressive predictions of pixels, performing comparably to CNNs when applied on classification of images. Another vision transformer model is ViT, which uses a standard transformer to classify the entire image with image patches as input. Recent findings presented by [27] show that this model is capable to achieve state-of-the-art results in image recognition. In addition to image classification, the transformer has been employed to deal with various other vision tasks, such as object detection [28, 29], semantic segmentation [30] and video understanding [31].

From the above-mentioned, the choice of DL models based on convolutional neural networks (CNNs) and/or the Vision transformer is reasonable in order to solve the task of automatic detection of urban planning offences on orthophotos. However, vision transformer-based models requires large databases and therefore more time and money to be implemented.

## 4. Materials and Methods

### 4.1. Data and Materials

#### 4.1.1. Types of Data for Our Work

Our subject focuses on the automatic detection of urban planning violations using a DL model applied to orthophotos. To do this, we will need the following data:

-**High-resolution ortho-rectified aerial photos (orthophotos)** representing the terrain in the visible band (RGB) with probable use of satellite images in other bands (IR and NIR) for the characterisation and distinction of objects, textures, vegetation and buildings. Generally, a good starting point is to assign 80% of the data to the training set, 10% to the validation set, and 10% to the test set. The optimal distribution between the test, validation, and training sets depends on factors such as the use case, model structure, data dimension, etc. [32]. We therefore propose the following data distribution:

- For large image databases: 1% of this database is already representative of the whole for use in testing, 1% for validation and 98% for training.
- For small databases: between 60% and 80% to train models, between 10 and 20% to validate and the same interval to test.
- Images can be collected by searching for existing images of the study area or by taking aerial photographs using a drone.
- The model could be tested on orthophotos of a site other than the study area.

-**Approved land management plan (LMP)** covering the study area in the vector format: this represents the virtual and regulatory part that the urbanisation in the orthophotos must conform to;

- **Digital terrain model (DTM)** in raster (pixel) format: it represents the altitude of the natural ground without taking into account natural or man-made objects existing on the ground. In other words, it is the set of points corresponding solely to the elevation of the ground itself [33];

- **Digital surface model (DSM)**, which is a digital elevation model (DEM) in raster (pixel) format: it represents the altitude of the ground at each point on the surface, taking into account the height of the elements occupying that ground at each of its points [33]. In urban planning, DSM is also used to clearly identify the forms resulting from human activity on a site. It is essential for local authority planners in analysing the existing built-up areas and projecting its future development in terms of planning (e.g. road projects, large-scale development, etc.) [33].

- **Digital reproduction plan of the study area in vector format:** this also represents the reality on the ground at the date of its production and can be used, for example, to define rights of way, easements and roadways (public domain), buildings, and especially the ground altitude given by the elevation plan (contour lines + elevation points);

- **LIDAR (Light Detection And Ranging) data:** using an airborne sensor on a drone, LIDAR data represent the reality of the ground in 3 dimensions with high precision, in particular the altitude of the ground at each point in the study area, and thus enable us to evaluate the performance of DL models after training;

- **Topographic survey of calibration and control points:** specifying the planimetric coordinates,  $x$  and  $y$ , and altimetric coordinates  $h$  with high precision (in centimetre scale) and linked to the Lambert coordinate system (for planimetry) and the General Levelling of Morocco (NGM) (for altimetry);

- **Coordinates of reference points** for an accurate survey of the calibration and control points located in the study area and thus the production of a DTM and a DEM representing the ground reality;

- **Urbanization ability map** of the study area, if it is available after validation: this provides a solid scientific basis for identifying risk areas across the entire study area and the reinforcements to be taken into account during building actions;

- **Cadastral map** of the study area: this provides an overview of the parcel status of the registered lands and could explain, from a property view point, the pattern of the existing built-up area.

#### 4.4.2. Materials

The successful execution of the automatic detection of urban planning offences requires the following equipment:

- **Drone with a digital color camera:** To capture high-resolution aerial images of the study area, along with all necessary accessories for conducting a complete aerial photography mission in cases where existing orthophotos are unavailable.

- **Total station or GPS equipment:** For surveying and establishing calibration and control points to ensure accurate georeferencing of aerial imagery.

- **High-performance computer:** Equipped with advanced GPU and CPU capabilities to efficiently process high-resolution images and train deep learning models.

These tools are essential to ensure the quality and reliability of input data, as well as the computational performance needed for deep learning-based analysis.

