Hybrid PSOBAGA-Tuned IARUMV2 for Segmentation of High-Resolution Remote Sensing Images

Original Scientific Paper

Babitha Lokula*
Department of ECE, Koneru Lakshmaiah Education Foundation,
Green fields,Vaddeswaram, Guntur District, Andhra Pradesh, India, 522302
babithakodishala@gmail.com

P V V Kishore
Department of ECE, Koneru Lakshmaiah Education Foundation,
Green fields,Vaddeswaram, Guntur District, Andhra Pradesh, India, 522302,
pvvkishore@kluniversity.in

L V Narasimha Prasad
Department of Electronics and Communication Engineering,
Institute of Aeronautical Engineering, Dundigal, Hyderabad-500043, India
lvnprasad@iare.ac.in

*Corresponding author

Abstract *– The complexity of spatial structures as well as fine boundary details are present in the high-resolution remote sensing images (HR-RSI). Due to the abovementioned characteristics, the segmentation of HR-RSI is a typical task. Enhanced U-Net models and traditional techniques are still facing problems like suboptimal feature extraction and low convergence rates for segmenting remote sensing images. In order to overcome the aforementioned challenges, this work suggested an improved U-Net architecture named Inception Attention Residual MobileNetV2 U-Net (IARUMV2). IARUMV2 incorporates inception modules, attention mechanisms, residual links, and a MobileNetV2 encoder to acquire multi-scale contextual information very effectively. For fine-tuning the hyperparameters of IARUMV2, a new hybrid optimization algorithm is introduced, which combines the Particle Swarm Optimization (PSO), Bat Algorithm (BA), and Genetic Algorithm (GA) in order to balance the exploitation, exploration, and convergence speed. The model is tested on two datasets, i.e., Massachusetts Buildings and WHU datasets. Experimental results on the Massachusetts Buildings and WHU Building datasets demonstrate the effectiveness of the proposed PSOBAGA-tuned IARUMV2 framework. The model achieves segmentation accuracies of 96.67% and 98.26%, respectively, with corresponding IoU values of 83.04% and 93.19%. Compared with existing approaches such as U-Net, MATU-Net, TCNet, and SegNet, the proposed method consistently achieves higher accuracy, precision, recall, and F1-score while exhibiting faster convergence and improved hyperparameter optimization efficiency showing its effectiveness for HR-RSI segmentation.*

Keywords*: Image segmentation, Hybrid optimization, Deep learning tuning, IARUMV2 architecture, PSOBAGA*

Received: March 26, 2026; Received in revised form: June 6, 2026; Accepted: June 10, 2026

1. INTRODUCTION

Image segmentation is nothing but dividing the image into subregions, which is the fundamental task in the field of computer vision. Image segmentation is used for object detection and recognition, especially in high-resolution remote sensing images (HR-RSI). For various applications, the efficient segmentation is very useful. The applications include urban planning, disaster monitoring, LULC (Land Use Land Cover) mapping and scene understanding [1]. Still, HR-RSI segmentation is critical task due to changing illumination, blurred object boundaries and complex spectral special characteristics. Traditional techniques such as

pixel and edge-based methodologies revolve around low-level features, preventing their robustness and scalability [2].

In contrast, deep learning based semantic segmentation achieves great status. These techniques use pixel-level labelling alongside improved understanding of contextual information [3]. U-Net-based approaches proved decent performance segmentation of HR-RSI. But still, they are facing the issues such as boundary information loss, less convergence rate and poor contextual modelling [4]. To tackle aforementioned issues, this work advised an improved U-Net architecture, titled as IARUMV2. IARUMV2 integrates Inception modules, attention mechanisms, residual connections, and MobileNetV2 for efficient and proper feature extraction. Furthermore, hyperparameter selection is very important for better deep learning performance. The traditional methods were utilized manual tuning of hyperparameters which is computationally expensive [5]. To automate this process different metaheuristic optimization algorithms, such as GA, PSO and their combinations were utilized [6]. By taking the inspiration from the No Free Lunch statement [7], this research suggested a new optimization strategy called PSOBAGA. Then, PSOBAGA is used to tune the hyperparameters of IARUMV2 to balance exploration and exploitation.

1.1. CONTRIBUTIONS

The main contributions of this research are:

- A new U-Net-based architecture named IARUMV2 is suggested to improve spatial feature retention and enhance semantic segmentation accuracy for HR-RSI.
- The PSOBAGA optimization technique is used to tune the IARUMV2 hyperparameters. It provides less computational cost, quicker convergence and enhanced exploration.
- Experimental evaluation proves that PSOBAGA outperforms traditional optimization methods. It effectively avoids local optima and ensuring robust segmentation performance.

