ACTA IMEKO
ISSN: 2221-870X
December 2023, Volume 12, Number 4, 1 - 8

Evaluation of spatial patterns accuracy in identifying built-up areas within risk zones using deep learning, RGB aerial imagery, and multi-source GIS data

Alessandro Vitale[1], Carolina Salvo[1], Francesco Lamonaca[2,3]

[1] *Department of Civil Engineering, University of Calabria, Via P. Bucci, Cubo 45 B , 87036 Rende (CS), Italy*
[2] *Department of Computer Engineering, Modeling, Electronics and Systems, University of Calabria, Via P. Bucci, Cubo 42D, 87036 Rende (CS), Italy*
[3] *National Research Council of Italy, Institute of Nanotechnology (CNR-NANOTEC), Rende, Italy*

ABSTRACT
In the presence of natural disasters that increasingly affect urban centers, innovative methodologies that can support all the subjects and bodies involved in the disaster management system are increasingly important. This task can be enhanced in urban settings by automatically assessing at-risk buildings through satellite and aerial imagery. However, creating and implementing models with robust generalization capabilities is crucial to achieving this goal. Based on these premises, the authors proposed a deep learning approach utilizing the U-Net model to map buildings within known landslide-prone areas. They trained and validated the U-Net model using the Dubai Satellite Imagery Dataset. The model's prediction accuracy in adapting its results to urban environments in Italy, different from those involved in the training and validation stages, was tested using natural color orthoimages and diverse geographic information system (GIS) data sources.
The outcomes indicate that the model's predictions are better in contexts with denser urban fabric. The level of accuracy in dispersed urban shapes worsens as building footprints cover a small portion of the total image area. Overall, the results demonstrate that the suggested methodology can effectively identify buildings in landslide risk zones, demonstrating noteworthy adaptability, making the proposed platform a tool that can be instrumental for decision-makers and urban planners in pre-disaster and post-disaster stages.

Section: RESEARCH PAPER

Keywords: Building identification; accuracy spatial patterns assessment; remote sensing; RGB images; deep learning; U-net segmentation; geographic transferability; landslide risk; spatial planning

Citation: Alessandro Vitale, Carolina Salvo, Francesco Lamonaca, Evaluation of spatial patterns accuracy in identifying built-up areas within risk zones using deep learning, RGB aerial imagery, and multi-source GIS data, Acta IMEKO, vol. 12, no. 4, article 27, December 2023, identifier: IMEKO-ACTA-12 (2023)-04-27

Section Editor: Laura Fabbiano, Politecnico di Bari, Italy

Received November 16, 2023; **In final form** December 4, 2023; **Published** December 2023

Corresponding author: Alessandro Vitale, e-mail: alessandro.vitale@unical.it

1. INTRODUCTION

Automatically delineating the footprints of buildings and infrastructure from high-resolution imagery is paramount, particularly in urban planning applications, with a strong focus on disaster management planning [1], [2]. Identifying accurately buildings and infrastructure is critical for monitoring urban development [3]-[5].

In contemporary literature, developing dependable and precise methods for extracting buildings has emerged as a prominent and intricate research concern, garnering increased attention [6]-[9]. As a discipline, remote sensing provides invaluable insights to researchers and authorities grappling with this challenge.

Over the past few years, numerous methodologies for extracting buildings and infrastructure have been developed and applied, encompassing parametric and non-parametric classifiers, shadow-based techniques, edge-based approaches, and object-based methods [10]-[12]. Before Deep Learning (DL) emerged, the remote-sensing research community transitioned its focus away from neural networks. Instead, it focused on methods like support vector machines (SVM) and ensemble classifiers, including random forest (RF), for tasks like image classification. For instance, Belgiu and Drăguţ [13] employed and compared supervised and unsupervised multi-resolution segmentation methodologies in conjunction with RF classification to extract buildings from high-resolution satellite images. The comparison of image object geometries was conducted using empirical

discrepancy methods, also known as supervised segmentation evaluation, to gauge the geometric disparities between the generated image objects and a set of reference data. The employed global accuracy metrics are the area fit index (AFI), quality rate (Qr), and Root Mean Square (Dij). These metrics take into consideration the entirety of the imagery for evaluation purposes. Specifically, the Dij metric combines two key aspects: under-segmentation (USeg) and over-segmentation (OSeg) metrics. It evaluates the degree of "closeness" between the image objects and the reference data. OSeg occurs when the image objects are smaller than the reference objects, while USeg occurs when the image objects are larger. Following a revised object-based segmentation approach, Chen et al. [14] identified buildings using three machine-learning classifiers (AdaBoost, RF, and SVM). To assess the automatically detected results by overlapping them with the ground-truth map, these regions are categorized into four groups: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). Accuracy, Precision, Recall, Specificity, and F_1-score have been employed as metrics to gauge the overall model performance.

Starting in 2014, the remote-sensing community has redirected its focus toward DL models, which can autonomously acquire features directly from unprocessed satellite imagery. This shift eliminates the necessity for manual feature engineering and the processing of extensive data volumes. DL technology has found widespread application in remote sensing in recent years. Its potential lies in its capacity to surmount the limitations of conventional classification algorithms.

The Convolutional Neural Network (CNN) is the primary choice among various deep learning technologies for computer vision tasks [15]. Starting in 2014, algorithms based on CNN for semantic segmentation [16] have found applications in numerous pixel-wise analysis tasks concerning remote sensing imagery. Algorithms like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep neural networks (DNNs) have been demonstrated to deliver superior outcomes in remote sensing imagery classification and segmentation compared to conventional machine learning techniques [17]-[20]. Among the CNN-based network architectures, the U-Net model stands out due to its unique U-shaped design, the inclusion of skip connections, and multi-resolution feature integration. These attributes bolster its performance in semantic segmentation tasks, facilitating precise and comprehensive land cover classification. Consequently, U-Net has become a preferred choice for mapping and monitoring applications [21]-[26].

