UDC 528.8:004.932.1 **doi: 10.32620/reks.2024.2.04**

Hasmukh KORINGA, Miral PATEL

Government Engineering College, Rajkot, Gujarat, India

AUTOMATED, QUICK, AND PRECISE BUILDING EXTRACTION FROM AERIAL IMAGES USING LL-UNET MODEL

The subject matter of this article is the detection and semantic segmentation of buildings from high-resolution aerial images. It extracts building images from similar characteristics of roads and soil objects. It is used for various applications such as urban planning, infrastructure development, and disaster management. The goal of this study is to develop a fast, accurate, and automatic building detection model based on the semantic segmentation LL-UNET architecture that is optimized and tuned with proper hyper parameter settings. The tasks to be solved are as follows: collect remote sensing building dataset that is divided into three parts of training, validation, and testing; apply data augmentation on the training dataset by vertical flip, horizontal flip, and rotation methods; further pass into the bilateral filter to remove noises from the images; optimize LL-UNET model by various optimizer methods and tuned hyper parameter by proper selection value, the method is compared by the performance metrics recall, precision, and accuracy. The following results were obtained: the model is evaluated under the training loss curve and accuracy curve of different optimizers SGD, ASGD, ADAM, ADAMW, and RMSProp; it measures the training time, mean accuracy, and mean IOU parameters during the training phase; the testing phase is evaluated by precision and recall; the method is compared by visualizing the result of LL-UNET + different optimizers; and the proposed method is compared with the existing method by common evolution parameters metric. Conclusions. The scientific novelty of the results obtained is as follows: 1) the LL-UNET effectively segmented the building remotely sensed images in the limited number of training samples available; 2) the loss function of the model observed under hyper parameter selection of the optimizer, learning rate, batch size, and epochs; which makes an optimal model to extract the building in an accurate and fast manner from the complex background; 3) the proposed model results compared with a well-known model of the building extraction under the common evaluation metrics of F1 score.

Keywords: *Deep Learning; Semantic Segmentation; LL-UNET; Buildings.*

1. Introduction

1.1. Motivation

Buildings extraction from aerial images is the process of automatically identifying and extracting the boundaries of buildings from high-resolution aerial or satellite images. The extracted building footprints can be used for a variety of applications, such as urban planning, disaster management, city development, land management, environmental monitoring, and 3D modeling. The results of building extraction from aerial images depend on several factors, such as the quality and resolution of the image and the choice of algorithm. The process of building extraction typically involves a series of steps, including image preprocessing, feature extraction, and classification. Classification involves the use of machine learning algorithms to classify each pixel in the image as either building or non-building. Comparing and recognizing similar features between two or more photos is known as image feature matching technology as suggested in [1]. The fine segmentation of remote sensing is not thought to be sufficient for most segmentation techniques currently in use; however, because of the Similar Buildings feature extraction from aerial images can be challenging due to factors such as varying building sizes and shapes, shadows, occlusions, and inherent uncertainties in aerial images, which have extremely complex resolutions, which frequently cause ambiguity among some geographic entities during the segmentation process [2].

These problems drive our research, which suggests a better method for building extraction based on deep learning less layered UNET architecture (LL-UNET) to overcome these obstacles. However, recent advances in deep learning and computer vision techniques have led to significant improvements in the accuracy and efficiency of building extraction methods. Several research studies have used deep learning methods for the extraction of fast and accurate buildings from aerial images.

1.2. Objectives

This study aims to develop an automatic, fast, and accurate building detection model based on aerial images. To obtain this, we constructed a model based on

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

deep learning that detects the buildings and background from the testing dataset. To accomplish the research objective, the following tasks have been developed as under.

1. Training, Validating, and Testing the building detection model based on deep neural network.
2. Preprocessing and data augmentation of the building detection dataset.
3. Developing a model for feature extraction and semantic segmentation using optimizers of SGD, ASGD, ADAM, ADAMW, and RMSProp.
4. Comparative analysis to assess the performance of the different models.

2. Literature Review

Urban areas of any country have high population density and crowded human-built structures such as roads, buildings, and other infrastructures. Nowadays, satellites are used for Earth observation through satellite images. These images are used to monitor and assess various natural resources and manmade infrastructures. Among these, one of the complex tasks is to detect buildings from aerial images. Perfect segmentation and classification of buildings from aerial images is a complex and challenging task. High-resolution satellite images provide high-quality images and help to classify images into various classes such as buildings and background. Manual detection of buildings from aerial images is a tedious task. The automatic approach of building detection from aerial images is highly demanding. The most commonly used algorithms for building or road extraction from aerial images include object-based image analysis (OBIA) [3], deep learning [4, 5], and traditional image processing techniques such as thresholding [6, 7] and segmentation [8, 9]. Detection of buildings can be used for city planning, urban mapping, urban change detection, natural disaster damage assessment, rescue systems, geographic information system (GIS) updates, environmental monitoring, virtual augmented reality, and defense systems [10]. Varying color textures, construction, and different shapes make the task more difficult.

In earlier days, building detection and classification from satellite images used texture filtering proposed by [11]. This method was not accurate in studying the rapid growth of urban areas and city development. The automatic building detection from high-resolution images and LiDAR data is done by the threshold-based method and object-based classification methods suggested by [6]. However, object-based classification requires higher procedure complexity, whereas threshold-based methods need to define only two threshold levels. A. Femin and K. S. Biju [12] proposed a CNN base building detection method from satellite images irrespective of different textures, different rooftops of various colors, and different shapes of rooftops. S. Kala and M. K. Jayakumar [13] proposed three-step methods to segment the image into buildings and backgrounds. First, preprocess the input image for noise reduction. The Adaptive FCM-based clustering method was then used for segmentation. In the last step, post-processing is performed to remove small objects like roads from the segmented image. Ryuhei and Shuhei suggested a multitask UNet model for the detection of various-sized buildings and roads from deep globe building datasets [14]. A method for building detection from high-resolution satellite images using improved data labeling and improved UNet with a batch normalization layer added to the UNet [15]. Haiying Wang and Miao [16] proposed building detection from aerial images using a deep residual. The methods have the advantages of UNet, atrous spatial pyramid pooling, residual learning, and focal loss. The modified residual UNet model reduces the learning parameters and degradation of the network. The model failed in some cases, such as misclassification of concrete floors, shadows considered as buildings, and poor identification effect of adjacent buildings. Wojciech Sirko [17] proposed a supervised semi-automatic UNet model for satellite image semantic segmentation that classifies each pixel of the image as building or non-building. The connected component method is used to determine the threshold level. Super-pixel-CRF for weakly supervised construction of semantic segmentation using initial deep seeds as guidance presented [18].