1.2. PRELIMINARIES

Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) are extensively utilized metaheuristic optimization algorithms in engineering applications. GA offers strong global search capability through crossover and mutation but often suffers from slow convergence. While PSO converges faster, it is prone to premature convergence due to reliance on local and global best solutions. To mitigate this issue, the BA is combined with PSO. This combination explains random perturbations and echolocation-inspired adaptive search. This combination also enhances exploration and avoids local optima. Therefore, attain a better balance between exploration and exploitation.

There is a lack of empirical research within the academic field of IT security that tries to measure the amount of human influence [2]. Some existing empirical studies analyze user perception, behavior and attitude towards computer ethics and information security [3-5], as computer security and computer ethics are important components of the management information system [6].

1.2.1. Genetic algorithm:

A genetic algorithm is incorporated in the hybrid PSOBAGA to enhance the global exploration. The diversity tournament selection size=2 is used to help fitter individuals. The uniform crossover with a probability of 0.8 is used to permit the efficient recombination of solutions. The Gaussian mutation with a probability of 0.05 is used to prevent premature convergence. The size of the population is maintained at 30. The algorithm evolves until it reaches the stop condition or until convergence. It provides the exploration-exploitation balance.

1.2.2. Particle Swam Optimization:

Particle Swarm Optimization is basically invented by Eberhart and Kennedy. It is a population-based metaheuristic algorithm. It utilizes the behavior of the bird flocking as well as swam intelligence. In this, particles are exploring the search space. By doing this process it always updates their positions with Pbest (separate best) and Gbest (global best) experiences. It allows steady convergence and also sharing of information. Each particle position and velocity updates are represented with equations (1) and (2):

$$V_{id}(t+1)=w.V_{id}(t)+c_1.rand().(P_{id}(t)-X_{id}(t)) +c_2.rand().(P_{gd}(t)-X_{id}(t)) \quad (1)$$

$$X_{id}(t+1)=X_{id}+V_{id}(t+1) \quad (2)$$

where X_{id} (t)denotes the position of the i^{th} particle in the d^{th} dimension at iteration t, V_{id} (t) denotes the corresponding velocity, P_{id} (t) represents the personal best position of particle i, P_{gd} (t) denotes the global best position of the swarm, wis the inertia weight, c_1 and c_2 are the cognitive and social acceleration coefficients, respectively, and r_1 and r_2 are uniformly distributed random numbers in the interval [0, 1].

Although PSO is computationally efficient, it may suffer from premature convergence and velocity explosion [8]. To address this, a constriction factor proposed by Clerc *et al*. [8] is incorporated, as defined in Equations (3) and (4):

$$V_{id}(t+1)=\lambda\{V_{id}(t)+c_1r_1(P_{id}(t)-X_{id}(t)) +c_2r_2(P_{gd}(t)-X_{id}(t))\} \quad (3)$$

$$\lambda = \frac{2-\zeta-\sqrt{\zeta^2-4\zeta}}{2} \quad (4)$$

Where λ is the constriction factor and $\zeta= c_1+ c_2$. In this work, c_1=c_2=1.5, and the inertia weight w is linearly

decreased from 0.9 to 0.4, ensuring a balanced exploration-exploitation trade-off and improved convergence stability.

1.2.3. Bat Algorithm:

The Bat Algorithm is a meta-heuristic inspired by the echolocation behavior of microbats, utilizing frequency tuning, loudness attenuation, and pulse emission rate control to balance global exploration and local exploitation. BA-inspired strategies enhance PSO by improving exploration and escaping local optima. In the hybrid framework, frequency, velocity, and position are updated using Equations (5)-(7), while loudness and pulse emission rate are updated using Equation (8).

$$fi=fmin+(fmax-fmin).rand() \quad (5)$$

$$v_i^t=v_i^{t-1}+(x_i^{t-1}-gbest).fi \quad (6)$$

$$x_i^t=x_i^{t-1}+V_i^t \quad (7)$$

$$A_i^{t+1}=\alpha.A_i^t, \quad r_i^{t+1}=r_i^0\,[1-\exp(-\gamma t)] \quad (8)$$

In this work, $f_i \in [0,2]$, with loudness and pulse rate adaptively updated using α=0.9 and γ=0.9 Where x_i^t and V_i^t denote the position and velocity of bat i at iteration t; f_i is the frequency controlling the search range; g_{best} is the global best solution; A_i is the loudness, decreasing as convergence progresses; r_i is the pulse emission rate, increasing near optimal solutions; α and γ control the adaptation rates; and $rand()$ is a uniform random number in [0,1]. Effective diversification in early stages and intensification near convergence are enabled by this adaptive mechanism, by which global optimality is improved.