Furthermore, in recent times, by harnessing the substantial potential of remote sensing (RS), geographic information systems (GIS), and advanced deep learning techniques such as U-Net, researchers and professionals can capitalize on the unique advantages of each component to enhance the precision and effectiveness of land cover mapping and analysis across diverse applications. These applications encompass urban and regional planning and administration, sustainable management of natural resources, and the detection of global environmental changes [27]. For instance, Li et al. [28] introduced a U-Net-based semantic segmentation method to extract building footprints from high-resolution multispectral satellite images of four cities (Las Vegas, Paris, Shanghai, and Khartoum). They integrated geographic information system (GIS) map datasets like OpenStreetMap, Google Maps, and MapWorld into their approach. The Precision, Recall, and F_1-score are computed as accuracy measures. Using the Massachusetts building dataset, Alsabhan and Alotaiby [29] compared building extraction results from high-resolution satellite images using U-Net and Unet-ResNet50 models. As an accuracy metric, the IoU, also referred to as the Jaccard Index, is the employed metric, serving as a straightforward yet highly valuable and efficient evaluation measure. The IoU metric calculates the overlap between the predicted segmentation area and the underlying area by dividing it by the combined area where the predicted segmentation and actual area intersect.

Recognizing the lack of methodologies that investigate the spatial accuracy and transferability of the results obtained from urban features mapping, the authors used remote sensing, deep learning, and GIS in a combined way to identify buildings at potential risk locations through RGB aerial images and enhance the model's spatial transferability. To gauge the methodology's effectiveness in identifying at-risk buildings, this study applies a semantic segmentation-based approach for the semi-automated assessment of buildings within landslide-prone areas. This research intends explicitly to investigate the statistical levels of accuracy of the segmentation model, also considering any differences in accuracy in the prediction due to the different urban morphologies of the application areas.

The authors advocate a deep learning methodology founded on the U-Net model, a Convolutional Neural Network (CNN) that lends itself to more precise segmentation, even with limited training images. The model undergoes rigorous training and validation using the "Semantic Segmentation of Aerial Imagery" dataset, comprising 72 Dubai satellite images. Subsequently, its adaptability and transferability are tested by identifying and segmenting buildings within landslide-prone zones in the Calabria Region, Southern Italy. It's important to note that the accuracy of the U-Net model's segmentation capabilities depends heavily on the quality of the input data (RGB images), the representation of spatial features within these images, and the training process, including data augmentation techniques employed.

The development of a universally applicable platform, confident in its performance across various geographical contexts, can play a crucial role in enhancing community resilience through better-informed urban planning and emergency response strategies and assumes paramount significance as it obviates the need for repetitive training and validation procedures, thereby expediting the estimation of damaged buildings.

The remainder of this paper unfolds as follows: Section 2 furnishes insights into the study area, along with details concerning the training, validation, and testing datasets. Section 3 presents the proposed methodology, encompassing data preparation and the semantic segmentation model deployed for extracting building footprints. Building extraction results are expounded upon in Section 4, while Section 5 delves into the discussion and presentation of results for the case study. Finally, Section 6 offers a summary of the research's conclusions.

2. DATASET AND STUDY AREA

2.1. Dataset for Training and Validation and Data Augmentation Process

The authors utilized the "Semantic segmentation of aerial imagery" dataset for the training and validation phases of the proposed U-Net model. This dataset, created in collaboration with the Mohammed Bin Rashid Space Center in Dubai, UAE, is in an open-access format [30], and it comprises 72 high-resolution satellite images of Dubai, each accompanied by corresponding semantic segmentation masks. These satellite

images are categorized into six land cover classes, encompassing water, land, road, building, vegetation, and unlabelled areas.

The training set employs 65 images in this study, while the validation set comprises seven images.

The accuracy of deep learning models, particularly in Computer Vision (CV) techniques, is intricately linked to the training data's quality, quantity, and contextual significance. As stated in the introduction section, one of the significant challenges in deep learning approaches for image semantic segmentation tasks revolves around the limited availability of data suitable for training and validation. There has been an increasing emphasis on employing data augmentation techniques to generate well-qualified training and validation datasets to address this constraint. The primary objective of data augmentation is to enhance the adequacy and diversity of training and validation data by creating synthetic datasets [31]. It can be conceptualized as deriving the augmented dataset from a distribution closely resembling the original one, thus enabling a more comprehensive configuration.

U-Net models have many parameters, making them robust but prone to overfitting, especially when the amount of training data is limited. Data augmentation artificially increases the size of the training dataset by introducing variations, helping the model to generalize better to unseen data.

In the context of this research, the authors employed a range of data augmentation strategies for both the training and validation datasets, applying them to each image and mask randomly. Data augmentation exposes the model to various features, shapes, sizes, and orientations. This prevents the model from learning irrelevant patterns and enhances its generalization ability to different urban forms. Indeed, there may be an imbalance between the number of building pixels and non-building pixels in urban aerial imagery. Data augmentation can help to balance the dataset by emphasizing rare cases, such as small buildings or buildings with unusual shapes.