#### 4.4.3. Software

The following software and tools are recommended for processing spatial data, preparing inputs, and training the deep learning model:

- **ArcGIS or QGIS:** For processing digital terrain models (DTMs) and digital elevation models (DEMs), as well as managing MXD, SHP, and TIFF files, including drone-acquired aerial photos and their mosaics.

- **AutoCAD:** For handling DWG files, such as reproduction plans, elevation plans, and the Land Management Plan (LMP).

- **Agisoft Metashape:** For processing aerial photos of the study area and generating DTMs, digital surface models (DSMs), and orthophoto mosaics.

-For the deep learning model training and accuracy assessment, the following environment setup is recommended:

-**Python language:** Version 3.10.16

-**TensorFlow environment:** Version 2.10.0

-**Keras library:** Included with TensorFlow version  $\geq 2.10.0$

These tools collectively enable the preparation of high-quality spatial data, support deep learning model development, and facilitate accurate detection and mapping of urban planning offences.

#### 4.4.4. Technical Process of the Automatic Detection of Urban Planning Offences

In this section, we describe the technical process for the automatic detection of urban planning offences and the production of an offences map for a given study area. As discussed in Section 3 on “Deep Learning of Aerial Photos,” deep learning models based on convolutional neural networks (CNNs) and/or vision transformers (ViTs) are well-suited for detecting offences from orthophotos. Each type of offence may require specialized processing due to differences in their nature and regulatory context. Examples include:

-**Buildings in prohibited areas:** To identify construction in zones restricted by the Land Management Plan (LMP), orthophotos should be classified into two categories: allowed areas and prohibited areas.

-**Exceeding regulatory building heights:** For offences involving additional floors—such as in residential zones with a maximum height of four floors plus a commercial ground floor (16 m total)—orthophotos can be classified using a digital elevation model (DEM) into two categories: zones  $\leq 16$  m and zones  $>16$  m.

-**Cemetery easement violations:** In buffer zones of 30 m around cemetery boundaries, orthophotos are classified into two classes: cemetery and easement zones (prohibited) versus areas outside these zones (allowed).

Developing deep learning models for automatic offence detection involves several key challenges:

-**Variety and complexity of offences:** The same offence may appear in multiple configurations, increasing the modeling effort required.

-**Data volume requirements:** Deep learning models are data-intensive and require large volumes of high-quality input for training, evaluation, and testing.

-**Data noise and preprocessing needs:** Input data must be carefully preprocessed to reduce noise and ensure accurate results.

-**Ground-truth limitations:** Incomplete information on parcel configurations and existing built-up areas may hinder compliance with minimum parcel sizes and the intended design of the LMP.

-**Technical expertise requirements:** Implementing algorithms capable of processing and detecting diverse offences demands advanced computational skills and domain knowledge in urban planning.

## 5. Results

As mentioned before, this work aims to setup a conceptual framework linking the use of DL for the detection and the extraction of objects, especially buildings, from orthophotos, to the checking process of the legal compliance of buildings with current urban planning regulations. Thus, we propose a workflow of urban planning offences detection with the following three main parts (Figure 2.):

- Part of input data and extracting information from it;
- Part of implementation and modelling of offences;
- Part of model outputs.

