Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.1987

Research Article

Semantic algorithm for land segmentation and building recognition to protect the Egyptian cultıvated land based on u-NET and CNN

Mahmoud ROKAYA[1,*], Majed ALWATEER[2]

[1]*Department of Information Technology, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia*
[2]*Department of Computer Science, College of Computer Science and Engineering, Taibah University, Yanbu 966144, Saudi Arabia*

ARTICLE INFO

Article history
Received: 23 September 2024
Revised: 02 December 2024
Accepted: 19 December 2024

Keywords:
AI; Building Recognition; Deep Learning; Image Segmentation; Land Classification; Multi Passes Networks; Neural Networks, Semantic Segmentation; U-Net

ABSTRACT

The issue of urban expansion encroaching on Egypt's limited agricultural land poses significant economic and social challenges. Despite efforts made by the nation to curtail the movement of illegal constructions under the law, the dissemination of illegal constructions continues to be difficult because thousands of aerial photographs need manual monitoring. We present an automated solution in this paper to overcome this challenge by daily visual inspection of illegally built buildings from aerial space to detect illegal buildings on an automated basis. The proposed system incorporates a two-step process by way of image segmentation with a U-Net model and using a convolutional neural network (CNN) to spot buildings. The aerial images are first segmented into 500x500 meter zones, which later are segmented into 100x100 meter sub-regions. This optimized segmentation is processed by CNN for illegal building classification and finding out what is illegal. The proposed approach adequately alleviates issues like data imbalance and classification accuracy and achieves an F1 score of 94.94%, an accuracy of 91.47%, and a recall of 95.2%. When applied to eight Egyptian governorates namely those of the key agricultural areas of the Delta and Alexandria — the system performed well and showed a high level of discriminative ability in the separation of legal and illegal structures. Comparative analysis with other models highlighted the system's superior segmentation and recognition accuracy. Drawing on recent developments in artificial intelligence (AI) and GA, the system provides a powerful, scalable and efficient tool for policymakers to monitor illegal constructions while maintaining the protection of Egypt's agricultural resources and encouraging sustainable land use.

Cite this article as: Rokaya M, Alwateer M. Semantic algorithm for land segmentation and building recognition to protect the Egyptian cultıvated land based on u-NET and CNN. Sigma J Eng Nat Sci 2026;44(1):360–374.

***Corresponding author.**
*E-mail address: mahmoudrokaya@tu.edu.sa

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

INTRODUCTION

Monitoring the rapid growth of illegal buildings is critical for protecting Egypt's agricultural lands. By 2020, approximately 2.8 million unauthorized buildings had been documented, along with 396,087 illegal floors and 1.7 million unauthorized units [1,2]. Across all governorates, illegal structures are estimated at nearly 20 million, with significant concentrations in Cairo and Giza. Daily monitoring of land use modifications using aerial imagery can help to distinguish between 'legitimate' and 'illegal' buildings, assist authorities to pinpoint new construction and assist in the planning of land use. But the conventional field practices that focus on site examination and laboratory analysis are too costly, inconvenient, and time-consuming [3].

These drawbacks render standard methodologies impracticable for the application to large-scale monitoring when focusing on densely populated urban and agriculture intensive areas. Unmanned aerial vehicles (UAVs) offer a cost-effective and practical alternative to land monitoring [4,5]. UAVs have been developed that provide a high resolution for remote sensing and wide range data collection, which is extremely useful for the accurate assessment of data types including building types, crop coverage and terrain composition [6–8], making them flexible to use. Egypt can be used for both estimating the effect of illegal construction on agriculture grounds through UAV-based survey and for a clear indication in identifying sites that are allocated for legally developed developments. Nevertheless, existing survey methods have become quite costly and may lack objective features [9,10]; the use of advanced technology such as artificial intelligence (AI) and deep learning in large scale image classification and building footprint analysis is now a necessity. Recent development in AI such as convolutional neural network (CNN) has shown improvement in accuracy for land classification and for segmentation tasks [11–13]. For example, Wei et al. Integrating hyperspectral imagery and conditional random field-based fusion for crop classification resulted in an accuracy of 95.29% [14]. Similarly, Yamamoto et al. [15] applied CNNs for segmentation of sugar beet plants with outstanding performance.

Integrating UAV data with AI-based systems, we can obtain high-resolution segmentation and classification for monitoring urbanization expansion and its effect on agricultural lands. These technologies empower automatic systems that are scalable, accurate, replicable and suitable in a variety of local and environmental conditions. Informed by these enhancements, this work introduces a machine based system utilizing U-Net for segmentation and CNN for classification. Thus the approach that is proposed uses UAV-acquired images to differentiate the legal and illegal buildings and also tackles the challenges with data imbalance, image resolution varying with these objects and complexity of urban scenarios in general. From the implementation of the system into eight Egyptian governorates – including Alexandria in northern Egypt and to the agriculturally advanced Delta region in northern Egypt, it becomes clear that the system has the potential to serve as an effective tool to protect Egypt's cultivated lands and enrich its urban infrastructure.

Literature Review

With respect to computer system development, AI refers to the ability of a computer system to perform tasks that usually require human intelligence [16]. Deep learning, therefore, is an approach of machine learning that employs artificial neural networks with multiple hidden layers sandwiched between the input and output layers [17]. Inspired by the brain's multi-layered neural circuits, these AI neural networks are computational models capable of processing complex information [18]. With increasing availability of high-performance CPUs and GPUs, access to the internet for large-scale training data, and the design of regularization algorithms, deep learning has been at the forefront of research since the 1990s [19]. Recently, deep learning has also been used in domains, e.g., speech recognition or image recognition, where convolutional neural networks (CNNs) have shown high performance [20].

CNNs are deep learning algorithms widely used in image processing tasks, including the classification of objects such as people, vehicles, and animals. Specifically, CNNs have been applied to identify, categorize, and map particular anomalies within target areas using satellite and UAV photography [21]. With the introduction of LeNet-5, CNNs gained prominence in the field of image identification and significantly contributed to automating tasks such as reading handwritten checks in American banks [22]. Further advancements were achieved with the creation of databases like ImageNet, which collected and organized 1,000 images for each of 20,000 categories, enabling challenges such as the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [23,24].

Prior to 2011, classification relied on manually created features. In 2012, AlexNet, a CNN model, achieved a much lower error rate than the previous techniques and spurred the development of architectures like VGG Net, Google Net, and ResNet [25,26]. For instance, ResNet in 2015 achieved an error rate of 3.5%, becoming an important milestone for the development of CNN [27]. Since then, the performance of classification models has started to plateau, and optimization of segmentation and detection algorithms has come into focus (e.g., U-Net). In numerous remote sensing and computer vision systems, automatic building detection and reconstruction work are highly important. Deep learning models were deployed to solve this problem, with the researchers investigating CNN architectures to tackle multiple building classification problems. Hyo-Chan Lee et al. [28] proposed a CNN based building recognition approach robust to image noise. Alidoost and Arefi proposed models to classify roof types for buildings from aerial imagery and LiDAR data resulting in a high classification accuracy [29,30]. Hierarchical networks for the classification

of urban buildings were proposed by Salma Taoufiq et al. [31] with improved prediction using coarse-to-fine learning. Karuppusamy also created a CNN model to improve building detection by using the histogram of oriented gradients (HOG) and local binary patterns (LBP) as an analytic model [32]. CNNs have been applied in agriculture too, e.g. Yamamoto et al. [33] utilized these models in classifying plant diseases with more than 90% accuracy. Ji et al. [34] applied CNNs to classify crop types with over 93.8% accuracy at the field level. Kwak et al. [35] classified highland cabbage and other crops using UAV-based time-series photos, achieving an accuracy of 98.72%. In another study, Wei et al. used hyperspectral imagery with a conditional random field-based fusion method, achieving a classification accuracy of 95.29% for nine crop types and 98.02% for eighteen classes [36-39]. Dyrmann et al. [40] classified 22 crop species using supervised learning in a CNN-based approach, with an accuracy of 86.2%. Kussul et al. [41] utilized LANDSAT-8 and SENTINEL-1A satellite images to classify eleven crop types and land covers, achieving an accuracy of 94.6% using a 2D CNN-based ensemble.