These augmentation operations were applied to the input images and their semantic segmentation masks. One of the strategies involves under-sampling, where an image is randomly cropped multiple times, each with a crop size of 512 × 512. This multiple cropping operation prioritizes the selection of regions with the highest sum of pixel values, thereby reducing the frequency of non-building entities in the samples. Another strategy adopted is a cost-sensitive approach. It involves constructing the loss function by combining the losses for each category with the total loss while assigning different weights based on the category proportions.

Various transformations were applied to enhance the model's robustness, including random flipping, rotation, translation, side view transformations, and zooming for the input images and their corresponding masks. Additionally, all input images underwent a random manipulation, adjusting their saturation, contrast, brightness, changes in colour space and band order, and introducing Gaussian noise and random filtering operations [32].

2.2. Study Area and Testing Dataset

The Calabria Region, situated in southern Italy, as of 2022, has an approximate resident population of 1,842,615 people, spread across 404 municipalities. Many of these municipalities have a population of 3,000 inhabitants. Due to its specific geological characteristics, the entire region is highly susceptible to various natural hazards, such as hydrogeological and seismic risks. Consequently, the rapid and accurate extraction of buildings within high-risk areas from remote sensing images is crucial for emergency management applications.

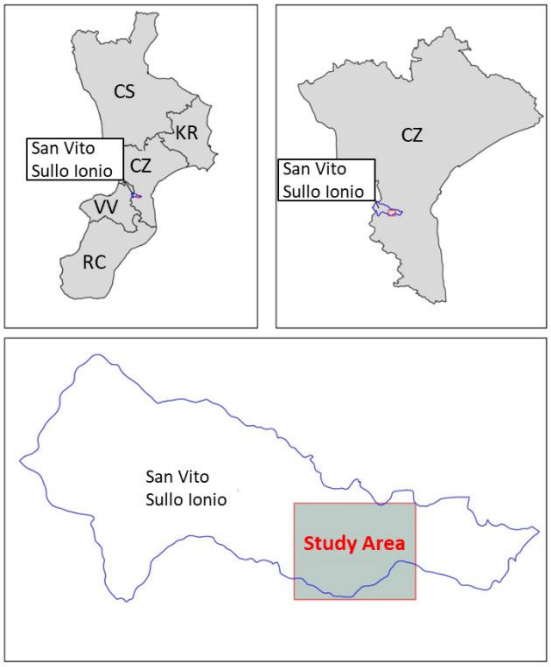

Figure 1. Study area framework.

The authors utilize information concerning landslide risk, including its location and severity level, as defined by the Hydrogeological Structure Plan, a tool used in Italy for basin planning, which addresses different types of risks, including landslides, floods, and coastal erosion.

The trained and validated model is applied to the urban area of San Vito sullo Ionio within the Calabria Region (Figure 1). The inhabited region is concentrated in the southeast area of the municipal borders. Woods and agricultural land occupy the rest of the surface. This municipality is exposed to landslide risk, and its urban form is polycentric and highly dispersed. It comprises a principal denser compact nucleus representing the municipality's core and a diffused rural area around the central core. This case study is particularly suitable for testing the capability of the model to classify and segment buildings in dense and dispersed urban regions. Moreover, it allows the identification of eventual spatial patterns in the model's accuracy.

For the testing stage of the model, the dataset employed consists of 0.423-meter pixel resolution (approximately 1-foot) natural colour orthoimages covering the considered municipality. The building polygons predicted by the model are compared with those retrieved from the Regional Technical Cartography (CTR) [33], which represents, in this study, the ground truth dataset. As previously mentioned, landslide risk areas are localized, considering the Hydrogeological Structure Plan [34].

3. MATERIALS AND METHODS

3.1. Summary of the Methodological Approach

The central component of the proposed methodology revolves around a deep learning model based on a Convolutional Neural Network (CNN) [35], [36] designed specifically for automatically delineating built-up areas and subsequently identifying buildings at risk of landslides. A set of parameters or weights characterizes this neural network. A training phase is conducted to assess the mapping function's effectiveness. During this training phase, input

tiles are passed through the network, and a loss function is computed. The network's weights are adjusted iteratively to minimize this loss. The architecture of the neural network plays a crucial role in its performance. In this research, we employ a U-Net architecture. The U-Net model is trained using the "Semantic segmentation of aerial imagery" method and utilizes the IoU-based loss function. The authors incorporate various data augmentation techniques to enhance the diversity and sufficiency of the training and validation data, creating a new synthetic dataset. These augmentation methods are applied to the input images and their corresponding semantic segmentation masks.

Performance evaluation is carried out using Recall, Precision, and F_1 metrics. The output of the trained model is a Semantic segmentation mask that categorizes segmented elements into six classes: water, land, road, buildings, vegetation, and unlabelled. For this study, we focus solely on identifying buildings. Subsequently, the model is tested on an urban area within the Calabria region to assess its ability to classify accurately and segment buildings vulnerable to natural hazards.

The output Semantic segmentation mask produced by the model is imported into a Geographic Information System (GIS) platform and vectorized. The authors utilize the open-source QGIS Desktop software as our GIS platform for this research. To evaluate the model's accuracy when applied to the case study, the authors compare the building polygons generated by the model with building polygons from the Regional Technical Map, calculating the F_1-score metric. Specifically, the comparison is made for the entire built-up area and the buildings within landslide risk zones. For the whole built-up area, the authors, using a buffering spatial analysis, evaluated the difference in accuracy in the model's predictions between the centre of the municipality under study, characterized by a greater density of buildings, and the rural outskirts, depicted mainly through scattered houses.