Recently, we proposed a large-scale building extraction framework based on super-resolution (SR) and instance segmentation using an open-sourced dataset with a considerably lower-resolution image [19]. To reduce the computational burden of building extraction, the method suggests that the scene classification model based on the ConvNeXt-based network can successfully distinguish between images with and without buildings. When dense building structures are present in the remote sensing photos, it is impossible to separate the building and background from the input images. [20] suggested a method to detect the building features from the highly dense area of the building. It enhances the image by super-resolution followed by polygonization techniques, which improve accuracy and robustness for building outline extraction in these highly dense complex scenes. [21] suggested using the uncertainty rank approach to map the encoder-decoder network with a high relative uncertainty feature for building extraction. Although it improves the accuracy of building detection, the uncertainty rank method weight is not adaptable. [22] and [23] suggested a building interpretation model that not only extracts the building but also detects changes in the building that indicate the rate of urban area development.

In this paper, we propose semantic segmentation for building extraction using the open-source dataset of Massachusetts [24]. The proposed method used LL-UNET-based semantic segmentation architecture for building extraction from high spatial resolution aerial images. The LL-UNET-based deep learning method combined with data augmentation and preprocessing steps to increase the F1 score by 10% more than the existing methods.

3. Materials and Methods of Research

3.1. Proposed Method

Building detection from aerial images is a challenging task because of occlusion and similar-type type characteristic objects present in the RSI. We attempted to extract building images from the input as described in Fig. 1. It has several stages to detect the building data from RSI, such as data preprocessing, where the input image size is larger so it is resized to 512x512 pixels. Further input images are taken by satellite or by drone, so they are noisy images. Therefore, a bilateral filter is used to remove noise and preserve the contrast of the image [25]. Furthermore, De-noising input image dataset is split into training, validation, and testing sets. As a limited number of training samples are available, it is augmented to increase the training dataset size. The proposed model based on the semantic segmentation architecture UNet [26] was originally developed for biomedical image segmentation applications but has since gained popularity for a variety of semantic segmentation applications [27 – 29]. The proposed method is based on UNet but has less convolution layered to detect the building from the aerial images. Due to the smaller number of convolution layers, it detects the building data from the input image quickly. However, the building is not a tiny component, so it detects accurately when layered are reduced from the original UNet.

The LL-UNet architecture and skip connection are shown in Fig2. Originally, UNet had four units of Unet-down sample, four units of Unet-uP sample, and two units of the Unet-Middle block. The suggested less layered method has three units of Unet-down, three units of Unet-uP, and one unit of Unet-Middle block. This causes a total of seven convolution layers to be reduced.

The detailed layer of each block of the Unet-down, Unet-up, and Unet-Middle blocks is shown in Fig. 3. The Unet-down sample consists of two 3x3 convolution layers, each followed by a Leaky ReLu activation function that introduces nonlinearity into the network for better generalization. The after Max Pooling function with stride = 2 is used to half the image size. During the Unet-down sample, the network learns the features of the input image.

Skip connection allows us to bring lost information of images previously done during the Unet-down sample block to the Unet-up sample block, which gives better image segmentation.

The Unet-middle block consists of two convolution layers followed by the Leaky Relu activation function that bridges the Unet-down and Unet-up sample blocks.

The Unet-up sample consist of the 2x2 de-convolution unit followed by two convolution units with a filter size of 3x3 and the same padding. The concat function concatenated the features convoluted output of the Unet-up sample block and the Unet-down sample unit by the skip connection. This provides the features of images that are lost due to the depth of the network.

The Hyperparameter of the UNet network is tuned by several epochs, learning rate, batch size, and loss function along with different types of optimizers such as SGD, ASGD, ADAM, ADAMW, and RMSProp [30, 31]. During the Training Phase, Ground Truth and Predicted image mask are compared concerning the chosen loss model of the Less Layered UNET, and accordingly, the weight parameter and learning rate of the proposed method are calculated by the back propagation algorithm under the consideration of various optimization techniques. After training and validation

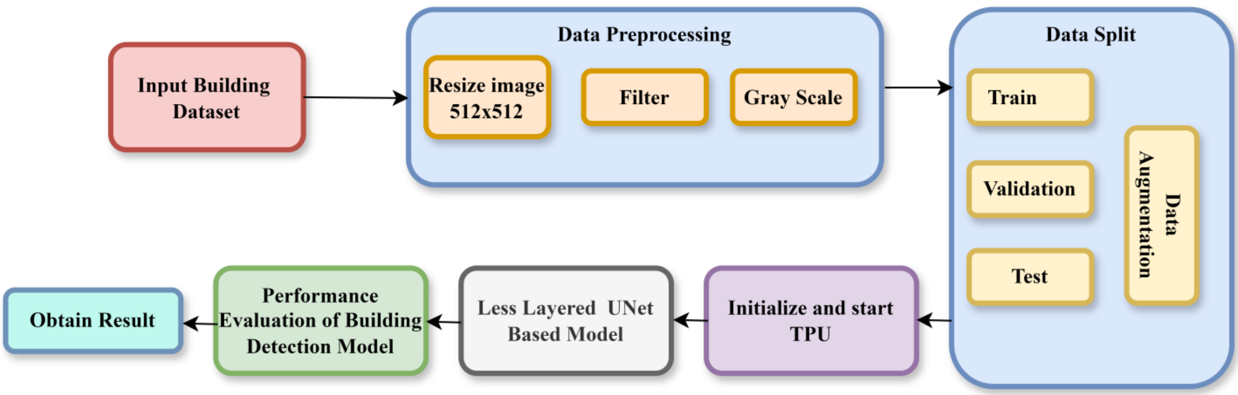


Fig. 1. Block Diagram of Building Detection System from Aerial images

Radioelectronic and Computer Systems, 2024, no. 2(110)
ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