2. RELATED WORK

Due to the effective segmentation ability of U-Net, numerous studies have improved its performance by incorporating attention mechanisms, residual connections, and multi-scale feature extraction. The U-Net architectures with attention mechanisms like U-Net [9], UMV2 [10], MACU-Net and MARes UNet [11], Unified U-Net [12], and DHAU-Net [13] enhance the contextual representation. The multi-scale and lightweight models like MATU-Net [14], TCNet [15], CT-U-Net [16], and altered U-Net variants focus on feature fusion and boundary refinement.

Similarly, this research focuses on emphasizing the crucial role of hyperparameter optimization to enhance the segmentation performance using deep learning techniques. To perform the semantic segmentation, numerous population-based and hybrid optimization techniques were proposed. Yanyan Fan *et al*. proposed a PSO-SSA algorithm [17]. This method achieves high optimization accuracy and improved search speed but still suffers from high computational cost and low early search speed. Han YAN *et al*. designed a GPPSO strategy [18]. It shows an enhanced convergence rate and less computational cost. Still, this approach lacks joint optimization of the CNN (Convolutional Neural Network) framework and hyperparameters. Marcello Serqueira *et al*. implemented a HBRKGA and Hyper-BRKGA algorithm [19]. It attains the improved exploration and exploitation capability while performing hyperparameter tuning. These methods do not optimize architecture and are applicable to only a few applications. Umesh Kumar Lilhore *et al*. developed a Bayesian optimization algorithm [20]. It provides enhanced hybrid architecture for improved feature extraction, optimization of hyperparameters, and high diagnostic performance. It still suffers from high computational cost and practical deployment. Mustafa Acikkar *et al*. proposed a hybrid optimization strategy called PSOGS [21]. This strategy shows enhanced computational efficiency and greater prediction accuracy. But it incurs scalability and data diversity problems. Leila Zahedi *et al*. designed a novel hyperparameter tuning algorithm called HyP-ABC [22]. This algorithm gained an enhanced classification accuracy, and there is a huge reduction in ttraining time. It still suffers with time complexity. Asel Sagingalieva *et al*. developed a hybrid quantum ResNet for car classification [23]. It achieves high convergence efficiency and strong classification accuracy. It's facing issues like complexity and scalability. Hussam N. Fakhouri implemented a novel hybrid success history intelligent optimizer with Gaussian transformation (SHIOGT) [24] for CNN hyperparameter tuning. It attains high optimization performance and a robust exploration-exploitation balance. Still, it's facing sensitivity and overfitting problems. Jian-Yu Li *et al*. proposed surrogate-assisted algorithms [25]. Even though many advances are introduced, the existing hybrid optimizers still suffer from premature convergence, high computational complexity, and imbalance between exploration and exploitation [26].

Deliang Li *et al*, implemented a Transformer-based building extraction framework using SegFormer has demonstrated promising by effectively capturing multiscale contextual information. However, the model's dependence on computationally intensive Transformer operations and large annotated datasets remains a challenge for real-time and resource-efficient remote sensing applications [27]. Zirui Shen *et al*, proposed a DS-SwinUNet enhances local and global feature learning through a dual-scale Transformer block, improving segmentation performance over Swin-UNet. Though, its increased architectural complexity leads to higher computational requirements [28]. Pelagia Drakopoulou *et al*, designed a Mask2Former-based Transformer architectures have shown strong performance in coastal land-form segmentation by effectively capturing complex spatial features, although they require extensive training data and high computational cost [29]. The various state of the art works, its contributions, strengths and limitation are discussed in Table1.

Although recent Transformer-based and hybrid CNN–Transformer architectures have significantly improved segmentation performance through enhanced global

context modeling and multi-scale feature extraction, several limitations remain. Many methods suffer from high computational complexity, inadequate boundary preservation, excessive parameter requirements, and dependence on manually tuned hyperparameters. Furthermore, optimization strategies employed in existing segmentation frameworks often exhibit premature convergence and insufficient exploration–exploitation balance. To address these issues, the suggested PSOBA-GA combines PSO, BA, and GA. This PSOBAGA achieves better convergence and effective hyperparameter tuning in the U-Net-based frameworks.