3.2. U-Net model training and validation

The authors selected the Adam optimization method to train the U-Net semantic segmentation model. Choosing an appropriate loss function is pivotal for ensuring model accuracy and typically hinges on data characteristics and class definitions. By adjusting the learning rate during training, Adam leads to faster convergence than methods with a constant or manually decayed learning rate. This can be crucial when training complex models like U-Net, where the computation cost is high. Furthermore, this optimization method is well-suited for problems with non-stationary objectives, common in urban feature segmentation due to the varying shapes, sizes, and textures of buildings in urban landscapes.

This study considered the IoU-Based Loss, also known as the Jaccard Index IoU-Based Loss.

The IoU (Intersection over Union), reported in equation (1), also known as the Jaccard index, is a widely adopted region-based performance metric for image segmentation tasks. The IoU is naturally robust to class imbalance (common in building segmentation, as building footprints often cover a small portion of the total image area). It does not treat every pixel equally but focuses on the proportion of correctly predicted building pixels to the total predicted and actual urban feature pixels. By focusing on the intersection over union, the Jaccard loss emphasizes the quality of the overlap between the predicted and ground truth segments. This is crucial for segmentation tasks, as it encourages the model to predict segments that closely match buildings' actual shapes and boundaries. Its values range from 0 (indicating no similarity) to 1 (indicating complete agreement between predictions and ground truth labels). To incorporate IoU into the optimization process, the authors computed the IoU-based loss function based on equation (2).

$$IoU = \frac{TP}{TP + FN + FP}, \tag{1}$$

$$IoU_{\text{Based}}Loss = 1 - IoU\,. \tag{2}$$

In Equation (1), 'TP' represents True Positives (correctly predicted buildings), 'FP' signifies False Positives (non-building entities wrongly identified as buildings), and 'FN' denotes False Negatives (actual buildings incorrectly classified as non-buildings).

The optimization process employed the Adam optimization algorithm with an initial learning rate of 0.001 and default hyperparameters β_1 = 0.9 and β_2 = 0.999. Validation was conducted during training, and the loss was monitored. The learning rate was halved if the validation loss did not improve within four epochs. To prevent overfitting, training was halted if the validation loss showed no improvement for 100 epochs.

Training and validation processes took place in the Google Colaboratory Pro Plus environment, using 52 GB of RAM and the NVIDIA Tesla P100 GPU.

The project was implemented in Python, with PyTorch and TensorFlow as the deep learning framework. (GitHub repository) [37].

3.3. Evaluation of the General U-Net Model's Prediction Accuracy

In this research, the prediction accuracy performance of the U-Net model was evaluated using several metrics, including precision, recall, and the F_1-score [38], that are required to provide a comprehensive assessment, especially when dealing with an imbalanced number of samples.

Precision assesses the proportion of correctly classified samples among those predicted as positive. Recall quantifies the proportion of correctly classified samples among all truly positive samples. The F_1-score, the harmonic mean of precision and recall, offers a comprehensive performance measure. As previously mentioned, the terms True Positive (TP) represent positive samples correctly predicted by the model, True Negative (TN) denotes negative samples correctly identified as negative by the model, and False Positive (FP) represents negative samples incorrectly identified as positive. False Negative (FN) refers to negative samples mistakenly predicted as positive. Equations (3) to (5) present the mathematical expressions for these performance measures:

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

$$Precision = \frac{TP}{TP + FP}, \tag{4}$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \tag{5}$$

3.4. Evaluation of U-Net Model Spatial Accuracy in the Testing Phase

The U-Net model is employed in a case study, resulting in a predicted mask for built-up areas, focusing on the overall built-up area and the buildings located within landslide risk zones. Subsequently, this mask is imported into the QGIS desktop

platform, where it is transformed into a vector layer and compared with the label mask from the ground truth dataset.

Assessing a deep learning model's transferability involves applying the model to geographic regions significantly distant from the areas used for training and validation. This testing of geographic transferability is valuable in eliminating spatial autocorrelation patterns between the training and testing data.

The main challenge consists of applying the model to a case study with an urban form variable passing from the central nucleus from the rural area, completely different from the urban forms used in the training and validation phases of the model.

The authors conducted a multi-phase analysis to evaluate the spatial patterns of the model's accuracy.

In the first phase, a comparison was made between the actual and predicted built-up area, considering all the buildings in the inhabited centre of the case study. The objective is to identify any differences in the accuracy of the predictions related to the density of the built-up areas. In particular, the authors intend to verify how the model behaves in the segmentation of aerial images in which the density of building footprints is predominant compared to other landscape features. For this purpose, the authors conducted a multi-ring buffering analysis in the QGIS environment by dividing the inhabited centre of San Vito sullo Ionio into concentric circular buffers with a constant radius of 350 meters, starting from the central urban nucleus with the highest built-up density and moving towards the rural areas. The urban centre area with the greatest density of public and private services was considered the central nucleus of the circular buffers. In small urban centres, a 350-meter radius will likely capture a homogenous set of buildings with similar architectural styles, materials, and heights, providing a more controlled environment for the U-Net model to operate. Indeed, as the case study considered, larger buffers might result in diminished returns for small municipalities due to the sparsity of buildings outside the central area.

Through the spatial analysis of the data, the authors calculated the actual built surface for each buffer considered. They compared this with the built surface predicted by the U-Net model, obtaining various values of the F_1-score. This score, as defined in Equation (5), is used for classification and segmentation tasks to evaluate the model's ability to classify instances correctly into different categories. In this way, it is possible to outline the trend of the model's accuracy with the variation of the urban form expressed in terms of the density of the built-up areas.