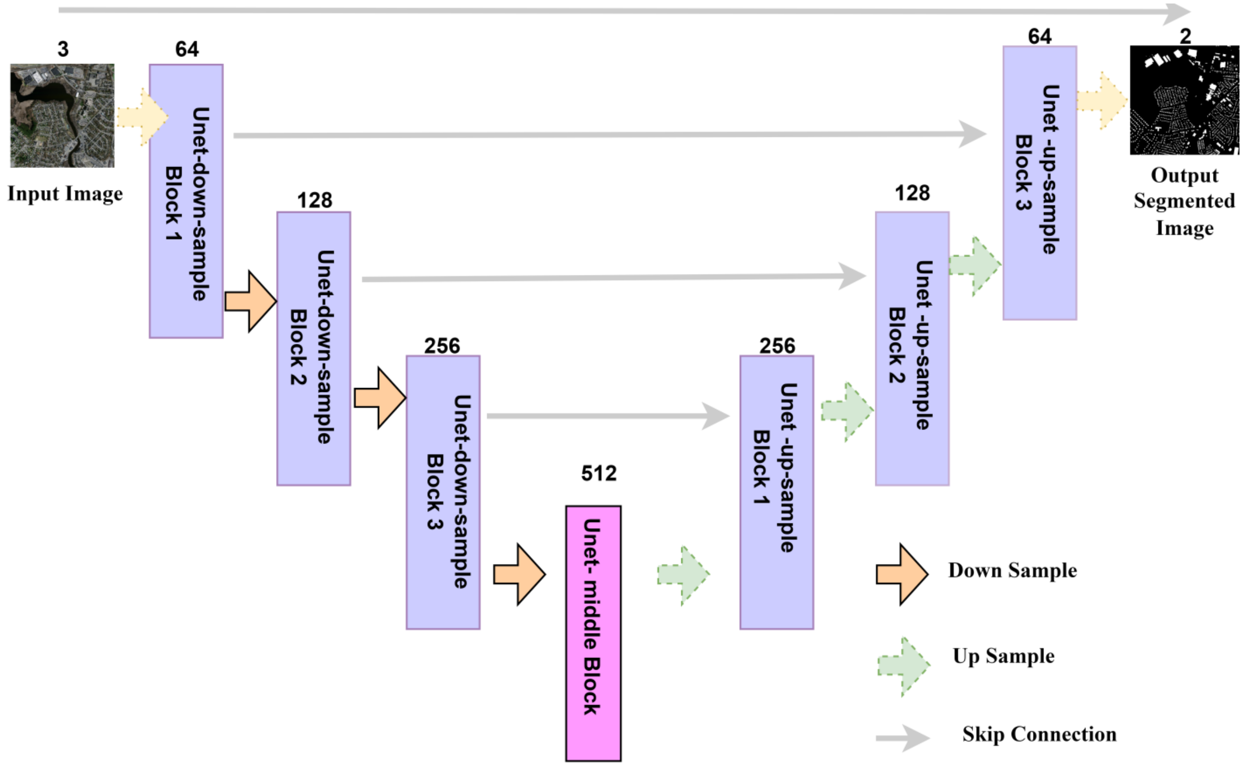
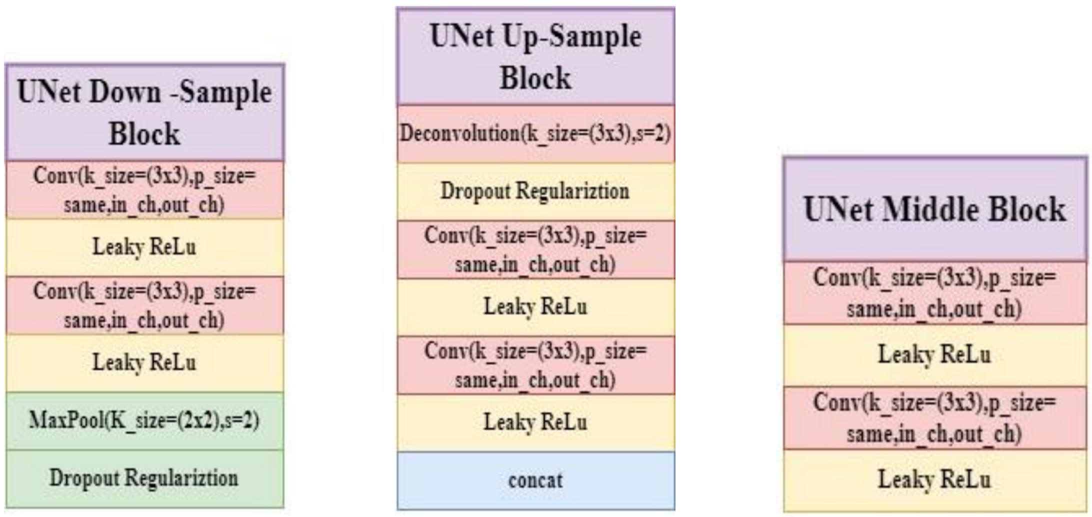


Fig. 2. LL-UNET Architecture



(a) (b) (c)

Fig. 3. Detailed layer of UNet architecture Block

phase, the trained model is saved as a less layered building detection model into the disk, and testing was performed by loading testing dataset images and measuring the model mean accuracy, precision, and recall parameter for building detection.