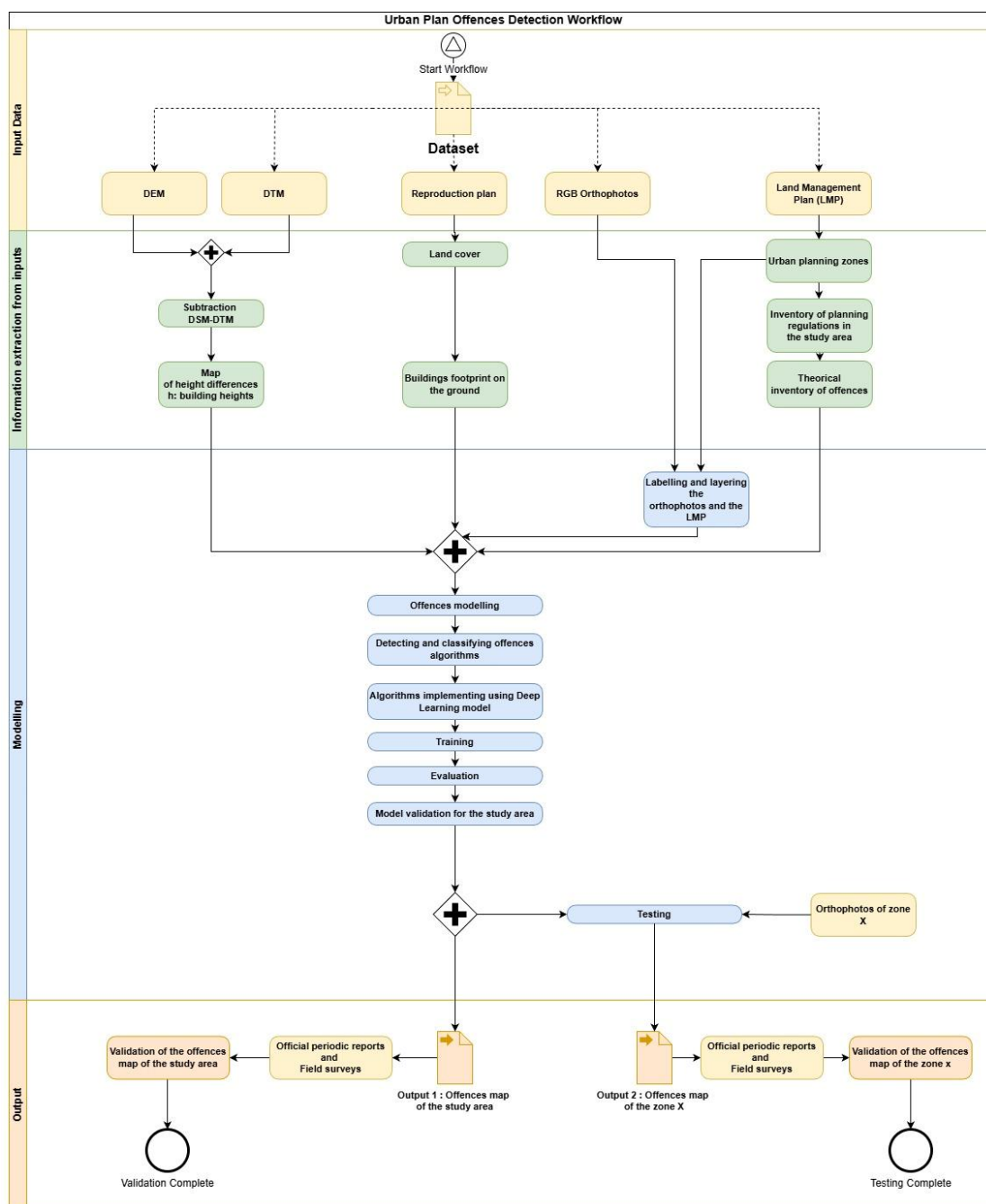


Figure 2. The workflow diagram (Source: Authors' construction, 2026).

### 5.1. Part of Input Data and Extracting Information from It

- The Land Management Plan** serves as the primary reference for identifying and cataloguing the urban planning regulations applicable to each zoning designation within the study area. It provides detailed specifications regarding permitted land use, occupancy conditions, and the linear and surface dimensions that must be respected by both public authorities and the population. Based on this regulatory inventory, a corresponding **theoretical inventory** of potential urban planning violations can be derived, with at least one type of violation associated with each planning rule and defined according to its specific requirements. For instance, cemetery easements, where all construction activities are prohibited, are typically defined by a 30-meter buffer zone; any construction within this zone constitutes a single type of violation. In contrast, regulations governing R+2 residential zones may be breached through multiple forms of non-compliance, such as the addition of extra floors beyond the permitted height or failure to respect the minimum parcel size requirements.

- Inventory of offences from official sources: This inventory is compiled using field reports from provincial monitoring committees, periodic reports from the relevant urban agencies, or through photo interpretation of time-series high-resolution aerial images. It provides insight into the most frequent types of offences occurring within the study area. It is important to distinguish this empirical inventory from the theoretical inventory, which is derived from a systematic listing of the LMP's land management regulations and the corresponding potential violations for each regulation.

- **RGB orthophotos:** at the end of the above-mentioned mission, RGB orthophotos of the entire study area are obtained with high spatial resolution, in the accuracy range centimetres, thus enabling, via orthophoto processing software on a computer (e.g. AgiSoft Metashape), the generation of a **mosaic**, a DEM and a DTM. Subtracting these last two elements will allow us to deduce a **map of the differences in altitude h** of objects on the ground, in particular **the height of buildings**, which can be controlled according to the LMP rules.

- **Reproduction plans:** These are obtained through detailed surveys of all existing features within the study area and represent the ground-truth land cover with precise X and Y coordinates. Altimetric information is provided through elevation plans, which include contour lines and spot elevation points. From reproduction plans, it is possible to extract the footprint of each surface element, particularly the footprints of buildings regulated by the LMP. The LMP specifies regulatory parameters for each parcel, including the minimum parcel size and the maximum allowable number of floors. The altimetry information is based on the general levelling of Morocco (NGM).