To provide insights into how the model's performance changes over different spatial scales, the authors calibrated the coefficients of a power decay function involving F_1-score and buffer radius in the multi-ring buffer analysis to evaluate the spatial variability of the U-Net model's classification and segmentation accuracy. The mathematical expression of the used decay function is the following:

$$F_1 = a \cdot DFC^b. \tag{6}$$

In Equation (6), a and b are constants that define the specific characteristics of the decay function. At the same time, DFC represents the distance from the central nucleus of the municipality, i.e., the buffer radius.

This function suggests that the buffer radius influences the F_1-score's non-linearly. The exponent b dictates the nature of this relationship:

- if $b > 0$, the F_1-score increases as the buffer radius increases, suggesting better model performance at greater distances from the centre;
- if $b < 0$, the F_1-score decreases as the buffer radius increases, indicating a decline in model performance with distance;
- if $b = 0$, the F_1-score remains constant regardless of the buffer radius, implying uniform performance across distances.

The parameter a acts as a scaling factor for the F_1-score. Its value adjusts the overall level of the scores. The authors considered the coefficient of determination R^2 to measure how well the decay model replicates the observed outcomes.

In the second phase, the authors tested the accuracy of the model, considering exclusively the buildings located in landslide risk areas to verify the capabilities of the proposed model to identify specific elements scattered over the territory. Also, the authors used the F_1-score as a metric in this case.

The described methodology evaluates the model's ability to extend its predictions to geographic areas it has yet to encounter during training.

4. RESULTS

4.1. Training and Validation Results for the U-Net Model

The U-Net model, commonly employed in semantic segmentation tasks as discussed in Section 3.2, underwent training and validation using the IoU-Based Loss.

The best-performing model reached convergence at epoch 91, achieving a Precision value of 0.63, a Recall value of 0.84, and an F_1-score of 0.72. The validation loss reached the value of 0.634. Once it was established that the model had reached a good level of accuracy in the training and validation phases, the authors applied the U-Net to the case study for building classification and segmentation.

4.2. Spatial Accuracy Results of the U-Net Model in the Testing Phase

The trained and validated U-Net model was applied to the urban area of San Vito sullo Ionio, as detailed in Section 2. The representation of the various vectorial elements within the aerial images was developed using the QGIS platform, allowing the authors to process the geospatial statistics necessary to determine the spatial patterns of the model accuracy.

In Figure 2, the predicted building mask generated by the U-Net model is depicted in orange.

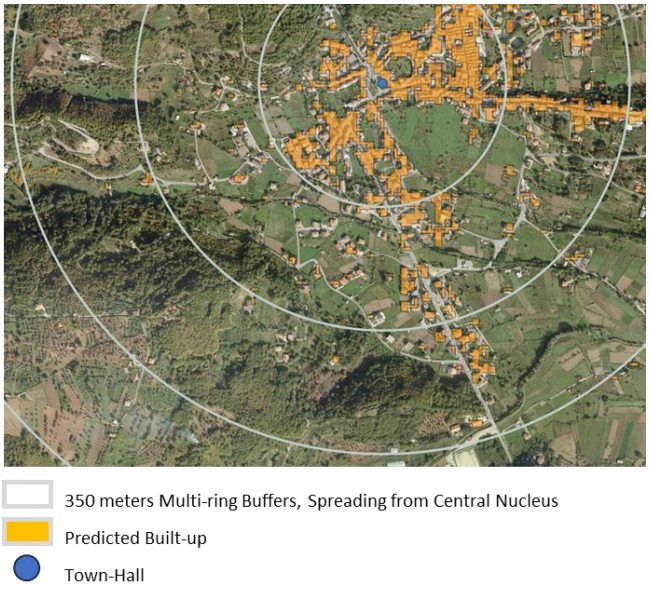

Figure 2. Predicted built-up area's mask and multi-ring buffers representation.

Table 1. The trend of F_1-score values concerning the built-up area surface and the distance from the central nucleus of San Vito sullo Ionio.

Buffer Radius (m)	Built-up Area (ha)	F_1-score
350	1.810	0.994
700	0.113	0.806
1050	0.127	0.678
1400	0.005	0.444
1750	0.015	0.642
2100	0.004	0.456

As described in the methodological section, starting from the central urban nucleus with the highest built-up density and moving toward the rural areas, the authors carried out a multi-ring buffering analysis considering circular buffers of 350 meters constant radius.

As shown in Table 1, the entire municipality of San Vito sullo Ionio is contained in a buffer with a maximum radius of 2100 meters. The central nucleus enclosed within a 350-meter radius, in which there is a built-up area of approximately 1.810 ha, is characterized by a high average F_1-score value of 0.994, which indicates an excellent level of accuracy of the U-Net model in building classification and segmentation.

Moving from the more densely built central core to the rural outskirts, the number of buildings becomes increasingly smaller, as does the built-up area, and the F_1-score average value becomes 0.806, 0.678, 0.444, 0.642 and 0.456 in the 700, 1050, 1400, 1750 and 2100-meter radius buffer, respectively. The building footprints decrease in number and density to the detriment of the other elements of the territory, passing from the central core to the periphery. In this circumstance, the model overestimates the surface area occupied by existing buildings, generating a more significant number of FPs.

This trend is confirmed considering the calibrated decay model functional form (Figure 4) reported as follows:

$$F_1 = 11.651 \cdot DFC^{-0.416} \tag{7}$$

The coefficient of determination R^2 reached the value of 0.84, which is relatively high, suggesting that approximately 84% of the variance in the F_1-score is predictable from the buffer radius (Figure 4). This implies a strong relationship between the buffer radius and the U-Net model's performance.