3.2. Building the Training Dataset

Two open source Massachusetts and WHU building datasets were used to train and validate the model. The Massachusetts Buildings Dataset comprises 151 aerial images covering 2.25 km2 in the Boston region, with an image size is 1500 × 1500 pixels. The dataset was divided into a training set of 137 images, a validation set of 4 images, and a test set of 10 images. The dataset considers urban and suburban areas of buildings of all sizes, including high-rise and single-storied buildings. Every image has been resized to have a resolution of 1 pixel /m^2. To improve the accuracy of the evaluations, target maps for the dataset's test and validation sections were manually corrected.

The WHU building extraction dataset is a subset of aerial images. A total of 8188 non-overlapping RGB images have a size of 512 x 512 pixels. The images were taken above Christchurch; New Zealand, with a spatial resolution varying between 0.0075 and 0.3 meters. A total of 600 images were randomly taken from a 8188 images. The taken images were divided into three categories: training set (540), validation set (170), and test set (30).

3.3. Hyper Parameter Selection

Before the training process starts, a set of variables known as hyper parameters is often chosen through experimentation [32]. The learning rate, batch size, and number of epochs are typically among them. Hyper-parameter tuning is the process of achieving optimal hyper-parameter settings [33]. Through experimentation with different permutations, the values of the various hyper parameters employed in the proposed work are displayed in Table 1. Ultimately, the parameter values that yielded the best results for building semantic segmentation were chosen. Vertical and horizontal flip operations were performed on the original image for data augmentation using a random probability of 0.5. Further rotation by 45^0 and 90^0 was used to randomly augment the image set with a probability of 0.75.

Table 1
Hyper parameter-tuned value for building segmentation

Hyper-Parameter	Value
Model	LL-UNET
Initial Learning Rate	10^{-4}
Image Size	512x512 pixel
Batch Size	4
Target Labels	2
Data Augmentation	Vertical Flip, Horizontal Flip, Rotation = 45^0
Loss Function	Cross Entropy Loss
Optimize Algorithm	SGD, ASGD, ADAM, ADAMW, RMSProp

3.4. Performance Metrics

Performance metric measurement accuracy, precision, recall, IOU, and F1-score are generally used for semantic segmentation effectiveness. It is derived from the test dataset's confusion matrix. It defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{3}$$

$$\text{F1-score} = 2 * \text{Recall} \frac{\text{Precision}}{\text{Precision} + \text{Recall}}, \tag{4}$$

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \tag{5}$$

where TP, TN, FP, and FN indicate the number of pixels of True Positive, True Negative, False Positive, and False Negative respectively. Accuracy shows how the model performs overall in all classes. Recall is the ratio of the number of correctly identified positive samples to all positive samples in the test set. The precision of an image is defined as the proportion of accurately predicted positive samples among the entire expected positive sample. The F1 score represents the harmonic average value of precision and Recall. IOU represents the expected and original image overlap ratios.

4. Results and Analysis

This section describes the findings of semantic segmentation of buildings from the WHU and Massachusetts aerial datasets that are publicly accessible.

4.1. Segmentation of the building

The input image for testing, validation, and training has a dimension of 1500 by 1500 pixels. Images entered into the network were scaled to 512x512 throughout the training phase. The LL-UNET model is trained with a batch size of 4, 64 hidden layers, and a 10^{-4} learning rate over 100 epochs. Initially, the model is trained by traditional gradient-based SGD and ASGD optimization techniques. In addition, adaptive estimation of momentum-based optimization methods, including ADAMW, RMSProp, and ADAM, was used to train the suggested model.

4.2. Classical optimizer and adaptive momentum-based Training

The optimization techniques SGD and ASGD were used to train the first version of the model. Graph of Training Loss vs. Epochs are shown in Fig. 4. As Fig. 4 illustrates, training loss is excessive during the first few epochs. The model is trained using both classical and adaptive momentum-based optimizers, which alter the loss behavior according to the optimizer.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Fig. 4, a illustrates how, in the SGD and ASGD algorithms, the training loss dropped to about half of the initial loss after a progressive number of epochs. Initially, the loss was excessively large. Moreover, the loss reduction profile for the ADAMW, RMSProp, and ADAM optimizers is also displayed in Fig. 4, b. At training for the 100th epoch, ADAM has less loss. It is approximately 18% less than the traditional method based on optimizers.

The accuracy curves for several optimizers employed in the training and validation stages are displayed in Fig. 5. These results demonstrate that for different numbers of epochs, the training and validation accuracy for SGD and ASGD are nearly equal. However, as the epochs went by throughout the training and validation phase, as illustrated in Fig. 5, the accuracy curve for the ADAMW, RMSProp, and ADAM-based optimizer progressively increased. Thus, it can be seen from Fig. 5 that an adaptive momentum-based optimizer works well when building detection models from aerial images.

4.3. Performance Metrics Comparison for the Proposed Method

Various performance metrics are used during the training and testing phases to measure building segmentation effectiveness. After training the building detection model, the average training accuracy (mAccuracy) and average intersect over union (mIOU) and Training Time (TT) [34] are calculated and noted in Table 2. It was concluded that ADAM has the highest accuracy and IOU, whereas ADAMW has the lowest IOU and accuracy. On the other hand, RMSProp has an IOU of 66.12% and an accuracy of 84.87%.

Table 2

Performance Metrics Comparison for Various Optimizers During the Training Phase

Optimizer	Performance Metrics		
	mAccuracy (%)	mIOU (%)	Training Time (TT) (sec)
SGD	65.02	23.44	5066.64
ASGD	70.23	21.23	5527.43
ADAM	84.66	66.42	5000.79
ADAMW	82.63	64.67	4980.23
RMSprop	84.87	66.12	5126.87

The Building Detection model's testing phase performance metrics, recall and precision, are computed in Table 3. This reveals that the precision of SGD and ASGD is close to 54% and the recall is almost 0%. Consequently, it was not possible for these optimizers to identify buildings from the aerial images. Additionally, ADAM has the highest precision and recall, whereas RMSProp has a recall parameter of 93.78% and precision of 76.38%. As a result, the proposed model employed the ADAM optimizer-based technique to identify the building from aerial photos. The visual result is displayed in Fig. 6 also conveys the results of the proposed strategy.