Table 1. Critical Analysis of Recent Remote Sensing Image Segmentation Methods (2023–2026)

Method	Year	Dataset	Key Contribution	Strengths	Limitations
SegFormer	2023	ADE20K, Cityscapes, Remote Sensing Benchmarks	Introduced a hierarchical Transformer encoder with lightweight MLP decoder for semantic segmentation.	Strong global context modeling, fewer parameters, excellent scalability.	Boundary localization is less precise for small objects; computationally demanding for large images.
DeepLabV3+	2023	WHU Building, Massachusetts Buildings	Combined Atrous Spatial Pyramid Pooling (ASPP) with encoder-decoder architecture.	Captures multi-scale contextual information effectively.	High computational complexity and limited long-range dependency modeling.
Swin-UNet	2024	ISIC, Synapse, Building Extraction Datasets	Integrated Swin Transformer blocks within U-Net architecture.	Excellent feature representation and global dependency learning.	Large memory requirements and slower training.
HRFormer	2024	LoveDA, Potsdam, Vaihingen	Maintained high-resolution representations throughout the network.	Preserves spatial details and object boundaries.	Increased model complexity and computational cost.
Mask2Former	2024	COCO, ADE20K, Remote Sensing Datasets	Unified semantic, instance, and panoptic segmentation using masked attention.	State-of-the-art segmentation accuracy and flexibility.	Heavy architecture requiring extensive computational resources.
SAM-Based Segmentation	2025	SA-1B, Remote Sensing Adaptations	Utilized foundation-model-based promptable segmentation.	Strong zero-shot generalization capability.	Requires prompt engineering and struggles with domain-specific remote sensing objects.
TransUNet	2025	WHU Building, Massachusetts Buildings	Combined CNN feature extraction with Transformer contextual modeling.	Balances local and global feature learning.	Increased parameter count and training complexity.
SegNeXt	2025	LoveDA, Potsdam	Introduced convolutional attention mechanisms for efficient segmentation.	Lightweight architecture with competitive accuracy.	Reduced effectiveness for highly heterogeneous scenes.
UPerNet-Transformer	2026	ISPRS Vaihingen, Potsdam	Feature Pyramid Network integrated with Transformer backbone.	Strong multi-scale representation and semantic understanding.	Computationally expensive and sensitive to hyperparameter selection.
Hybrid CNN–Transformer Networks	2026	WHU Building, Massachusetts Buildings, LoveDA	Combined CNN local feature extraction with Transformer global context learning.	Improved segmentation accuracy and boundary preservation.	Increased training time and hardware requirements.

3. ENHANCED MULTI-OBJECTIVE U-NET

This section gives the summary of the suggested model framework. It also gives the hybrid optimization strategy, which incorporated PSO, BA, and GA. The complete workflow of the suggested method is in Fig. 1. Firstly, total applied images are resized to a 224X224 resolution and normalized to the range [0,1] to confirm data consistency. These pre-processed images are applied to IARUMV2, whereas its hyperparameters are optimized with PSOBAGA. The model gives the segmented images as output. The efficiency of the model is measured with loss metrics as well as accuracy. And also, with the key training parameters like batch size and learning rate repeatedly optimizing with the hybrid strategy.

3.1. OVERVIEW OF IARUMV2

In semantic segmentation, the encoder-decoder frameworks with multi-scale skip connections exhibit strong efficiency. It is done with the process of preserving spatial detail while catching contextual information. The suggested IARUMV2 enhance the normal U-Net with the integration of Inception, Attention, Residual connections and a MobileNetV2 based encoder.

This model is suggested to mitigate limited accuracy, slow convergence, and blurred boundaries challenges in HR-RSI segmentation. MobileNetV2 is a lightweight model, and it extracts the features very effectively. This is done through inverted residual blocks and depthwise separable convolutions. It also reduces computational cost. Skip connections are used to connect the encoder bottlenecks to equivalent decoder stages. It also enables the efficient fusion of low-level and high-level features. Inception and residual blocks improve the multi-scale contextual modeling and stability of training. The attention gates in the decoder highlight the relevant regions as well as perform the refining of object boundaries. Because of the integration of above

mentioned blocks in the proposed method, it achieves better performance. The improved performance parameters are accuracy, efficiency, and robustness. The framework of IARUMV2 model is depicted in Fig. 2.