Considering the built-up area in landslide risk zones, Figure 3 reveals a relatively low level of landslide risk, with only 0.434 ha of built surface falling in landslide risk areas. The landslide risk zones from the official Basin Authority dataset are shown in red, and the predicted building shapes falling within landslide risk zones are highlighted in violet.

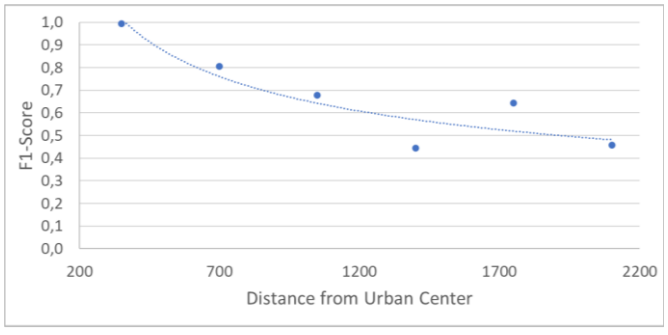

Figure 4. The functional relationship between F_1-score and the distance from the denser built-up core.

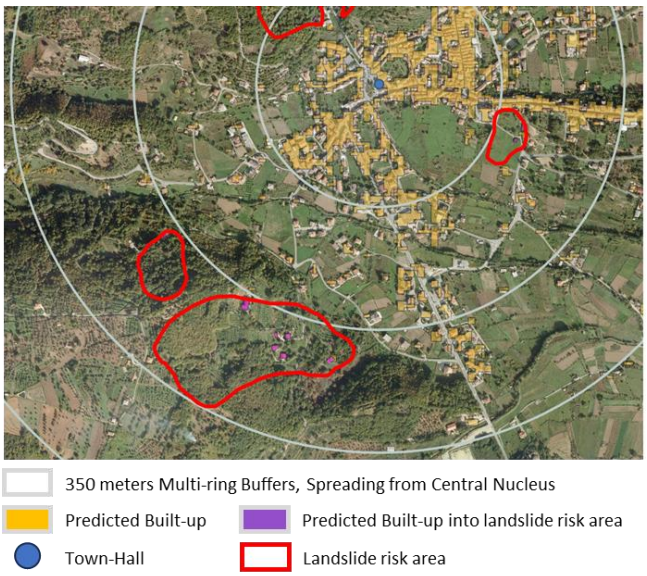

Figure 3. Predicted built-up area's mask and multi-ring buffer representation considering the landslide-risk zones.

Nonetheless, the model predicts the footprint of these buildings with a reasonably good level of accuracy, reaching an F_1-score value of 0.75. This result is perfectly consistent with the spatial accuracy analysis considering the entire built-up area. The landslide risk areas correspond with the buffers of 350, 700, and 1050 meters radius.

5. DISCUSSION OF RESULTS

A significant challenge of this research lies in applying the methodology for the classification and segmentation of built-up surfaces located in a geographical area, distinct from the training and validation dataset regarding geomorphology and territorial features. Nevertheless, the model demonstrated a commendable level of geographical adaptability considering the case study, indicating a low spatial autocorrelation pattern between the training and testing data. This adaptability is crucial for model generalization, particularly for disaster management operations, allowing for rapid, accurate estimations of buildings needing monitoring or those damaged in catastrophic events without extensive retraining.

The authors, employing a multi-ring buffer analysis and calibrating a decay function, considered the variation in segmentation accuracy of the U-Net model, measured by the F_1-score, across different buffer zones to identify spatial patterns or discrepancies in model performance.

The high value of the R^2 coefficient obtained from the decay function calibration process indicates that the model's performance varies significantly across different spatial scales (buffer radii). The results demonstrated that the urban form influences the U-Net model's ability to classify and segment built-up areas. Denser urban environments, with their distinctive, repetitive patterns and reduced complexity, are more conducive to accurate classification and segmentation by the U-Net model. Indeed, buildings and other structures often have more defined, repetitive patterns in densely built-up urban environments and are closely situated. These distinct features make it easier for the U-Net model to identify and segment the urban fabric. The compact nature of dense urban areas reduces the complexity of the image, as there is less variation in the types of objects and less

open space to differentiate from built structures. This uniformity aids the model in recognizing and delineating built-up areas.

In contrast, scattered or dispersed urban areas present a challenge for the U-Net model. The irregular distribution of buildings, interspersed with natural landscapes or other non-urban elements, creates a more complex scenario for the model to classify and segment accurately. This complexity can lead to reduced accuracy due to the difficulty of distinguishing isolated structures from their surroundings.

The results are confirmed, considering the built-up area in landslide risk zones. The reasonably good accuracy value is probably linked to the fact that the landslide risk areas are located in the areas of the municipality characterized by a higher density of built-up areas than the other areas of the case study.

The resolution of RGB aerial images also plays a critical role. In denser urban areas, even lower-resolution images provide sufficient information for effective segmentation. In contrast, higher resolution is necessary in scattered urban regions to capture the finer details of isolated structures.

For these reasons, future research will focus on expanding the dataset sample for training and validation and comparing deep-learning architectures to improve building extraction performance in complex urban settings. Enhancing the accuracy of building extraction employing metrology assessment and integrating accurate information into GIS has great utility for many urban planning tasks, which along with disaster management, include land use classification, infrastructure development, and environmental assessment.

6. CONCLUSIONS

This research introduces a novel method employing deep learning to detect built-up areas at risk in natural disasters through satellite and aerial imagery analysis. Central to this approach is the U-Net model, trained with "Semantic segmentation of aerial imagery". The authors enhanced the diversity and adequacy of training and validation data by employing various data augmentation techniques on the input images and their associated semantic segmentation masks.