Table 3

Performance Metrics Comparison for Various Optimizers During the Testing Phase

Optimizer	Performance Metrics	
	Recall (%)	Precision (%)
SGD	0	53.39
ASGD	0	52.24
ADAM	94.92	78.84
ADAMW	92.56	73.62
RMSprop	93.78	76.38

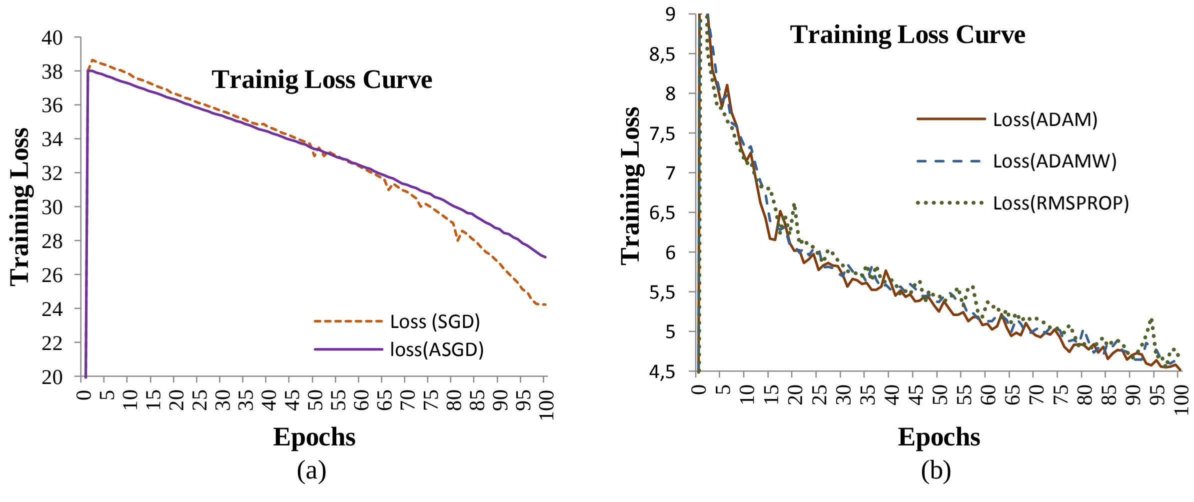


Fig. 4. Training Loss vs Epochs for (a) SGD, ASGD (b) ADAM, ADAMW, and RMSProp during the training Phase