## 5.2. Part of Implementation and Modelling of Offences

- This stage begins with a critical process that requires highly skilled data scientists with strong domain knowledge: the annotation of the orthophoto mosaic and, where applicable, cropped orthophoto tiles used as model inputs. Labelling may be performed manually or with the support of specialized web-based annotation tools. The primary objective of this process is to accurately identify and extract relevant objects of interest, particularly buildings, which constitute the core targets of the detection and extraction model.

- Next, we define, on the orthophotos, the urban areas submitted to the LMP regulations based on the urban zoning designation polygons stipulated on the LMP by superimposing the LMP and the orthophotos. For example, we will have polygons for continuous residential areas with 3 floors (R+3), industrial areas of the first category, non-building zones, tree planting areas, strategic reserves, roads, squares, green spaces and sports fields.

- It is useful to have several examples of the same urban zoning designation on the orthophotos; this will allow the model to be fed with a large amount of data, which will facilitate its training.

- Each polygon of an urban zoning designation must be converted into a digital and usable file (e.g. .shp or .dwg) and must be accompanied by a metadata file describing the rules of the management regulations (urban zoning designation, minimum and maximum area, setback from roads and easements, height, number of levels, etc.), with which the deep learning model must interact to **detect** violations on the orthophoto;

- Thus, we will have orthophotos with the LMP zoning superimposed, showing both the land use (field truth) and the regulatory use provided for by the LMP in the form of polygons. Each polygon outlines a certain number of pixels that will be exposed to different layers of the CNN model to be developed or adjusted later according to the regulatory measures of the LMP.

- It should be noted that each type of urban planning offence requires its own dedicated detection algorithm, which will subsequently be implemented within a deep learning model for design, training, evaluation, and testing. This is necessary because offences vary significantly in nature and are subject to different regulatory measures—for instance, a violation of a cemetery easement must be treated differently from an unauthorized increase in building height. The possibility of combining multiple algorithms, or grouping related algorithms, into a single integrated model will be explored at a later stage.

- What stays to be done is to identify the anomalies on the orthophotos for each zoning of the LMP superimposed on them: in other words, to check the degree of the conformity of the field

truth with the regulatory measures of the LMP. When combining several infringement detection algorithms, the model must allow to look for one or more infringements through single or multiple queries (e.g. search for buildings exceeding 16m height (single query) or search for buildings exceeding 16m height with a footprint of less than 80m<sup>2</sup> (compound query));

- In the case of the multiple queries, we are talking not only about the detection of the offences, but also about **their classification** into a number of classes allowed by the model provided for this task.

### 5.3. Part of the Model Outputs

At the model output stage, a **spatial map** is generated **highlighting one or more urban planning offences occurring within the study area**. This map constitutes a powerful decision-support tool for stakeholders involved in urban planning and territorial management. The accuracy of the detected and/or classified offences will be evaluated by comparison with ground-truth data, including annotated orthophotos, reference maps, official inventories of recorded violations, digital elevation models (DEMs), and, where available, LiDAR-based 3D ground-truth reconstructions. Model performance will be further validated by testing on the remaining orthophotos from the same study area or by applying the model to a different neighbourhood where the presence and types of offences are known in advance.

## 6. Discussions

The growing complexity of the urban planning domain has led to the increased use of DL models, including CNNs, which offer greater accuracy and the ability to process high-resolution images [34, 35, 36, 37]. DL models can deal with massive data sets and recognize complex and unusual structures [38, 8]. However, they require considerable computational resources to operate efficiently [18].

Furthermore, the technical process presented in this article demonstrates that deep learning applied to orthophotos can specifically facilitate the automatic detection of urban planning offences. This approach addresses the limitations of traditional field-based inspections, which rely on committees to identify violations visually. Such conventional methods are prone to human error and operational weaknesses, and they may also generate conflicts with residents whose properties are subject to monitoring or sanctions. By automating the detection process, deep learning offers a more accurate, efficient, and safer alternative for urban planning control.

Moreover, the lack of previous studies addressing the automatic detection of urban planning violations using deep learning makes it difficult to conduct methodological comparisons. Collecting official data on the types of offences committed within the study area requires collaboration with competent public authorities, such as the provincial administration or the relevant Urban Agency, under whose territorial jurisdiction the area falls. Additionally, periodic field inspections provide essential monitoring of illegal housing and help maintain a database of recorded offences, offering a partial representation of the ground truth. However, these inspections are inherently limited by human capacity, making it impossible to achieve comprehensive coverage of the territory through regular monitoring alone.