The study also addressed the spatial variation assessment of the model's accuracy, which depends on the density of the built-up area. Understanding this relationship can guide efforts to optimize the U-Net model, especially if specific spatial scales have suboptimal performance.

The model demonstrated commendable accuracy and geographical adaptability in its application to different urban forms. Factors influencing this adaptability include varied urban layouts, the scope and position of landslide risk zones, diverse land use types (like residential, commercial, and industrial), geomorphological territory features, construction styles, and materials leading to different RGB spectrum values for identical object classes.

The method rapidly estimates areas needing monitoring or likely to be damaged in catastrophic events. This makes it an invaluable asset for disaster management and planning, functioning as an early warning system to pinpoint critical areas needing regular surveillance. It is also helpful for evaluating damage to buildings and infrastructure post-disaster, utilizing satellite, aerial, or UAV-derived images, especially in less accessible zones. Given its demonstrated geographical adaptability, this tool can assist decision-makers in swiftly devising effective strategies and countermeasures for disaster response. The spatial accuracy assessment in building extraction operations proposed in this paper aims to provide an accurate and up-to-date database about the built environment, which plays a crucial role in urban and disaster management planning.

REFERENCES

[1] A. Gupta, S. Watson, H. Yin, Deep learning-based aerial image segmentation with open data for disaster impact assessment, Neurocomp. 439 (2021), pp. 22-33.
DOI: 10.1016/j.neucom.2020.02.139

[2] V. Nastro, D. L. Carnì, A. Vitale, F. Lamonaca, M. Vasile, Passive and active methods for Radon pollution measurements in historical heritage buildings, Measurement, vol.114, 2018, pp. 526-533.
DOI: 10.1016/j.measurement.2016.09.002

[3] W. Li, C. He, J. Fang, J. Zheng, H. Fu, L. Yu, Semantic Segmentation-Based Building Footprint Extraction Using Very High-Resolution Satellite Images and Multi-Source GIS Data, Remote Sens. 11 (2019) 403.
DOI: 10.3390/rs11040403

[4] I. Tudosa, F. Picariello, E.a Balestrieri, L. De Vito, F. Lamonaca, Hardware Security in IoT era: the Role of Measurements and Instrumentation, Proc. of the IEEE Int. Workshop on Metrology for Industry 4.0 and IoT, Naples, Italy, 4-6 June 2019, pp. 285-290.
DOI: 10.1109/METROI4.2019.8792895

[5] P. Daponte, F. Lamonaca, F. Picariello, L. De Vito, G. Mazzilli, I. Tudosa, A Survey of Measurement Applications Based on IoT, Proc. of the IEEE Int. Workshop on Metrology for Industry 4.0 and IoT, Brescia, Italy, 16-18 April 2018, pp. 1-6.
DOI: 10.1109/METROI4.2018.8428335

[6] L. Ivanovsky, V. Khryashchev, V. Pavlov, A. Ostrovskaya, Building detection on aerial images using U-NET neural networks, Proc. of the 24th Conf. of the Open Innovations Association (FRUCT), Moscow, Russia, 8-12 April 2019.
DOI: 10.23919/FRUCT.2019.8711930

[7] C. Scuro, F. Lamonaca, R. Codispoti, D. L. Carnì, R. S. Olivito, Experimental and numerical analysis on masonry arch built with fictile tubules bricks Measurement: Journal of the International Measurement Confederation, vol.130, 2018, pp. 246-254.
DOI: 10.1016/j.measurement.2018.08.001

[8] F. Lamonaca, A. Carrozzini, D. Grimaldi, R.S. Olivito, Improved Monitoring of Acoustic Emissions in Concrete Structures by Multi-Triggering and Adaptive Acquisition Time Interval, Measurement, vol.59, 2015, pp. 227-236.
DOI: 10.1016/j.measurement.2014.09.053

[9] F. Lamonaca, P. F. Sciammarella, C. Scuro, D. L. Carnì, R. S. Olivito, Synchronization of IoT Layers for Structural Health Monitoring, Proc. of the IEEE Workshop on Metrology for Industry 4.0 and IoT, Brescia, Italy, 16-18 April 2018, pp. 89-94.
DOI: 10.1109/METROI4.2018.8428329

[10] G. Cheng, J. Han, A survey on object detection in optical remote sensing images, ISPRS J. Photogramm and Rem. Sens. 117 (2016), pp. 11-28.
DOI: 10.1016/j.isprsjprs.2016.03.014.

[11] Z. Ziaei, B. Pradhan, S.B. Mansor, A rule-based parameter aided with object-based classification approach for extraction of building and roads from WorldView-2 images, Geocarto Int. 29 (2014), pp. 554-569.
DOI: 10.1080/10106049.2013.819039

[12] A. O. Ok, Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts, ISPRS J. Photogramm and Rem. Sens. 86 (2013), pp. 21-40.
DOI: 10.1016/j.isprsjprs.2013.09.004

[13] M. Belgiu, L. Drăguţ, Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery, ISPRS J. Photogramm and Rem. Sens. 96 (2014), pp. 67-75.
DOI: 10.1016/j.isprsjprs.2014.07.002