3.2. PSOBAGA

The suggested PSOBAGA architecture sets the PSO parameters. The PSO parameters are inertia, social and cognitive constants, swarm size, and maximum iterations. It also initializes the hyperparameters of the model. The hyperparameters are the learning rate and batch size. It also sets GA parameters such as crossover and mutation rates. Evaluation of a particle's fitness and updating of Pbest and Gbest have also been done. Repositioning of stagnant particles is performed by BA; otherwise, typical PSO updates are applied. Based on a PSO-derived population, GA evolves through mutation and crossover. It also updates the global best solutions. The PSOBAGA improved the exploration and exploitation balance. It also preserves the population diversity. It also performs strong hyperparameter optimization over a wide search space. Finally, all these provide the performance improvement of HR-RSI semantic segmentation.

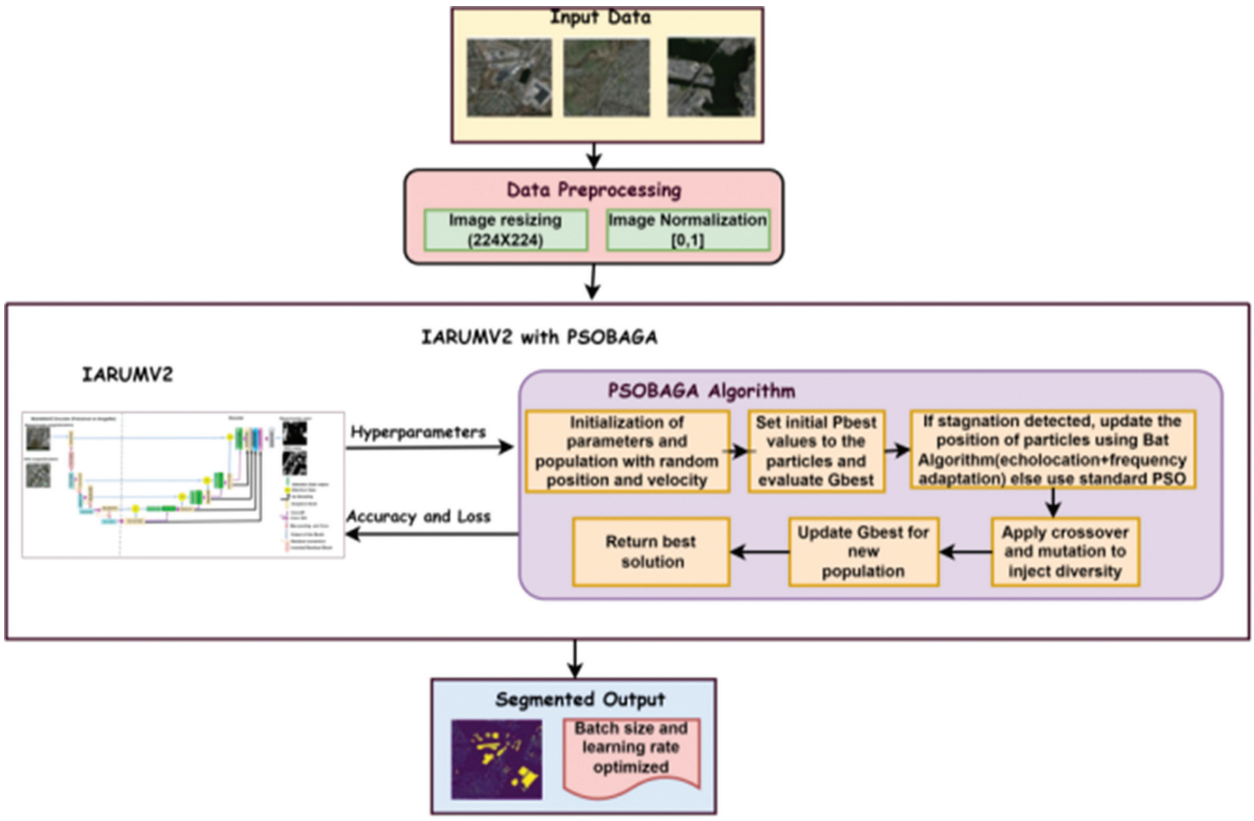


Fig.1. Structural Diagram of the Proposed Model and Optimization Strategy

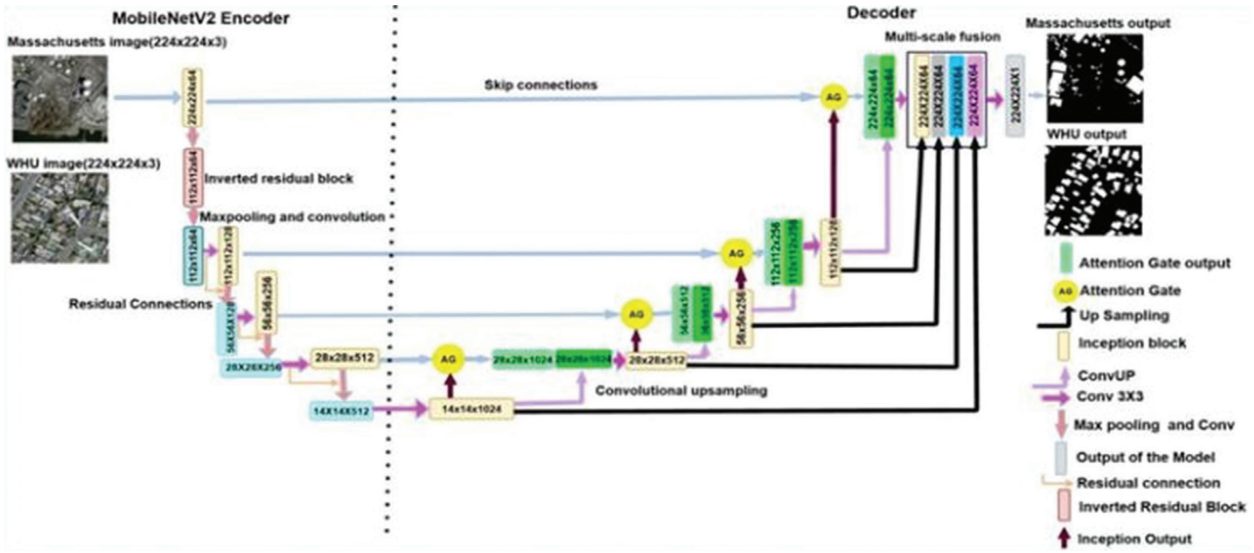


Fig. 2. Structure of IARUMV2

3.2.1. Hybrid Execution Strategy and Iteration Control

The PSOBAGA incorporates the PSO, BA and GA combination in the process of iteration. Through the standard PSO, particles are evolved. If the stagnation is spotted, BA announces adaptive movements and stochastic jumps. Which enhances the escape local optima and diversity. The updated particle fitness is authorized, before performing GA operations. To avoid premature convergence, it refines the population through crossover, mutation and selection. The one iteration counter rules the following process: 1. An occurrence of PSO update for every iteration. 2. BA is triggered upon stagnation. 3. The periodic application of GA. This process confirms effective monitoring of convergence as well as reduced computational redundancy.