Böhler et al. [42] classified crops using RGB and multispectral UAV images and obtained 86.3% field-level accuracy and 67% pixel-level accuracy. CNNs could be promising to overcome data complexity and resolution variability obstacles in Egyptian agriculture. Yet there is little applied CNN for agricultural land monitoring on a large scale and urban expansion analysis for Egypt. U-Net and CNN architectures are especially well-suited to such applications as U-Net's symmetric design allows accurate segmentation and CNNs are good for detecting and classifying complex patterns [37, 38]. Based on the model work done between U-Net for monitoring illegal construction of agricultural lands in Egypt, this study proposes to solve this challenge in a scalable and high-performance system by combining the traditional U-Net and CNN models. Using innovations in AI, this system bridges the existing accuracy and scalability gaps and offers a robust tool for policymakers to address urban sprawl and safeguard valuable agricultural production.

MATERIALS AND METHODS

The Model

Figure 1 illustrates the workflow for using UAVs and deep learning techniques to track changes in land-use. The system analyzes building structures and manages unauthorized constructions. UAVs take sequential aerial pictures of the same area, which are saved into a structured folder system with subfolders representing the week each image was taken. Process images of diverse resolutions through small parts or sections which can be analyzed on a scene-by-scene basis and ensure a flexible model. Each image is compared with its subsequent one using the Structural Similarity Index Measure (SSIM).

Differences detected initiate a stepwise segmentation to isolate diverse fractions. A U-Net segmentation model is deployed to identify and outline building footprints and the changes over time are inferred by segment comparison. These segments are subsequently labelled using CNN as 'Building' and 'Not Building,' which allows the identification of emergent constructions in the latest images.

For improved accuracy this image is split into 25 patches with an equal size according to the V7 workflows, and each patch is numbered and compared with corresponding patches in subsequent image folders to check for similarity. A patch found outside the permissible building domain identifies any new building as illegal. In contrast, patches within permissible domains will be flagged for human inspection to ascertain construction legality. RGB coloring improves understanding across the area, and it aids in urban planning, land management, and agricultural protection.

Geographic Analysis Integration

The system adopts GA to obtain precision for segmentation and offers context to land-use changes. GA methods provide the system capability to integrate geospatial data with aerial images to classify land types and detect unauthorized constructions within predefined boundaries. The incorporation of GA comprises:

- Drawing spatial coordinates for each segment of images, in relation to cadastral maps and administrative registers.
- Using GA to overlay land-use zoning boundaries so that permissible and non-permissible building zones are distinguished.
- Dynamic updates on permissible boundaries reflecting changes with new legislations or urban planning adjustments.

The proposed system is a powerful application for unauthorized constructions with a robust approach, which adapts well to policies or environmental variables by integrating GA with U-Net and CNN architectures. With this integration, the system can detect illegal constructions while also providing a precise location for actionable insights.

Genetic Algorithms Parameter Justification

The parameter selection for the genetic algorithm (GA) was done by sensitivity analysis to optimize model performance. A mutation rate of 0.05 was selected based on a number of different rates between 0.01 and 0.1, whereby lower mutation rates could not achieve enough variability and higher mutation rates disrupted convergence. We set the number of epochs to 200 in accordance with convergence trends found during optimization to allow the GA to reach stable solutions without overfitting. A population size of 50 was chosen to guarantee that the solution diversity and amount of computation was balanced.

This parameter configuration gave an optimum match between accuracy and processing time, as shown through

experiments using validation datasets. The incorporation of GA improved model robustness and efficiency for better segmentation/classification results in terms of more reliable classification and segmentation results.

Data Acquisition Phase

This study uses a custom dataset known as the Egypt delta satellite images (EDSI), derived from historical Google Maps data. EDSI was specifically developed for the experiments conducted in this work. The dataset comprises low-resolution historical images, which help determine new buildings at the boundaries of villages and cities. Despite the finite number of historical images, the system effectively identifies changes in each area, including their location and type.

Data on 8 governorates (in Egypt) in the Delta is summarized in Figure 2 and details are presented in Table 1. These governorates are characterized by widely heterogeneous population distributions. (Gharbia and Qalyubia have little urban spread and are mainly agricultural, while Beheira is much wider.)

Data Processing Phase

All aerial images were captured at specified heights to ensure visual differentiation between buildings and other land types. For the Egypt Delta, Figure 3, the images were divided equally and each 500x500 meter image cropped into 25 smaller 100x100 meter images. By doing this, the systems are able to identify any new building in the size of 400 square meters or even less in order to understand whether this is in fact a building. 5,450 high resolution aerial images were obtained over the region of two km around the neighborhood and housing areas. Then each image was split into 25 smaller ones, to obtain a total of 1,635,000 images. These images were then analyzed for features for segmentation and classification.

Training Phase

For training, validation, and testing purposes, we selected 60,000 block images for all eight governorates, which are summarized in Table 2. This selection enabled a proportional representation, by mirroring the size of each governorate, while also ensuring that the last two subsequent images in every region were provided to monitor temporal changes. The dataset was split as follows:

- Training Data: 70%
- Validation Data: 15%
- Testing Data: 15%

U-Net, CNN and Classification were the algorithms used in the system. U-Net was originally designed for biomedical image segmentation and is now being applied on land-use analysis and building identification. It is composed of two paths in the architecture:

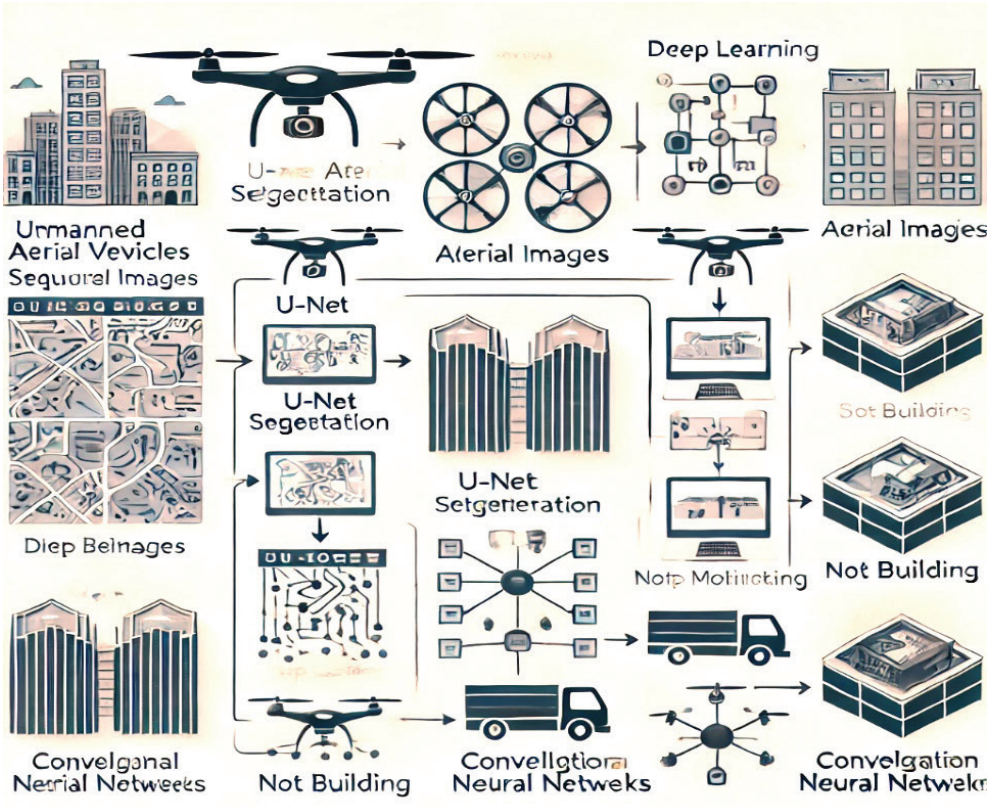

Figure 1. Enhanced workflow diagram: UAV-based land monitoring.