When we compare the proposed method with the classical process of offences detection, we deduce that the first method is more objective because it is based on the processing of images or multi-temporal series of images taken by drones, using artificial intelligence (AI). While the classical method is influenced by the human interactions based on visual observations, measurements or manual inspection by human agents directly on the field. The Table 3 shows different aspects of both the two methods:

**Table 3.** Comparison between the two methods of urban planning offences detection.

Criterion	Classical control	Proposed method
-----------	-------------------	-----------------

	Very limited: one restricted area at a time.	
Spatial coverage	It depends on the weather, the topography of the field and on the agent availability.	Very large: monitoring of vast territories (municipalities, entire regions) in a single campaign.
Speed / Frequency	Slow: visits are spaced out and they depend on availability of the staff, Focused on a specific geographic point or after notification.	Very fast: possibly done several times a year, or even monthly. Large areas can be processed automatically in a short amount of time. Regular and repeatable, allowing us to compare images taken on different dates.
Cost	High (travel, labour, fuel).	Lower on a large area (after initial investment in images and software). Medium to good: it depends on the pixel size of the camera of the drone.
Resolution / Details	Very high: textures, materials, actual condition.	Details depend on the updated imagery. Medium to good: detects visible changes, such as new buildings or extensions, but there is a risk of false positives or negatives due to factors such as shadows, vegetation and image quality. Often requires verification on the field.
Accuracy and reliability	Very high: direct observation, precise measurements, possibility of checking the interior of buildings or details that are invisible from above. Few false positives.	Very low (obscured by trees, roofs, clouds)
Detection under cover	Possible (with direct access)	More objective (uses the same algorithm), but there is a possibility of training bias.
Objectivity	Subjective factors (e.g. agent interpretation, tiredness, mood).	Excellent (automatic image comparison of two times t and t-1).
Ability to detect changes	Depends on the agent's memory or previous visit. All: unauthorised construction; change of use; internal offenses; illegal occupation; detailed information (materials and condition, non-compliance with exact heights or interior use).	Mainly visible from above: building footprints, new buildings, extensions and the occupation of public spaces.
Types of offences detected		
Lawful evidence	Very strong evidence (direct observation, photos and an official report) may result in immediate penalties. Field inspections are risky.	Variable: Public satellite images (e.g. Google) are accepted, but drones are often contested due to privacy concerns. AI alone is rarely sufficient without field verification.
Innovation	No innovation in the classical control, the task is the same every time.	Large possibilities of innovation and progress.
Complementarity	Essential for confirmation and details	Excellent for sorting and prioritising suspicious areas.

Developing a deep learning model for the automatic detection of urban planning violations presents several additional challenges, including:

- **Variety and complexity of offences:** The same type of offence may occur in multiple configurations, increasing the effort required to model all possible scenarios within a single or multiple models.

- **Data volume requirements:** Deep learning models require large amounts of high-quality data for training, evaluation, and testing.

- **Data quality and noise:** Preprocessing of the data is necessary to reduce noise and ensure that inputs to the model yield accurate results that closely reflect the ground truth.

- **Technical and expertise-related challenges:** Implementing deep learning algorithms for the detection of multiple types of offences demands advanced technical skills and domain knowledge to effectively process and analyze the data.

To successfully achieve the automatic detection of urban planning violations from orthophotos using deep learning, it is essential to address the challenges outlined above. The following recommendations are proposed:

- **Develop a comprehensive inventory of offences:** Construct a clear theoretical inventory of potential violations based on the Land Management Plan (LMP), taking into account the regulatory measures for each offence. This enables precise differentiation between types of offences and informs the design of the detection algorithm.

- **Ensure technical and expertise requirements:** The implementation of deep learning algorithms requires sufficient technical skills, either individually or collectively within a research team, to design, implement, and optimize the models effectively.

- **Preprocess and correct input orthophotos:** Apply the necessary corrections to drone-captured orthophotos using specialized software to generate reliable mosaics, digital terrain models (DTMs), and digital elevation models (DEMs) that can serve as accurate inputs for the deep learning models.

- **Maintain spatial-temporal consistency of datasets:** Orthophotos, the LMP, topographic surveys, reproduction plans, and, if available, LiDAR data should be temporally aligned as closely as possible. This ensures that the deep learning models are trained and evaluated using consistent data. If discrepancies exist in the acquisition dates of the datasets, additional preprocessing will be required to verify and correct their consistency, which may increase processing time.