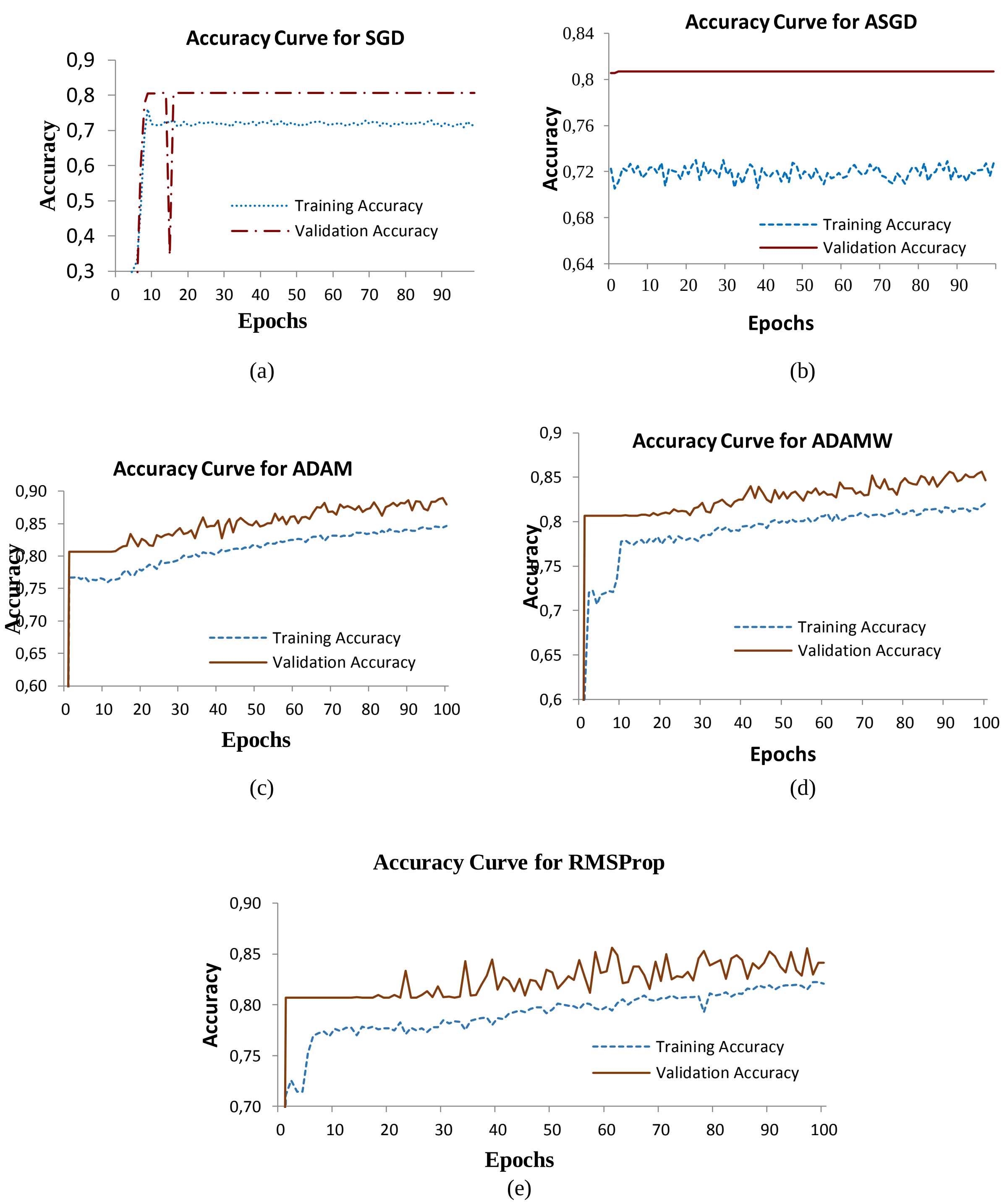


Fig. 5. Accuracy Curve for (a)SGD, (b) ASGD, (c) ADAM, (d) ADAMW, and (e) RMSProp during the Training and Validation Phase

4.4. Comparison with Existing Methods

The challenge of detecting buildings from aerial photos is difficult because the target building data appears to be similar to the background images. Cities these days are expanding quickly, so it is necessary to detect this increase from aerial photos. Numerous other things will benefit from this as well, such as city planning, infrastructure development management, and urban planning and management [4, 33]. Reliability, speed, accuracy, and automation are challenging to achieve when extracting a building. More time, money, and efficiency are required for people to identify a building using aerial images [15, 5]. However, by automatically extracting the building from the satellite photos, these issues are eliminated.

Reference [15] proposed three three-step procedures for using the morphological building index

Fig. 6. Visual Results for Building Detection

approach to identify the built-up area. It used five satellite photos with a spatial resolution of one meter and four spectral bands. [35] separated the vegetation and buildings from the LIDAR image using Convolutional Neural Network techniques based on the k closest approach. Subsequently, the same type of item building and backdrop were segmented using the Euclidean Clustering technique. The technique fails when there is less space between two buildings, but it works for high-resolution photos. Reference [16] presented a unique UNet-based model that uses an ADAM optimizer and an up-sampling pooling operator. High-resolution aerial image evaluation is performed using the proposed model. The approach works well for larger buildings; however, it falls short for smaller buildings.

However, Reference [14] proposed a method for building detection based on UNET along with VGG and RESNET. It is used to extract every size of the building from the remote sensing building images. The method used the building detection dataset provided in the DeepGlobe competition. All the existing methods along with our proposed method are compared under the common performance metrics F1-score, as shown in Table 4. The model [26] UNet architecture with a VGG16 backbone and ADAM optimizer has an F1-score of 83.55%. The suggested approach of the LL-UNET model F1-score of 86.13% performs better than the other methods. Furthermore, because there were fewer convolution layers in the original UNET, it took less time to train and detect. However, the model is trained with fewer samples.

Table 4
Results Comparison with the Existing Method

	F1-score (%)
[15]	66.40
[16]	70.00
[14]	73.99
[35]	80.87
[26]	83.55
ours	**86.13**

5. Conclusion

The main problems of building segmentation from aerial images are the size, shape, and color of buildings due to similar background elements. Deep Learning provides better automatic object segmentation from aerial images. Various deep learning-based architectures such as UNET, SEGNET, and CNN have been used by researchers to detect buildings from high-resolution aerial images. The proposed LL-UNET deep learning-based architecture extracts buildings more effectively from aerial images compared to existing methods. The semantic segmentation architecture-based implemented LL-UNET obtained high accuracy with fewer training samples and processing power. After the implementation of the LL- UNET model, it is trained, validated, and tested using the WHU and Massachusetts datasets. The proposed model is trained with hyper parameters of the Learning rate of 10^{-4}, a Batch size of 4, and hidden layers of 64 for 100 epochs. Furthermore, the model is trained by various optimizers to calculate the loss value between the predicted building and the actual building pixel. The model obtained 86.13 % optimum F1-score value using ADAM optimizer and Leaky ReLu activation function. The experimental results show that our proposed method improves the total F1 score by more than 5% compared with the existing methods. Furthermore, in the future, this model can be improved using an unsupervised deep learning algorithm to make it more generalized, fast, and accurate.

Contributions of the authors

Conceptualization, methodology, formulation of task, development of a model, writing and reviewing draft – **Miral Patel;** formulation of tasks, analysis of the method, development of the model, analysis of the result, writing and reviewing draft **– Hasmukh Koringa.**

Conflict of Interest

The authors declare that they have no conflict of interest about this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence methods while creating the presented work.

All the authors have read and agreed to the published version of this manuscript.

References

1. Chen, Q. Y., & Feng, D. Z. Feature matching of remote-sensing images based on bilateral local–global structure consistency. *IET Image Processing*, 2023, vol. 17, iss. 14, pp. 3909-3926. DOI: 10.1049/ipr2.12907.

2. Ma, X, Xu, J., Chong, Q., Ou, S., Xing, H., & Ni, M. FCUnet: Refined remote sensing image segmentation method based on a fuzzy deep learning conditional random field network. *IET Image Processing*, 2023, vol. 17, iss. 12, pp. 3616-3629. DOI: 10.1049/ipr2.12870.

3. Norman, M., Shahar, H. M., Mohamad, Z., Rahim, A., Mohd, F. A., & Shafri, H. Z. M. Urban building detection using object-based image analysis (OBIA) and machine learning (ML) algorithms. *IOP Conference Series: Earth and Environmental Science*, 2021, vol. 620, article no. 012010. 12 p. DOI: 10.1088/1755-1315/620/1/012010.

4. Alsabhan, W., Alotaiby, T., & Dudin, B. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. *Hindawi Computational Intelligence and Neuroscience*, vol. 2022, article no. 4831223. DOI: 10.1155/2022/4831223.