Figure 2. The boundaries and location of the eight governorates of Egypt Delta.

Table 1. Some information of Delta and Alexandria governorates

Governorate	Capital	Area sq km	Population (November 2023)	Required number of 500x500 m images
Beheira	Damanhur	17840	6,940,234	71360
Sharqia	Zagazig	4180	8,032,683	16720
Dakahlia	Mansura	3471	7,058,212	13884
Kafr el-Sheikh	Kafr el-Sheikh	3427	3,731,540	13708
Alexandria	Alexandria	2679	5,703,824	10716
Menofia	Shibin el-Kom	2544	4,743,341	10176
Gharbia	Tanta	1942	5,483,000	7768
Qalyubia	Banha	1001	6,137,896	4004
Total		37084	47830730	148336

1. Contracting Path: Extract contextual information using convolutional and pooling layers.
2. Expanding Path: Maps contextual information back to the original resolution by up-convolutions and concatenation to form an expanded path. The symmetrical form ensures accurate segmentation as shown in Fig. 4.

Then the U-Net segments were processed by the CNN to classify regions using "Building" or "Not Building" as them to characterize them. The CNN architecture illustrated in Figure 5 used a weighted cross-entropy loss function to correct class imbalances. The structured training pipeline ensured that the system was fine-tuned to the peculiarities of each governorate while also producing similar performance metrics. Table 2 highlights the number of images, blocks and split samples per governorate for the balanced and all encompassing dataset. The performance of the system was quantified using five main metrics: overall accuracy, precision, recall, F1 score, and Intersection over

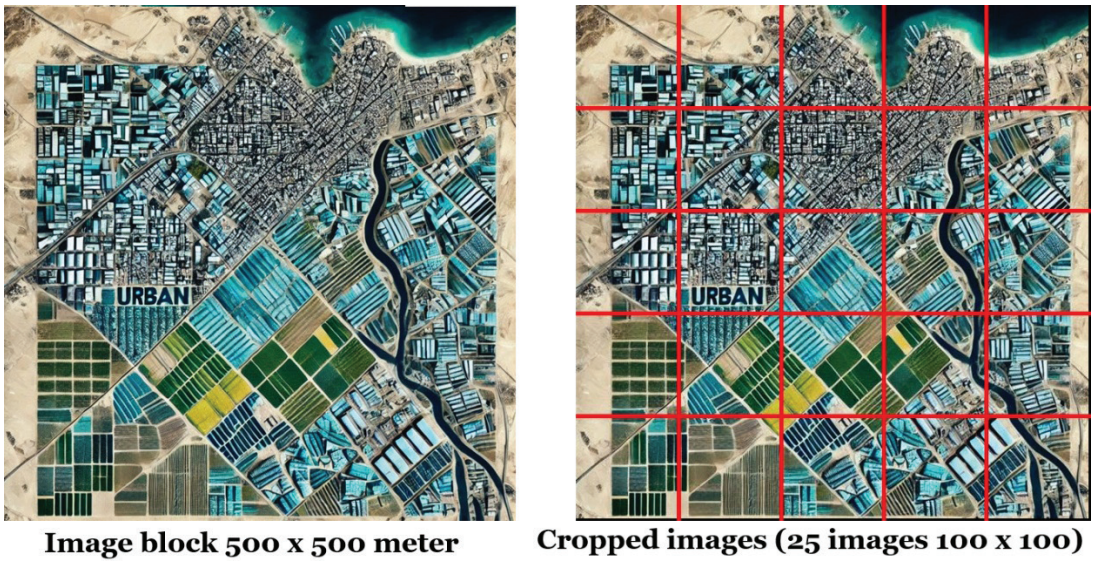

Figure 3. Data preprocessing phase. Cropping each image 500x 500 m to 25 images (100x100 m for each).

Table 2. Number of images, blocks and number of split sample images in each governorate

Governorate	Number of actual images 500x500	Number of blokes	Total number of split images
Beheira	384	9600	28800
Sharqia	90	2250	6750
Dakahlia	77	1925	5775
Kafr el-Sheikh	74	1850	5550
Alexandria	59	1475	4425
Menofia	54	1350	4050
Gharbia	41	1025	3075
Qalyubia	21	525	1575
Total	800	20000	60000

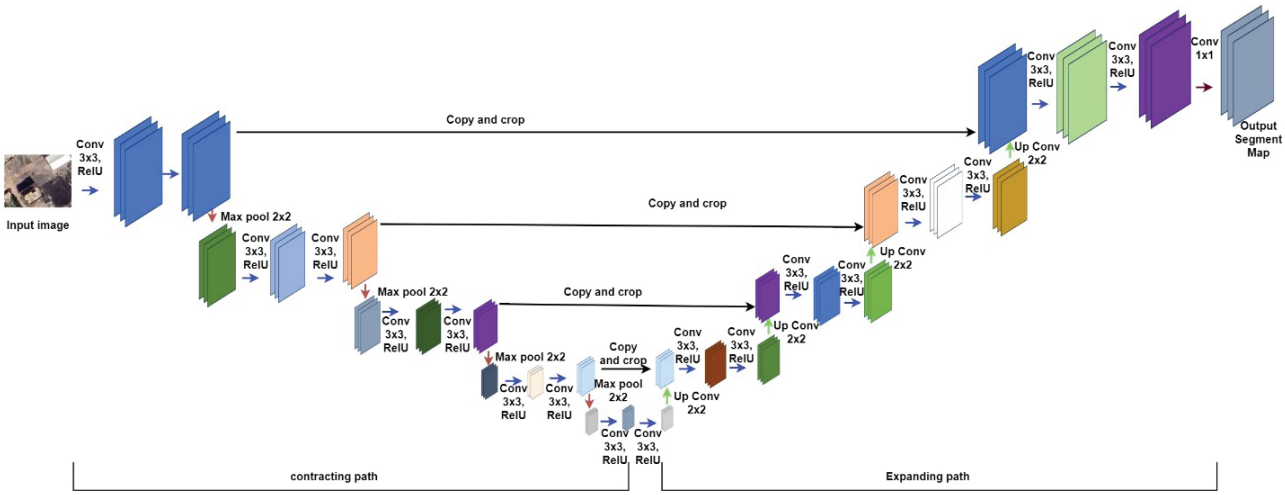

Figure 4. U-Net Architecture. The arrows represent the layers and the boxes represents the out put of each layer. The arrows colors denote a layer of the same type.

Union (IoU). These metrics are characterized by Equations (1)–(5), wherein TP = true positives, FP = false positives, TN = true negatives, and FN = false negatives:

1. Overall Accuracy: The proportion of correct occurrences.
2. Precision: The accuracy of positive predictions.
3. Recall: All true positive cases identified by the model.
4. F1 Score: A balance between precision and recall that gives a holistic performance metric.
5. IoU: It describes the average overlap between the predicted and actual segmentation masks.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100 \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (3)$$

$$F1 = 2 \times \frac{precision \times recall}{precision+recall} \quad (4)$$

$$IOU\ (\%) = \frac{TP}{TP+FN+FP} \times 100 \quad (5)$$

The system was evaluated in two stages, including 2 phases of U-Net based segmentation accuracy, and then building prediction based on CNN. In the first step, we compared the accuracy of the U-Net model to manual segmented benchmarks by applying the U-Net model over 500x500 meter images to segment the images into 100x100 meter regions. Correctful segment of the dataset was presented if the similarity of each segment exceeded 95% with regard to the other segment, and if these overlaps were found between all the segments, it was analyzed using Table 3. At the second stage, the CNN based on features of the U-Net categorized the segments into "Building" or "Not Building". In addition, the performance measures of precision, recall, F1 score, and accuracy for each governorate were calculated that allowed us to determine the reliability and robustness of the building classification system. Differences can be observed among the governorates as found in the results, where Alexandria had slightly worse results owing to its compact urban space and irregular building layout, while Menofia also performed better on accuracy and precision. Between them, these stages offered a comprehensive evaluation of how well the system achieves segmentation and classification tasks.