[14] R. Chen, X. Li, J. Li, Object-Based Features for House Detection from RGB High-Resolution Images, Remote Sens. 10 (2018) 451. DOI: 10.3390/rs10030451
[15] I. Ahmed, E. Balestrieri, P. Daponte, F. Lamonaca, A Method Based on Ellipse Fitting for Automatic Morphometric Parameter Measurements of Fish Blood Cells, proc. of IEEE Int. Symp. on Medical Measurements and Applications, Jeju, Korea, 14-16 June 2023, pp.1-6. DOI: 10.1109/MeMeA57477.2023.10171881
[16] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Boston, MA, USA, 7-12 June 2015. DOI: 10.1109/CVPR.2015.7298965
[17] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, B.A. Johnson, Deep learning in remote sensing applications: A meta-analysis and review, ISPRS J. Photogramm and Rem. Sens. 152 (2019), pp. 166-177. DOI: 10.1016/j.isprsjprs.2019.04.015
[18] H.C. Shih, D.A. Stow, Y.H. Tsai, Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping, Int. J. of Rem. Sens. 40 (2019), pp. 1248-1274 DOI: 10.1080/01431161.2018.1524179
[19] M. Carranza-García, J. García-Gutiérrez, J.C. Riquelme, A framework for evaluating land use and land cover classification using convolutional neural networks, Rem. Sens. 11 (2019) 274. DOI: 10.3390/rs11030274
[20] M. A. Shafaey, M. A. M. Salem, H. M. Ebied, M. N. Al-Berry, M. F. Tolba, Deep learning for satellite image classification, Proc. of the Int. Conf. on Advanced Intelligent Systems and Informatics, Cairo, Egypt, 1-3 September 2018.
[21] S. Mohajerani, P. Saeedi, Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery, Proc. of the IEEE Int. Geoscience and Remote Sensing Symp. IGARSS, Yokohama, Japan, 28 July - 2 August 2019, pp. 1029-1032.
[22] H. Ye, S. Liu, K. Jin, H. Cheng, CT-UNet: An Improved Neural Network Based on U- Net for Building Segmentation in Remote Sensing Images, Proc. of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10-15 January 2021, pp. 166-172. DOI: 10.1109/ICPR48806.2021.9412355
[23] N. He, L. Fang, A. Plaza, Hybrid first and second order attention Unet for building segmentation in remote sensing images, Sci. China Inf. Sci. 63 (2020) 140305. DOI: 10.1007/s11432-019-2791-7
[24] Y. Hou, Z. Liu, T. Zhang, Y. Li, C-Unet: Complement UNet for remote sensing road extraction, Sensors, 21 (2021) 2153. DOI: 10.3390/s21062153
[25] M. Francini, C. Salvo, A. Vitale, Combining Deep Learning and Multi-Source GIS Methods to Analyze Urban and Greening Changes, Sensors 23 (2023) 3805. DOI: 10.3390/s23083805
[26] M. Francini, C. Salvo, A. Viscomi, A. Vitale, A Deep Learning-Based Method for the Semi-Automatic Identification of Built-Up Areas within Risk Zones Using Aerial Imagery and Multi-Source GIS Data: An Application for Landslide Risk, Remote Sens. 14 (2022) 4279. DOI: 10.3390/rs14174279
[27] H. Miao, Z. Zhao, C. Su, C. Li, R. Yan, A U-Net-Based Approach for Tool Wear Area Detection and Identification, IEEE Trans. on Instrum. and Measur. 70 (2020), pp. 1-10. DOI: 10.1109/TIM.2020.3033457
[28] W. Li, C. He, J. Fang, J. Zheng, H. Fu, L. Yu, Semantic Segmentation-Based Building Footprint Extraction Using Very High-Resolution Satellite Images and Multi-Source GIS Data, Remote Sens. 11 (2019) 403. DOI: 10.3390/rs11040403.
[29] W. Alsabhan, T. Alotaiby, Automatic Building Extraction on Satellite Images Using Unet and ResNet50, Comput. Intell. Neurosci. 2022. DOI: 10.1155/2022/5008854
[30] Humans in the Loop, Semantic segmentation dataset. Online [Accessed 8 December 2023] https://humansintheloop.org/semantic-segmentation-dataset/
[31] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, F. Shen, Image Data Augmentation for Deep Learning: A Survey, Proc. of the Computer Vision and Pattern Recognition Conference (CVPR), New Orleans, Louisiana, 19-24 June 2022.
[32] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, Proc. of the 3rd International Conference for Learning Representations (ICLR), San Diego, CA, USA, 7-9 May 2015. DOI: 10.48550/arXiv.1412.6980
[33] Geoportale della Regione Calabria. Online [Accessed in March 2023] http://geoportale.regione.calabria.it/
[34] Autorità di Bacino della Regione Calabria. Online [Accessed in March 2023] https://www.regione.calabria.it/website/ugsp/autoritadibacino/
[35] I. Ahmed, E. Balestrieri, I. Tudosa, F. Lamonaca, Segmentation techniques for morphometric measurements of blood cells: Overview and research challenges, Measurement: Sensors, 2022, pp.1-22. DOI: 10.1016/j.measen.2022.100430
[36] D. L. Carnì, F. Lamonaca, Toward an Automatic Power Quality Measurement System: An Effective Classifier of Power Signal Alterations, IEEE Transactions on Instrumentation and Measurement, vol.71, 2022, pp.1-8. DOI: 10.1109/TIM.2022.3192258
[37] Alessandro Vitale on GitHub, A U-Net model for the automatic Segmentation of aerial images using Dubai's satellite imagery dataset. Onbline [Accessed in December 2023] https://github.com/ingegnerevitale/U-Net-Semantic-Segmentation-Aerial-Images
[38] N. Chinchor, B. M. Sundheim, MUC-5 evaluation metrics, Proc. of the Fifth Message Understanding Conference (MUC-5), Baltimore, MD, USA, 25-27 August 1993.