5. Kothari, A. M., & Patel, M. Road network extraction methods from remote sensing images: a review paperx. *International Journal of Next-Generation Computing*, 2022, vol. 13, no. 2, DOI: 10.47164/ijngc.v13i2.376.

6. Hermosilla, T., Ruiz, L. A., Recio, J. A., & Estornell, J. Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data. *Remote Sens.*, 2011, vol. 3, iss. 6, pp. 1188-1210. DOI: 10.3390/rs3061188.

7. You, Y., Wang, S., Ma, Y., Chen, G., Wang, B., Shen, M., & Liu, W. Building Detection from VHR Remote Sensing Imagery Based on the Morphological Building Index. *Remote Sens.*, 2018, vol. 10, iss. 8, article no. 1287. DOI: 10.3390/rs10081287.

8. Patel, M. J., & Kothari, A. M. Deep Learning-Enabled Road Segmentation and Edge-Centerline Extraction from High-Resolution Remote Sensing Images. *International Journal of Image and Graphics*, 2023, vol. 23, no. 06, article no. 2350058. DOI: 10.1142/S0219467823500584.

9. Parikh, Y., & Dr. Koringa, H. A Systematic Analysis of CMR Segmentation Using Deep Learning. *International Journal of Next-Generation Computing*,

2022, vol. 13 special iss. 3. DOI: 10.47164/ijngc.v13i3.825, 2022.

10. Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J., & Liang, D. Big Earth Data: a new challenge and opportunity for Digital Earth's development. *Int J Digit Earth*, 2017, vol. 10, iss. 1, pp. 1-12. DOI: 10.1080/17538947.2016.1264490.

11. Zhang, Y. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS Journal of Photogrammetry & Remote Sensing*, 1999, vol. 54, iss. 1, pp. 50-60. DOI: 10.1016/S0924-2716(98)00027-6.

12. Femin, A., & Biju, K. S. Accurate Detection of Buildings from Satellite Images using CNN. *2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE)*, Istanbul, Turkey, 12-13 June 2020, pp. 1-5. DOI: 10.1109/ICECCE49384.2020.9179232.

13. Kala, S., & Jeyakumar, M. K. A Proficient Satellite Image Building Detection Using Fuzzy and Neural Network Techniques. *International Journal of Engineering Research and Technology*, 2019, vol. 12, no. 1, pp. 89-96. Available at: https://www.ripublication.com/irph/ijert19/ijertv12n1_13.pdf (accessed 12.12.2023).

14. Hamaguchi, R., & Hikosaka, S. Building detection from satellite imagery using ensemble of size-specific detectors. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018, pp. 187-191. DOI: 10.1109/CVPRW.2018.00041.

15. Afanasyev, I., & Prathap, G. Deep Learning Approach for Building Detection in Satellite Multispectral Imagery. *International Conference on Intelligent Systems (IS)*, Funchal, Portugal, 2018, pp. 461-465. DOI: 10.1109/IS.2018.8710471.

16. Wang, H., & Miao, F. Building extraction from remote sensing images using deep residual U-Net. *European Journal of Remote Sensing*, 2022, vol. 55, no. 1, pp. 71-85. DOI: 10.1080/22797254.2021.2018944.

17. Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., & Quinn, J. Continental-Scale Building Detection From High Resolution Satellite Imagery. *arXiv preprint arXiv:2107.12283*, 2021. DOI: 10.48550/arXiv.2107.12283.

18. Moghalles, K., Li, H.-C., Al-Huda, Z., Raza, A., & Malik, A. Weakly supervised building semantic segmentation via superpixel-CRF with initial deep seeds guiding. *IET Image Processing*, 2022, vol. 16, no. 12, pp. 3258-3267. DOI: 10.1049/ipr2.12558.

19. Chen, S., Ogawa, Y., Zhao, C., & Sekimoto, Y. Large-scale individual building extraction from open-source satellite imagery via super-resolution-based instance segmentation approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, vol. 195, pp. 129-152. DOI: 10.1016/j.isprsjprs.2022.11.006.

20. Nguyen, V., Ho, T. A., Vu, D. A., Anh, N. T. N., & Thang, T. N. Building footprint extraction in dense areas using super resolution and frame field learning. *12 th International Conference on Awareness Science and Technology (iCAST)*, November 2023, pp. 112-117. DOI: 10.1109/iCAST57874.2023.10359290.

21. He, W., Li, J., Cao, W., Zhang, L., & Zhang, H. Building extraction from remote sensing images via an uncertainty-aware network. *arXiv preprint arXiv:2307.12309*, 2023. DOI: 10.48550/arXiv.2307.12309.

22. Wang, M., Chen, K., Su, L., Yan, C., Xu, S., Zhang, H., Yuan, P., Jiang, X., & Zhang, B. RSBuilding: Towards General Remote Sensing Image Building Extraction and Change Detection with Foundation Model. *arXiv preprint arXiv:2403.07564*, 2024. DOI: 10.48550/arXiv.2403.07564.

23. Liu, Y., Pang, C., Zhan, Z., Zhang, X., & Yang, X. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geoscience and Remote Sensing Letters*, 2020, vol. 18, no. 5, pp. 811-815. DOI: 10.1109/LGRS.2020.2988032.

24. Mnih, V. *Road and Building Detection Datasets. Massachusetts Buildings Dataset*. Available at: https://www.cs.toronto.edu/~vmnih/data/. (accessed 12.12.2023).

25. Paris, S., Kornprobst, P., Tumblin, J., & Durand, F. Bilateral Filtering: Theory and Applications. *Foundations and Trends in Computer Graphics and Vision*, 2009, vol. 4, no. 1, pp. 1-73. DOI: 10.1561/0600000020.

26. Ronneberger, O., Fischer, P., & Brox, T. UNet: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*, 2015, pp. 1-8. DOI: 10.48550/arXiv.1505.04597.