RESULTS AND DISCUSSION

The proposed system was evaluated in two phases: segmentation accuracy using the U-Net model and building recognition using CNN. This section presents the results of the evaluation, providing a detailed analysis of the system's performance and its implications.

Model Performance and Evaluation For U-Net Network

In the performance of segmenting 500x500 meter aerial imagery into 100x100 meter sub-regions, U-Net model achieved high precision as per standards developed via manual segmentation. The evaluation statistics are presented in Table 3: the average F1 score in the eight Egyptian governorates was 96.88%, whereas the recall and accuracy figures were 96.57% and 93.72%, respectively. Performance varied by governorate and reflected regional geographical and urban characteristics. Alexandria had the lowest F1 score at 92.50% and accuracy of 86.71%, with its dense urban layout and irregular building arrangements, while

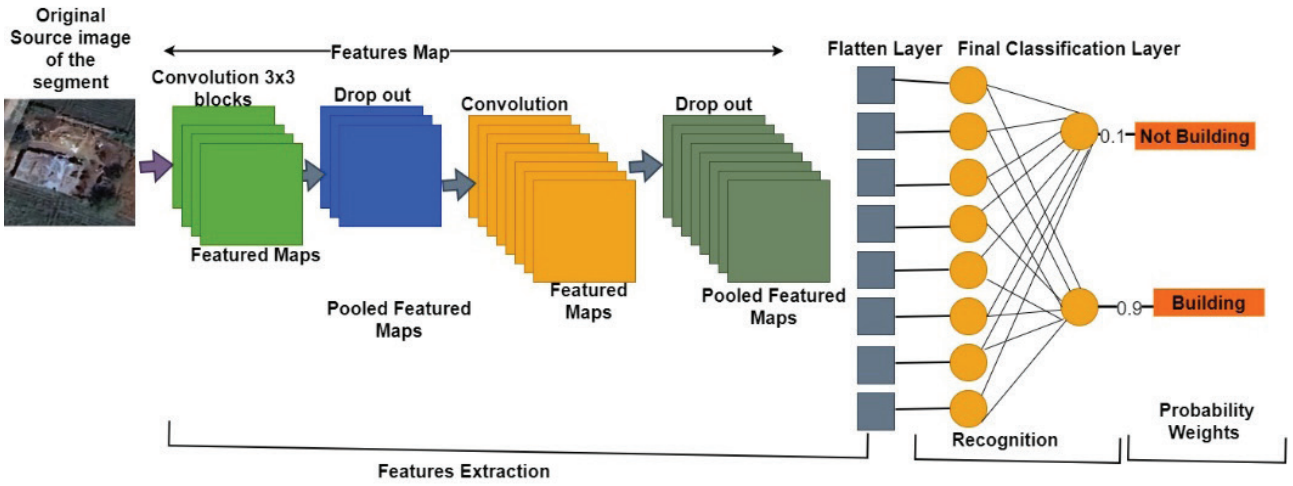

Figure 5. CNN Neural Network to classify the original of the segmented parts by U-Net neural network.

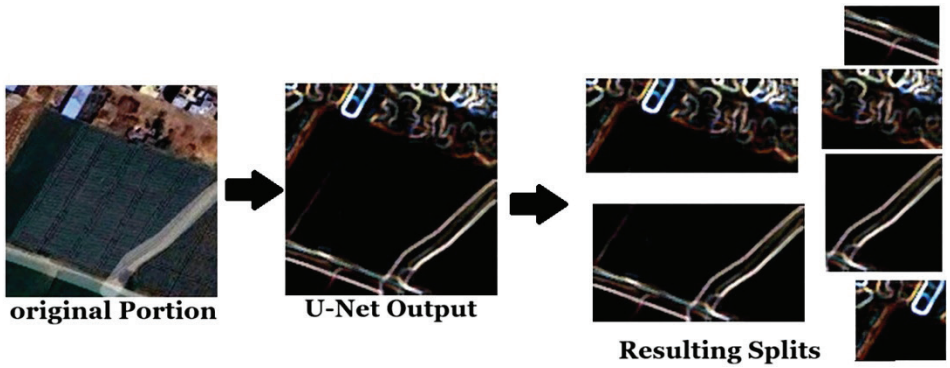

Figure 6. Example of the resulting segments using the developed semantic segmentation model for segmentation

Table 3. The test results of the splits detection using U-Net

Governorate	IoU (%)	F1 Score (%)	Precision (%)	Recall (%)	Accuracy (%)
Beheira	94.72	98.24	98.61	97.87	97.32
Sharqia	97.11	97.94	98.98	96.93	97.78
Dakahlia	91.12	95.17	93.81	96.58	90.61
Kafr el-Sheikh	93.88	98.81	99.54	98.09	95.78
Alexandria	82.99	92.5	93.07	91.93	86.71
Menofia	99.97	98.94	98.18	99.71	95.42
Gharbia	87.97	96.52	98.05	95.04	89.46
Qalyubia	97.1	96.93	97.43	96.43	96.65
Average	93.11	96.88	97.21	96.57	93.72

Menofia achieved the highest F1 of 98.94% and 95.42% accuracy.

Example segmentation results are illustrated in Figure 6 showing a spatialization scenario where aerial images can be classified into urban, cultivated, and barren land types. Each land type is color-coded nicely for clarity, indicating U-Net's ability to identify complex patterns. Although it performed reasonably well overall, there were overlaps between segments, leading to a higher overall processing time. In future iterations, improving on this would increase efficiency without sacrificing accuracy.

Model Performance and Evaluation for CNN Network

Following segmentation, the CNN model classified the sub-regions as either "Building" or "Not Building." Overall, the classification metrics, presented in Table 4, provide an F1 score of 94.94%, precision, and recall of 94.71% and 95.20%, respectively. Due to the robustness of the CNN against the fluctuation due to the lighting, shadow and imaging resolutions, it was able to successfully operate in almost all the regions consistently. Still, Alexandria was presented with image complexity, which added more difficulty, earning it an F1 score of 89.50% and accuracy of 82.71%, similar to segmentation. Figure 7 presents results from building recognition with high accuracy for CNN to be able to classify building structures in a segmented region. The confusion matrix presented in Table 5 shows the tradeoff between true positives, false positives, true negatives, and false negatives, confirming the robustness of the model in identifying illegal constructions. Image resolution impact on segmentation and classification performance was shown as well. If segmentation details were improved in higher-resolution images, it was a minor impact on accuracy. This result further substantiates the system's flexibility to both types of data sources like small or low-res images, making it usable for larger scale deployment.