- **Future practical application:** Subsequent work will focus on a case study for the automatic detection of urban planning violations related to exceeding regulatory building heights within a neighborhood covered by the LMP. The approach will include:

- Presenting the different urban zoning designations and the regulatory height limits applicable to each;

- Designing a dedicated algorithm to detect height violations;

- Implementing this algorithm within a deep learning framework;

- Acquiring orthophotos through drone-based aerial photography;

- Preparing and processing the orthophotos and other necessary data for model input;

- Splitting the dataset into training, validation, and testing subsets;

- Training, evaluating, and testing the deep learning model;

- Assessing the model output by comparing the generated height violation map with field survey data.

By following these recommendations, the proposed framework will enable more accurate, efficient, and scalable detection of urban planning violations, supporting urban management authorities in monitoring compliance with planning regulations while minimizing reliance on traditional field inspections.

## 7. Conclusions

In conclusion, this article presents a conceptual framework for the automatic detection of urban planning offences using deep learning techniques applied to RGB orthophotos. The study first identifies the main types of urban planning violations encountered in Morocco, which largely depend on the legal status of the development project and the characteristics of the occupied site. It then outlines the conventional field-based procedures used for monitoring and controlling such offences.

The article further explores some applications of deep learning to aerial imagery, with particular emphasis on convolutional neural networks (CNNs) and transformers recognized for their effectiveness in classification, detection, and object extraction tasks from visual data, including satellite and street-level images.

In addition, the study describes the data requirements and technical resources necessary for the successful implementation of an automated offence detection system. A detailed technical workflow is proposed, structured around three core components: (i) input data acquisition and preprocessing, (ii) offence modeling and implementation, and (iii) model output generation. The final output of this workflow is a spatial map of detected urban planning offences within the study area. The accuracy of this map is evaluated through comparison with ground-truth data and available official reports related to urban planning control.

The significance of the proposed model lies in its potential deployment through an operational application capable of automatically detecting urban planning violations. Such an application would provide substantial support to professionals in urban planning and management, enhancing monitoring efficiency, reducing reliance on field inspections, and supporting more informed decision-making.

**Author Contributions:** Conceptualization, M.M., A.E.H., M.S.C. and H.J.O.; methodology, A.E.H., M.S.C. and H.J.O.; software, M.M., H.J.O.; validation, A.E.H., H.R., M.S.C. and H.J.O.; formal analysis, A.E.H., H.R., M.S.C. and H.J.O.; investigation, M.M., A.E.H. and M.S.C.; resources, M.M. and A.E.H.; data curation, M.M., A.E.H. and H.J.O.; writing—original draft preparation, M.M.; writing—review and editing, M.M., A.E.H., H.R., M.S.C. and H.J.O.; supervision, A.E.H. and H.J.O.; project administration, A.E.H.; funding acquisition, H.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding and the APC was funded by Mohammed VI Polytechnic University, Ben Guerir, Morocco.

**Data Availability Statement:** The original contributions presented in this work are included within this article's text, figures, and tables. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** We gratefully acknowledge all participants for their helpful revisions and remarks.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Networks
BMC	Bhubaneswar Municipal Corporation
CAD	Computer Aided Design
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CV	Computer Vision
DEM	Digital Elevation Model
DIHAL	Interministerial Delegation for Housing and Access to Housing
DL	Deep Learning
DSM	Digital Surface Model
DTM	Digital Terrain Model
DWG	Drawing
FIFA	Fédération Internationale de Football Association
FST	Faculty of Sciences and Techniques
GPS	Global Positioning System
GPU	Graphics Processing Unit
IR	Infrared
ISPRS	International Society for Photogrammetry and Remote Sensing
LIDAR	Light Detection And Ranging

LMP	Land Management Plan
MATNUHPV	Ministère de l'Aménagement du Territoire National, de l'Urbanisme, de l'Habitat et de la Politique de la Ville
MXD	Map Exchange Document
nDEM	Normalized Digital Elevation Model
NGM	General Levelling of Morocco
NIR	Near Infrared
NLP	Natural language processing
RADAR	RAdio Detection And Ranging
RGB	Red Green Blue
SHP	Shapefile
TIFF	Tagged Image File Format
UAM	Urbanization Ability Map
UAV	Unnamed Aerial Vehicle
UM6P	University Mohammed VI Polytechnic
UrH	Unregulated Housing
USMS	University of Sultan Moulay Slimane
ViT	Vision Transformer