27. Patel, M. J., Kothari, A. M., & Koringa, H. P A novel approach for semantic segmentation of automatic road network extractions from remote sensing images by modified UNet. *Radioelectronic and Computer Systems*, 2022, no. 3, pp. 161-173. DOI: 10.32620/reks.2022.3.12.

28. Patel, M., & Koringa, H. P. Deep Learning Architecture U-Net Based Road Network Detection from Remote Sensing Images. *International Journal of Next-Generation Computing*, 2023, vol. 14, no. 3. DOI: 10.47164/ijngc.v14i3.1301.

29. Parikh, Y., & Koringa, H. Left Ventricle Segmentation using Bidirectional Convolution Dense Unet. *Journal of Integrated Science & Technology*, 2023, vol. 11, no. 1, article no. 417. Available at: https://pubs.thesciencein.org/journal/index.php/jist/article/view/417/297 (accessed 12.12.2023).

30. Gupta, A., Ramanath, R., Shi, J., & Keerthi, S. S. Adam vs. SGD: Closing the generalization gap on image classification. *OPT2021: 13th Annual Workshop on Optimization for Machine Learning*, 2021. Available at: https://opt-ml.org/papers/2021/paper53.pdf (accessed 12.12.2023).

31. Aatila, M., Lachgar, M., & Kartit, A. Comparative study of optimization techniques in deep learning: Application in the ophthalmology field. *The International Conference on Mathematics & Data Science (ICMDS) 2020, Journal of Physics: Conference Series*, 2021, vol. 1743. DOI 10.1088/1742-6596/1743/1/012002.

32. Yaloveha, V., Podorozhniak, A., & Kuchuk, H. Convolutional neural network hyperparameter optimization applied to land cover classification. *Radioelectronic and computer systems*, 2022, no. 1, pp. pp. 115-128. DOI: 10.32620/reks.2022.1.09.

33. Alshehhi, R. Marpu, P. R., Woon, W. L., & Mura, M. D. Simultaneous Extraction Of Roads And Buildings In Remote Sensing Imagery With Convolutional Neural Networks. *ISPRS Journal of Photogrammetry And Remote Sensing*, 2017, vol. 130, pp. 139-149. DOI: 10.1016/j.isprsjprs.2017.05.002.

34. He, K., & Sun, J. Convolutional neural networks at constrained time cost. *IEEE Conference on Computer Vision and Pattern Recognition*, *arXiv preprint arXiv:1412.1710,* 2015, pp. 1-8, DOI: 10.48550/arXiv.1412.1710.

35. Gamal, A., Wibisono, A., Wicaksono, S. B., Abyan, M. A., Hamid, N., Wisesa, H. A., Jatmiko, W., & Ardhianto, R. Automatic LIDAR building segmentation based on DGCNN and euclidean clustering. *Journal of Big Data*, 2020, no. 7, article no. 102. DOI: 10.1186/s40537-020-00374-x.

Received 17.07.2023, Accepted 15.04.2024

СЕМАНТИЧНА СЕГМЕНТАЦІЯ НА ОСНОВІ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ПОБУДОВИ ВИЯВЛЕННЯ НА ЗОБРАЖЕННЯХ ДИСТАНЦІЙНОГО ЗОНДУВАННЯ

Хасмух Корінга, Мірал Патель

Предметом статті є виявлення та семантична сегментація будівлі за зображеннями ДЗЗ високої роздільної здатності. Він витягує зображення будівель із схожих за характеристиками об’єктів доріг і ґрунту. Він використовується для різних застосувань, таких як міське планування, розвиток інфраструктури та боротьба зі стихійними лихами. **Мета** полягає в тому, щоб розробити швидку, точну та автоматичну модель виявлення будівель на основі семантичної сегментації архітектури Unet, яка оптимізована та налаштована з належним налаштуванням гіперпараметрів. **Завдання,** які необхідно вирішити: зібрати набір даних дистанційного зондування, який складається з трьох частин: навчання, перевірка та тестування; застосоване доповнення даних набору даних поїзда за допомогою методів вертикального перевертання, горизонтального перевертання, ротації; далі проходить у двосторонній фільтр для видалення шумів із зображень; оптимізувати модель Unet за допомогою різних методів оптимізатора та налаштованого гіперпараметра за належним значенням вибору, метод порівнюється за відкликанням показників продуктивності, точністю та точністю. **Отримано** наступні результати: модель оцінена за кривою втрат навчання та кривою точності різних оптимізаторів SGD, ASGD, ADAM, ADAMW та RMSProp; він вимірює час навчання, середню точність і середні параметри IOU під час фази навчання; фаза тестування оцінюється за точністю та запам’ятовуванням, метод порівнюється за допомогою результатів візуалізації Unet + інший оптимізатор і запропонований метод порівняно з існуючим методом за загальною метрикою параметра еволюції. **Висновки** Наукова новизна отриманих результатів полягає в наступному: 1) Unet ефективно сегментував побудову ДЗЗ в обмеженій кількості доступних навчальних зразків; 2) функція втрат моделі, яка спостерігається під час вибору гіперпараметрів оптимізатора, швидкості навчання, розміру партії та епох; що створює оптимальну модель для точного та швидкого виділення будівлі зі складного фону; 3) запропонована модель порівнюється з добре відомою моделлю виділення будівлі за загальною метрикою оцінки F1-score.

Ключові слова: глибоке навчання; семантична сегментація; UNet; побудова.

Хасмух Корінга – доцент Державного інженерного коледжу, Раджкот, Гуджарат, Індія.
Мірал Патель – доцент Державного інженерного коледжу, Раджкот, Гуджарат, Індія.

Hasmukh P Koringa – Assistant Professor, Government Engineering College, Rajkot, Gujarat, India,
e-mail: patelhasmukh4u@gmail.com, ORCID: 0000-0002-2521-3972.
Miral J Patel – Assistant Professor, Government Engineering College, Rajkot, Gujarat, India,
e-mail: mjpatelgec@gmail.com, ORCID: 0000-0003-0047-9552.