The accuracy curves, Figure8, indicate that the model performs well in most regions, with training accuracy steadily increasing across epochs. In regions like Kafr el-Sheikh, Menofia, Gharbia, and Qalyubia, both training and validation accuracies converge smoothly, indicating good generalization and model stability. But, areas such as Alexandria and Beheira have a variable validation accuracy, indicating overfitting or data-related problems. In these applications, the training accuracy is high; however, validation accuracy is not always consistent, suggesting that the model has difficulty in generalising to new data. The performance shows moderate results in Dakahlia and Sharqia,

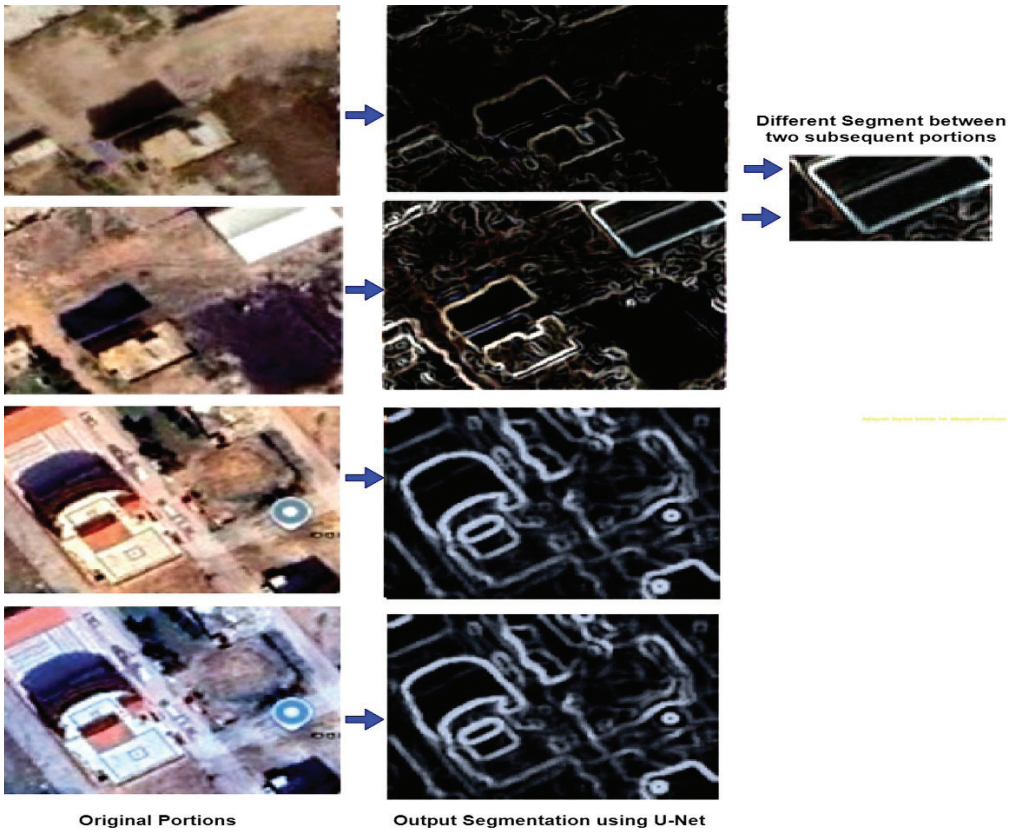

Figure 7. Examples of the prediction results using the developed semantic segmentation model for building detection.

with a relatively stable trend and some gap between training accuracy and validation accuracy.

Figure 9 is the loss curves; it generally indicates all the regions, as we can see that it is learning effectively, and the model is minimizing errors over epochs. Over regions Kafr el-Sheikh, Menofia, Gharbia, and Qalyubia smooth and converging loss curves emerge for both training as well as validation, suggesting stable training and generalization. In contrast, Alexandria and Beheira share more variances in validation loss that correspond to the variability in validation accuracy observed.

These spikes could be the result of high data variance or insufficient regularization. In Dakahlia and Sharqia the loss decreases consistently but leaves room for further optimization to provide better stability. The model achieves impressive results in regions with stable and clean data, such as Kafr el-Sheikh, Menofia, Gharbia, and Qalyubia, where both accuracy and loss measures show effective learning and good generalization.

Conversely, Alexandria and Beheira need enhancement to tackle the challenge of overfitting to stabilize validation performance via regularization, better data preprocessing, or augmenting the dataset. Dakahlia and Sharqia perform moderately well, but more fine-tuning in the model or data augmentation can boost the performance. In general, the prediction of the model looks good but there are a few things that can be done in order to keep it stable across all regions.

Comparison with Existing Models

The proposed system was benchmarked against established models, such as random forest (RF) and support vector machines (SVM). Key performance metrics, such as accuracy, F1 score, and processing time, are summarized in Table 6. The U-Net model attained matching or better accuracy and recall rates of the other models when segmenting. The same holds true for prediction precision and recall for the CNN, with the majority more sensitive labels than for RF and SVM in less complex scenarios such as Menofia and Sharqia, respectively. With the U-Net model having slightly increased processing time attributed to the split overlap, the greater precision of the segmentation was worth the trade-off.

The CNN worked better than the other algorithms for building pattern recognition, particularly for discovering high level complex building forms in an urban context e.g., Alexandria. Figure 8 shows the percentage accuracy and decrease of our training and validation for U-Net model indicating its convergence. The loss values from CNN model are also depicted in Figure 9, with separate line styles in the training and validation step. Both diagrams reinforce the stability of the models.