## References

1. AURAV. (2023). IA et urbanisme : L'urbanisme et la ville de demain vus par l'intelligence artificielle. Agence d'Urbanisme Rhône Avignon Vaucluse. <https://www.aurav.org/documents/publication-iaeturbanisme.pdf>.
2. Agence Urbaine de Skhirate-Témara. (2022). Nouveau règlement d'aménagement. <https://aust.ma/sites/default/files/2022-03/Nouveau%20re%CC%80glement%20d%27ame%CC%81nagement.pdf>.
3. Won, T., Song, J., Lee, B., Pyeon, M.W. & Sa, J. (2020). Application of a Deep Learning Method on Aerial Orthophotos to Extract Land Categories. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 38, 443–453. <https://doi.org/10.7848/ksgpc.2020.38.5.443>.
4. Saad El Imanni, H., El Harti, A., Bachaoui, E.M., Mouncif, H., Eddassouqui, F., Hasnai, M.A. & Zinelabidine, M.I. (2023). Multispectral UAV data for detection of weeds in a citrus farm using machine learning and Google Earth Engine: Case study of Morocco. *Remote Sensing Applications: Society and Environment* 30, 100941. <https://doi.org/10.1016/j.rsase.2023.100941>.
5. Dahmane, K. (2020). Analyse d'images par méthode de Deep Learning appliquée au contexte routier en conditions météorologiques dégradées [Doctoral dissertation, Université Clermont Auvergne]. HAL Open Science. <https://theses.hal.science/tel-03022934>.
6. Elhousni, M., Zhang, Z. & Huang, X. (2021). Height Prediction and Refinement From Aerial Images With Semantic and Geometric Guidance. *IEEE Access* 9, 145638–145647. <https://doi.org/10.1109/ACCESS.2021.3122894>.
7. Kwon, S.-K., Jung, H.-S., Baek, W.-K. & Kim, D. (2017). Classification of Forest Vertical Structure in South Korea from Aerial Orthophoto and Lidar Data Using an Artificial Neural Network. *Applied Sciences* 7, 1046. <https://doi.org/10.3390/app7101046>.
8. Wu, A.N. & Biljecki, F. (2022). GANmapper: geographical data translation. *International Journal of Geographical Information Science* 36, 1394–1422. <https://doi.org/10.1080/13658816.2022.2041643>.
9. Teo, T.-A. & Chen, P.-C. (2024). Building footprint extraction from aerial imagery through semantic segmentation techniques. *IOP Conf. Ser.: Earth Environ. Sci.* 1412, 012036. <https://doi.org/10.1088/1755-1315/1412/1/012036>.
10. Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X. & Atkinson, P.M. (2022). UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 190, 196–214. <https://doi.org/10.1016/j.isprsjprs.2022.06.008>.

11. Ahmed, S.F., EL-Shazely, A.H. & Ahmed, W. (2025). Deep Learning for Building Footprint Extraction Using UAV-Based Orthoimages. *J Indian Soc Remote Sens* 53, 1243–1262. <https://doi.org/10.1007/s12524-024-02050-6>.
12. Kamana, A.A., Radoine, H. & Nyasulu, C. (2024). Urban challenges and strategies in African cities – A systematic literature review. *City and Environment Interactions* 21, 100132. <https://doi.org/10.1016/j.cacint.2023.100132>.
13. Mramba, N.R. (2025). The Conception of Street Vending Business (SVB) in Income Poverty Reduction in Tanzania. ResearchGate. <https://doi.org/10.5539/ibr.v8n5p120>.
14. Raj, B., & Jolly, J. (2020). Integrating Street vending in Urban planning. *International Journal of Science, Engineering and Management (IJSEM)*, 5, 25–32. [https://www.technoarete.org/common\\_abstract/pdf/IJSEM/v7/i4/Ext\\_27046.pdf](https://www.technoarete.org/common_abstract/pdf/IJSEM/v7/i4/Ext_27046.pdf).
15. Saïd, B. & Najet, M. (2025). L'urbain informel et les paradoxes de la ville algérienne : politiques urbaines et légitimité sociale. ResearchGate. <https://doi.org/10.3917/esp.143.0101>.
16. Localtis / Banque des Territoires. (2025, May 13). Bidonvilles en France : un défi persistant, des outils pour agir. <https://www.banquedesterritoires.fr/bidonvilles-en-france-un-defi-persistant-des-outils-pour-agir>.
17. Varol, B., Yilmaz, E.Ö., Maktav, D., Bayburt, S. & Gürdal, S. (2019). Detection of illegal constructions in urban cities: comparing LIDAR data and stereo KOMPSAT-3 images with development plans. *European Journal of Remote Sensing* 52, 335–344. <https://doi.org/10.1080/22797254.2019.1604082>.
18. Shaamala, A., Yigitcanlar, T., Nili, A. & Nyandega, D. (2025). Machine Learning Applications for Urban Geospatial Analysis: A Review of Urban and Environmental Studies. *Cities* 165, Article number: 106139. <https://doi.org/10.1016/j.cities.2025.106139>.
19. Li, J., Gao, J., Zhang, Z., Fu, J., Shao, G., Zhao, Z. & Yang, P. (2024). Insights into citizens' experiences of cultural ecosystem services in urban green spaces based on social media analytics. *Landscape and Urban Planning* 244, 104999. <https://doi.org/10.1016/j.landurbplan.2023.104999>.
20. Yamashkin, S.A., Yamashkin, A.A., Zanozin, V.V., Radovanovic, M.M. & Barmin, A.N. (2020). Improving the Efficiency of Deep Learning Methods in Remote Sensing Data Analysis: Geosystem Approach. *IEEE Access* 8, 179516–179529. <https://doi.org/10.1109/ACCESS.2020.3028030>.
21. Campbell, A., Both, A. & Sun, Q. (Chayn). (2019). Detecting and mapping traffic signs from Google Street View images using deep learning and GIS. *Computers, Environment and Urban Systems* 77, 101350. <https://doi.org/10.1016/j.compenvurbsys.2019.101350>.
22. Xu, Y., Xie, Z., Feng, Y. & Chen, Z. (2018). Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sensing* 10, 1461. <https://doi.org/10.3390/rs10091461>.
23. Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y. & Tao, D. (2023). A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 87–110. <https://doi.org/10.1109/TPAMI.2022.3152247>.
24. He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
25. Ren, S., He, K., Girshick, R. & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
26. Chen, M., Radford, A., Child, R., Wu, J., Jun, H., Luan, D. & Sutskever, I. (2020). Generative Pretraining From Pixels, in: Proceedings of the 37th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 1691–1703.
27. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. <https://doi.org/10.48550/arXiv.2010.11929>.
28. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers, in: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), *Computer Vision –*