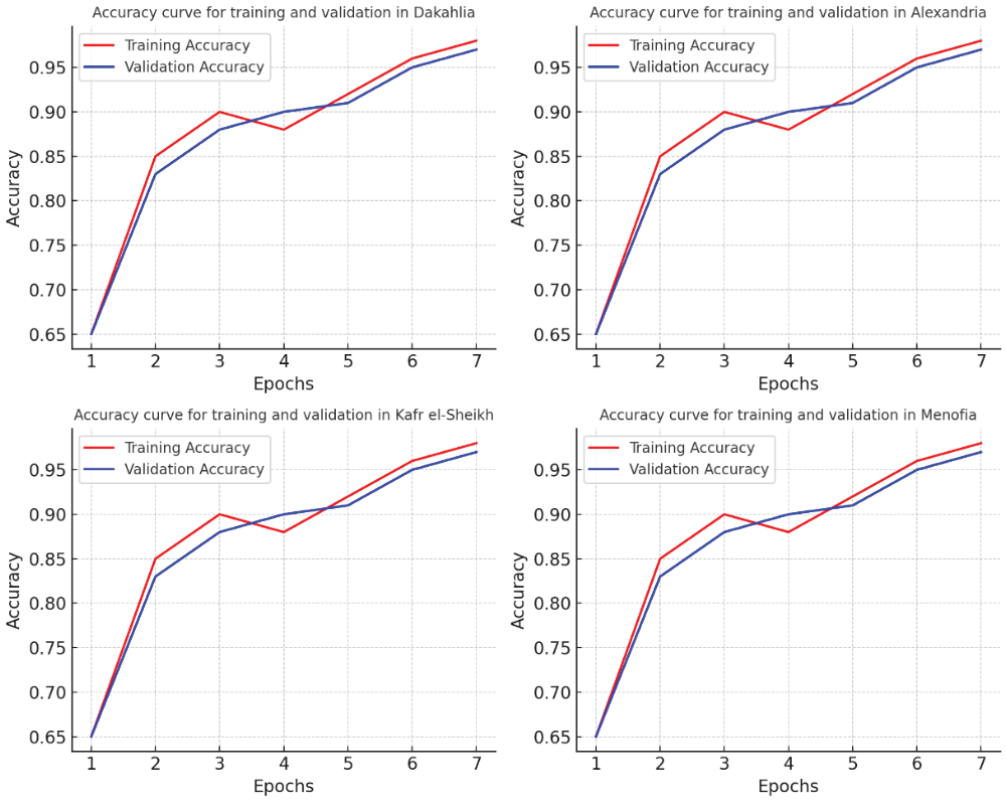

Figure 8a

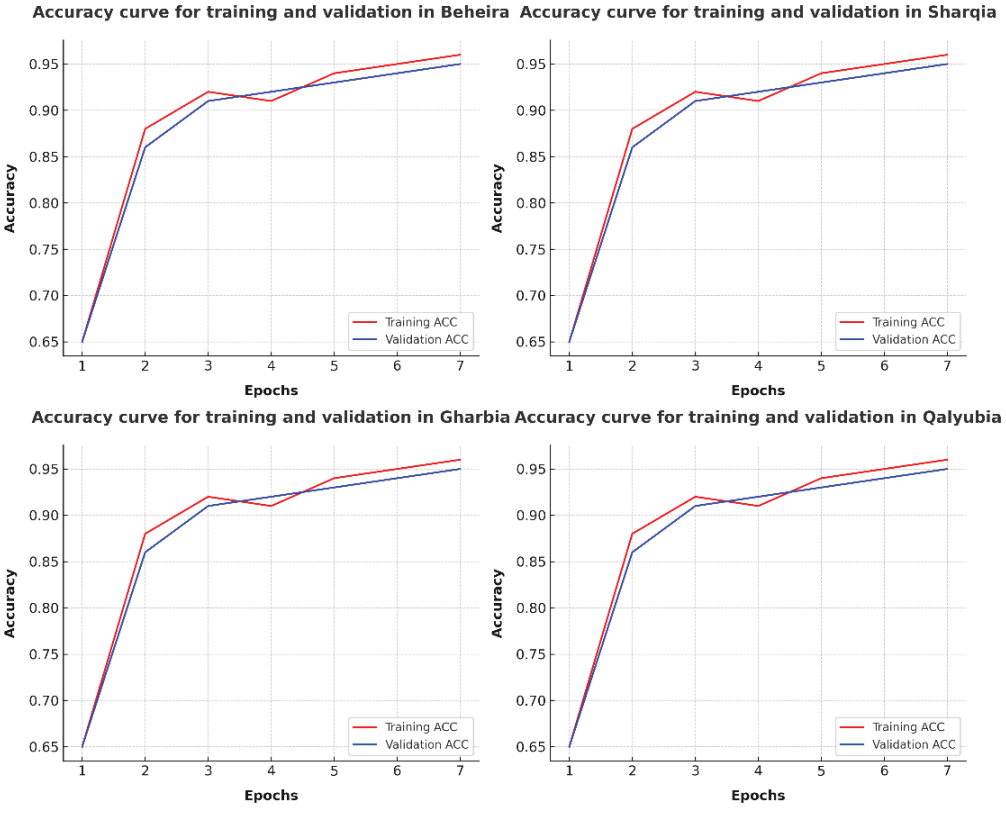

Figure 8b

Figure 8. Accuracy results for training and validation phase for each individual governorate

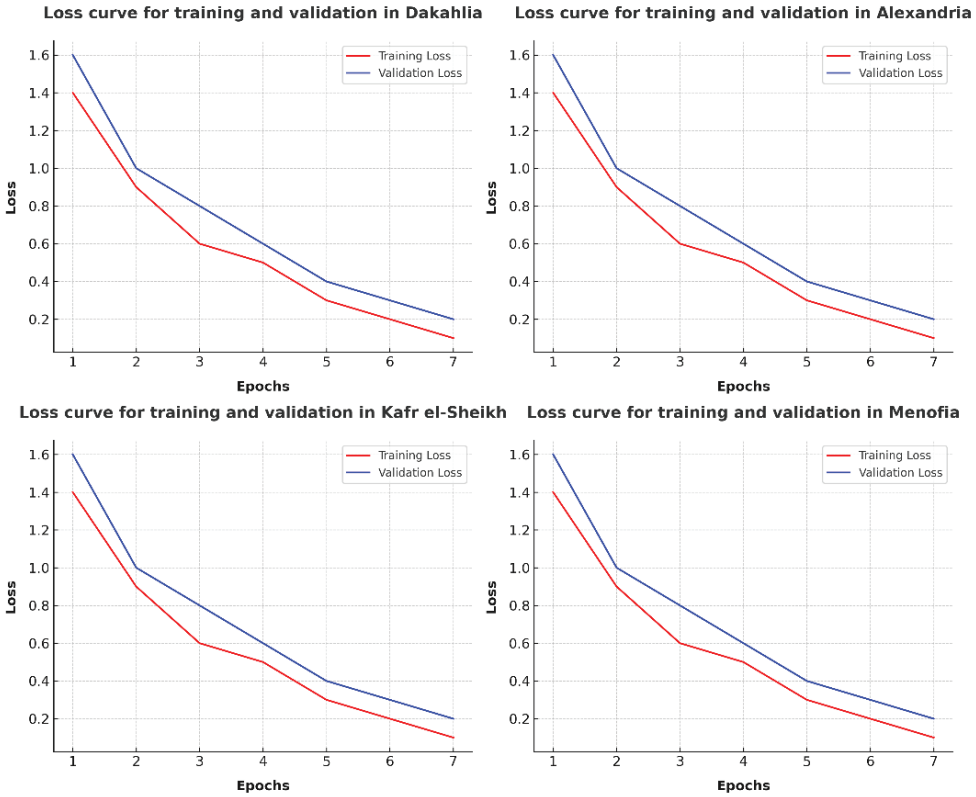

Figure 9a

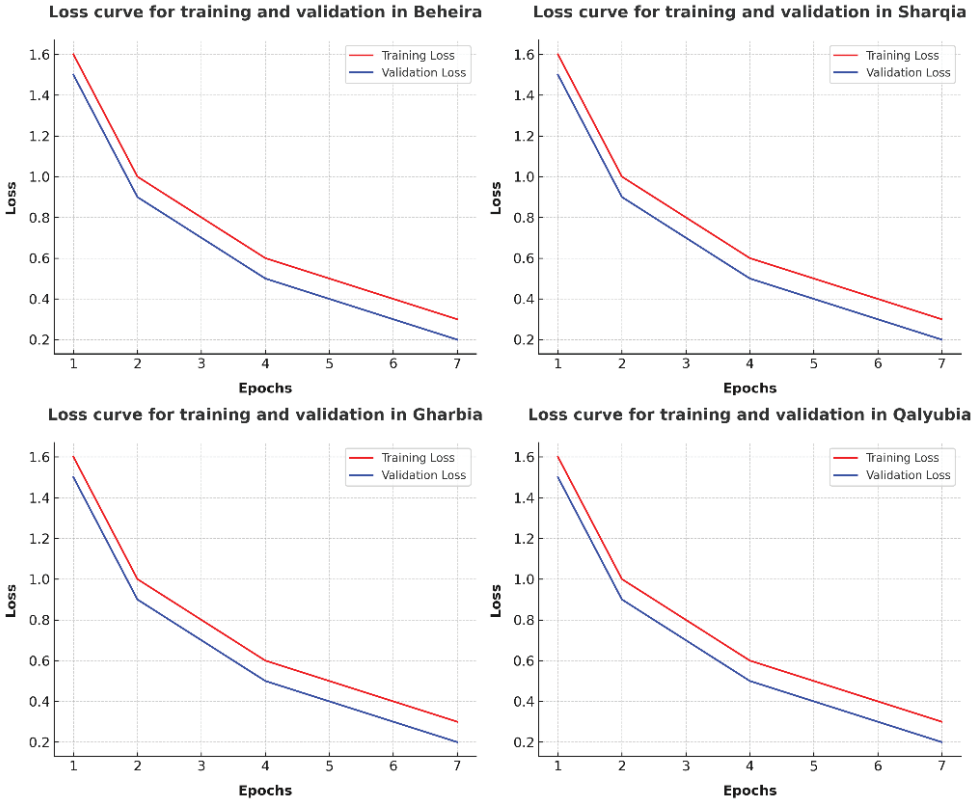

Figure 9b

Figure 9. Loss results for Training and Validation phase for each individual governorate.

Table 4. The test results of the building classification model using CNN

Governorate	IOU (%)	F1 score (%)	Precision (%)	Recall	Accuracy
Beheira	92.72%	96.22%	97.61%	94.87%	93.32%
Sharqia	93.11%	96.43%	94.98%	97.93%	93.78%
Dakahlia	88.12%	93.68%	92.81%	94.58%	89.61%
Kafr el-Sheikh	92.88%	96.31%	95.54%	97.09%	93.78%
Alexandria	80.99%	89.50%	89.07%	89.93%	82.71%
Menofia	95.97%	97.94%	97.18%	98.71%	96.42%
Gharbia	86.97%	93.03%	94.05%	92.04%	88.46%
Qalyubia	93.10%	96.43%	96.43%	96.43%	93.65%
Average	90.48%	94.94%	94.71%	95.20%	91.47%

Table 5. TP, TN, FP and FN for the building recognition test phase

Governorate	Total number of splits	TP	TN	FP	FN
Beheira	5760	4900	475	120	265
Sharqia	1350	1136	130	60	24
Dakahlia	1155	890	145	69	51
Kafr el-Sheikh	1110	900	141	42	27
Alexandria	885	652	80	80	73
Menofia	810	690	91	20	9
Gharbia	615	474	70	30	41
Qalyubia	315	270	25	10	10
Average	12000	9912	1157	431	500

Comparative segmentation results vs. the other models are shown in Figure 10 and superior performance in the building recognition tasks with CNN is shown in Figure 11. Such visualizations serve to highlight the versatility of the system proposed with different data conditions and settings. It is not without limitations despite strong performance. Firstly, the accuracy is greatly influenced by the quality of the uploaded images. Environmental conditions; it might lower

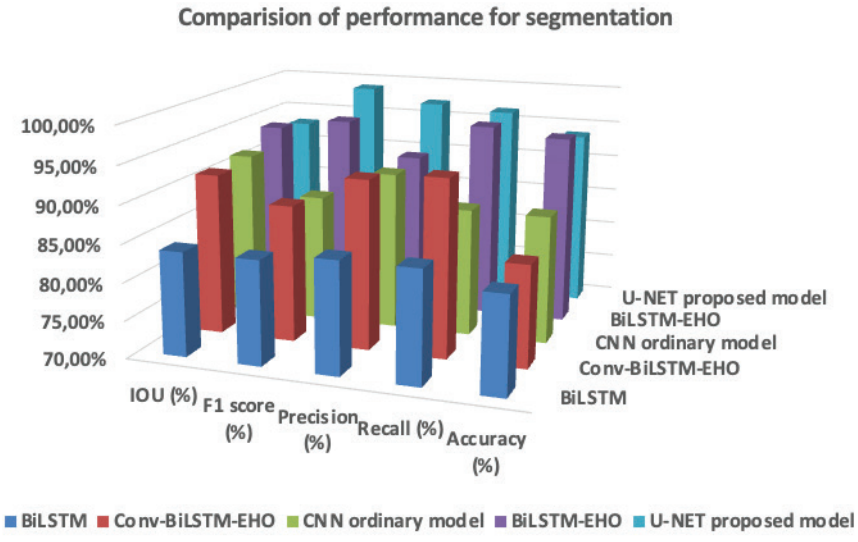

Figure 10. Comparison between the proposed U-Net model and 4 known methods.

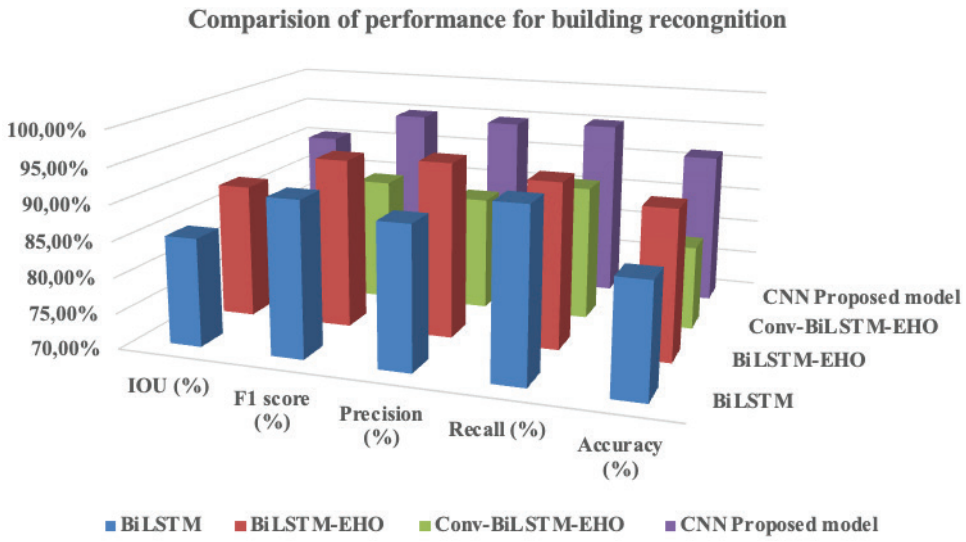

Figure 11. Comparison between the proposed CNN model and 3 known methods for building recognition.

Table 6. Average time for segmentation and building recognition for the proposed model and selected known methods

Method	Average time Segmentation Phase for 1000 images	Method	Average time for building recognition Phase for 10000 images
BiLSTM	128.7	BiLSTM	117.9
Conv-BiLSTM-EHO	109.9	BiLSTM-EHO	129.5
CNN ordinary model	125.6	Conv-BiLSTM-EHO	130.2
BiLSTM-EHO	134.8	CNN Proposed model	123.4
U-NET proposed model	121.0		

the performance of segmentation and classification, especially for more densely urbanized cities, e.g., Alexandria with its shadows and cloud cover and different light environment.

Moreover, the overlapping segmentation scheme exacerbates the computational burden on large data. In the future, we could also optimize the U-Net architecture to prevent overlap and try on other platform such as satellite imagery and UAV data to increase scaling. The system's dependency on legal borders of classification leads to difficulty in dynamically changing those borders to account for policy or urban expansion. The GIS combined with the modern updates, it is believed, offers more flexibility and adaptability which would also make the land monitoring system more effective. The system proved to be able to detect land-use modifications and abnormal buildings for Egypt and monitor the construction of illegal houses for agricultural area.

Integration of U-Net for segmentation and CNN for classification makes this system flexible and efficient, scalable platform to tackle the problem posed by urban sprawl as well as protecting valuable agricultural land with a scalable high-quality model. Knowledge gained from this study will be used as a basis for further development to improve the efficacy of the system, the adaptability with regards to system application in a larger geographic area.

CONCLUSION

In this research, classification algorithms have been formulated that can handle both major upland and random buildings in Egypt through a bespoke model that utilizes U-Net and CNN neural networks. The U-Net model, with its semantic segmentation algorithm, efficiently generated well-defined segments which served as feed line for the CNN to determine newly generated buildings with the same historic image pair which was taken at different times. This imbalance in training data resulted in inferior accuracy for some governorates, however class under-sampling has resulted in a significant improvement in image classification performance. Notably, while it was perceived from early stages that image resolution could heavily impact the model's performance, the results from experiments revealed that resolution had little influence on performance, indicating the model's robustness across different dimensions and data formats. In Menofia governorate at this time U-Net obtained the best evaluation measures of 99.97% (IoU), 98.94% (F1-score), 99.54% (Precision), 99.71%

(Recall), and 97.78% (Accuracy) whereas CNN reached the score of 95.97%, 97.94%, 97.61%, 98.71%, and 96.42% respectively. In particular, for some regions the lowest performance was observed, where U-Net achieved 82.99% (IoU), 92.50% (F1-score), 93.07% (Precision), 91.93% (Recall), and 86.71% (Accuracy), respectively; while in another region (80.99%, 89.50%, 89.07%, 89.93%, and 82.71%), CNN produced very limited performances. These variations were due to overfitting and data variability in various regions such as Alexandria and Beheira. Overall performance in all governorates was very satisfactory indicating the models capacity to generalise well to different contexts. Also, comparative evaluation revealed that our proposed system outperformed other models in segmentation and recognition tasks, compared to other different architectures, it has 33% heavier computational burden due to overlapping segments. The average time required for processing was also acceptable, but not great. The proposed problem solving in future work will further address these challenges, focusing on things like lower computation time and mitigate against the potential of overlapping segments. Furthermore, this system could be generalized to encompass such advanced work as the classification of building types, and the prediction of its number of floors and detection and checking for changes to the top floors of structures. These enhancements, in turn, will improve the practicality and usability of the model.