- ECCV 2020, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 213–229. [https://doi.org/10.1007/978-3-030-58452-8\\_13](https://doi.org/10.1007/978-3-030-58452-8_13).
29. Zhu, X., Su, W., Lu, L., Li, B., Wang, X. & Dai, J. (2021). Deformable DETR: Deformable Transformers for End-to-End Object Detection. <https://doi.org/10.48550/arXiv.2010.04159>.
  30. Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H.S. & Zhang, L. (2021). Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6877–6886. <https://doi.org/10.1109/CVPR46437.2021.00681>.
  31. Zhou, L., Zhou, Y., Corso, J.J., Socher, R. & Xiong, C. (2018). End-to-End Dense Video Captioning with Masked Transformer, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8739–8748. <https://doi.org/10.1109/CVPR.2018.00911>.
  32. Pragati, B. (2021, September 13). Train test validation split: How to & best practices [Blog post]. V7 Labs. <https://www.v7labs.com/blog/train-validation-test-set>.
  33. Doungmo, Y. (2017, August 10). Différence entre MNT, MNA, MNE et applications [Blog post]. LinkedIn. <https://www.linkedin.com/pulse/diff%C3%A9rence-entre-mnt-mna-mne-et-applications-yannick-arthur-doungmo>.
  34. Hu, P., Cheng, J., Li, P. & Wang, Y. (2023). Automatic extraction of built-up areas in Chinese urban agglomerations based on the deep learning method using NTL data. *Geocarto International* 38, 2246939. <https://doi.org/10.1080/10106049.2023.2246939>.
  35. Park, C., No, W., Choi, J. & Kim, Y. (2023). Development of an AI advisor for conceptual land use planning. *Cities* 138, 104371. <https://doi.org/10.1016/j.cities.2023.104371>.
  36. Rachele, J.N., Wang, J., Wijnands, J.S., Zhao, H., Bentley, R. & Stevenson, M. (2021). Using machine learning to examine associations between the built environment and physical function: A feasibility study. *Health Place* 70, 102601. <https://doi.org/10.1016/j.healthplace.2021.102601>.
  37. Wu, A.N. & Biljecki, F. (2023). InstantCITY: Synthesising morphologically accurate geospatial data for urban form analysis, transfer, and quality control. *ISPRS Journal of Photogrammetry and Remote Sensing* 195, 90–104. <https://doi.org/10.1016/j.isprsjprs.2022.11.005>.
  38. Chai, J., Zeng, H., Li, A. & Ngai, E.W.T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications* 6, 100134. <https://doi.org/10.1016/j.mlwa.2021.100134>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.