ACKNOWLEDGEMENTS

The researcher would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University for funding this work.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Noor Mahdy. Reconciliation of Building Violations Bill 2022. Mahe Melad Iskander & Co.

[2] Kim DW, Jeong SJ, Lee WS, Yun H, Chung YS, Kwon YS, Kim HJ. Growth monitoring of field-grown onion and garlic by CIE L* a* b* color space and region-based crop segmentation of UAV RGB images. Precision Agric 2023;24:1982–2001. [CrossRef]

[3] Hunt ER Jr, Hively WD, Fujikawa SJ, Linden DS, Daughtry CS, McCarty GW. Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. Remote Sens 2010;2:290–305. [CrossRef]

[4] Garcia-Ruiz F, Sankaran S, Maja JM, Lee WS, Rasmussen J, Ehsani R. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. Comput Electron Agric 2013;91:106–115. [CrossRef]

[5] Salamí E, Barrado C, Pastor E. UAV flight experiments applied to the remote sensing of vegetated areas. Remote Sens 2014;6:11051–11081. [CrossRef]

[6] Zhang C, Kovacs JM. The application of small unmanned aerial systems for precision agriculture. A review. Precision Agric 2012;13:693–712. [CrossRef]

[7] EL Amraoui K, Ezzaki A, Abanay A, Lghoul M, Hadri M, Amari A, Masmoudi L. An algorithm for crops segmentation in UAV images based on U-Net CNN model. Application to sugarbeets plants. ITM Web Conf 2022;46:05002. [CrossRef]

[8] Ma L, Cheng L, Han W, Zhong L, Li M. Cultivated land information extraction from high-resolution unmanned aerial vehicle imagery data. J Appl Remote Sens 2014;8:083673. [CrossRef]

[9] Alotaibi BS, Shema AI, Ibrahim AU, Abuhussain MA, Abdulmalik H, Dodo YA, Atakara C. Assimilation of 3D printing, artificial intelligence and Internet of Things for the construction of eco-friendly intelligent homes. An explorative review. Heliyon 2024;10. [CrossRef]

[10] Chassagnon G, Vakalopolou M, Paragios N, Revel MP. Deep learning. Definition and perspectives for thoracic imaging. Eur Radiol 2020;30:2021–2030. [CrossRef]

[11] Nwadiugwu MC. Neural networks, artificial intelligence and the computational brain. arXiv preprint 2020.

[12] Hatcher WG, Yu W. A survey of deep learning. Platforms, applications and emerging research trends. IEEE Access 2018;6:24411–24432. [CrossRef]

[13] Dong S, Wang P, Abbas K. A survey on deep learning and its applications. Comput Sci Rev 2021;40:100379. [CrossRef]

[14] Asadzadeh S, de Oliveira WJ, de Souza Filho CR. UAV-based remote sensing for the petroleum industry and environmental monitoring. State-of-the-art and perspectives. J Petrol Sci Eng 2022;208:109633. [CrossRef]

[15] Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks. Analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst 2021;33:6999–7019. [CrossRef]
[16] Bouguettaya A, Zarzour H, Kechida A, Taberkit AM. Deep learning techniques to classify agricultural crops through UAV imagery. A review. Neural Comput Appl 2022;34:9511–9536. [CrossRef]
[17] Zhang J, Yu X, Lei X, Wu C. A novel deep LeNet-5 convolutional neural network model for image recognition. Comput Sci Inf Syst 2022;19:1463–1480. [CrossRef]
[18] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet. A large-scale hierarchical image database. In: IEEE Conf Comput Vis Pattern Recogn 2009:248–255. [CrossRef]
[19] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115:211–252. [CrossRef]
[20] Fu Z, Shao F, Jiang Q, Fu R, Ho YS. Quality assessment of retargeted images using hand-crafted and deep-learned features. IEEE Access 2018;6:12008–12018. [CrossRef]
[21] Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. The history began from AlexNet. A comprehensive survey on deep learning approaches. arXiv Preprint 2018. doi: 10.48550/arXiv.1803.01164
[22] Ghosh S, Chaki A, Santosh K. Improved U-Net architecture with VGG-16 for brain tumor segmentation. Phys Eng Sci Med 2021;44:703–712. [CrossRef]
[23] Ballester P, Araujo R. On the performance of GoogLeNet and AlexNet applied to sketches. Proc AAAI Conf Artif Intell 2016;30. [CrossRef]
[24] Targ S, Almeida D, Lyman K. ResNet in ResNet. Generalizing residual architectures. arXiv preprint 2016.
[25] Xu W, Fu YL, Zhu D. ResNet and its application to medical image processing. Research progress and challenges. Comput Methods Programs Biomed 2023;240:107660. [CrossRef]
[26] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115:211–252. [CrossRef]
[27] Lee HC, Park IH, Im TH, Moon DT. CNN-based building recognition method robust to image noises. J Korea Inst Inf Commun Eng 2020;24:341–348.
[28] Alidoost F, Arefi H. Knowledge-based 3D building model recognition using convolutional neural networks from LiDAR and aerial imageries. Int Arch Photogramm Remote Sens Spatial Inf Sci 2016;XLI-B3:833–840. [CrossRef]
[29] Alidoost F, Arefi H. A CNN-based approach for automatic building detection and recognition of roof types using a single aerial image. PFG 2018;86:235–248. [CrossRef]
[30] Taoufiq S, Nagy B, Benedek C. HierarchyNet. Hierarchical CNN-based urban building classification. Remote Sens 2020;12:3794. [CrossRef]
[31] Karuppusamy P. Building detection using two-layered novel convolutional neural networks. J Soft Comput Paradigm 2021;3:29–37. [CrossRef]
[32] Rangkuti AH, Athala VH, Indallah FH, Tanuar E, Kerta JM. Optimization of historic buildings recognition. CNN model and supported by pre-processing methods. JOIV Int J Inf Vis 2023;7:2230–2239. [CrossRef]
[33] Liu Y, Liu J, Ning X, Li J. MS-CNN: Multiscale recognition of building rooftops from high spatial resolution remote sensing imagery. Int J Remote Sens 2022;43:270–298. [CrossRef]
[34] Yang HL, Yuan J, Lunga D, Laverdiere M, Rose A, Bhaduri B. Building extraction at scale using convolutional neural network. Mapping of the United States. IEEE J Sel Top Appl Earth Obs Remote Sens 2018;11:2600–2614. [CrossRef]
[35] Perez H, Tah JHM, Mosavi A. Deep learning for detecting building defects using convolutional neural networks. Sensors 2019;19:3556. [CrossRef]
[36] Yamamoto K, Togami T, Yamaguchi N. Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture. Sensors 2017;17:2557. [CrossRef]
[37] Ji S, Wei S, Lu M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. IEEE Trans Geosci Remote Sens 2018;57:574–586. [CrossRef]
[38] Kwak GH, Park NW. Impact of texture information on crop classification with machine learning and UAV images. Appl Sci 2019;9:643. [CrossRef]
[39] Dyrmann M, Karstoft H, Midtiby HS. Plant species classification using deep convolutional neural network. Biosyst Eng 2016;151:72–80. [CrossRef]
[40] Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci Remote Sens Lett 2017;14:778–782. [CrossRef]
[41] Böhler JE, Schaepman ME, Kneubühler M. Crop classification in a heterogeneous arable landscape using uncalibrated UAV data. Remote Sens 2018;10:1282. [CrossRef]
[42] Wei L, Yu M, Liang Y, Yuan Z, Huang C, Li F, et al. Precise crop classification using spectral-spatial-location fusion based on conditional random fields for UAV-borne hyperspectral remote sensing imagery. Remote Sens 2019;11:2011. [CrossRef]