3.2.2. Parameter Settings and Control Mechanisms

In the suggested method, the BA parameters defined as shown in Table 2: These parameters are opted for, depending on the convergence behavior presented in previous studies [23-25].

Table 2. PSOBAGA Parameter Settings and Control Mechanisms

Parameter	Value	Purpose
Population Size NP	30	Number of candidate solutions
Inertia Weight w	0.9 → 0.4	Exploration–exploitation balance
Cognitive Coefficient c_1	1.5	Personal learning factor
Social Coefficient c_2	1.5	Global learning factor
Frequency f_i	[0, 2]	BA search range control
Loudness A_i	[0.5, 1.0]	BA exploration intensity
Pulse Rate r_i	[0, 1]	BA local search probability
α	0.9	Loudness update factor
γ	0.9	Pulse rate update factor
Crossover Rate	0.8	GA recombination probability
Mutation Rate	0.05	GA diversity enhancement

In the proposed framework, PSOBAGA is employed to optimize the key training hyperparameters of the IARUMV2 model, including learning rate, batch size, dropout rate, and optimizer-specific parameters. Each candidate solution represents a unique hyperparameter configuration. The fitness of each solution is evaluated on the validation set using segmentation performance metrics. The optimization objective is to maximize validation accuracy while minimizing validation loss. Through the combined exploration and exploitation capabilities of PSO, BA, and GA, the algorithm efficiently searches the hyperparameter space and identifies configurations that promote faster convergence and improved segmentation performance. The Hyperparameters Optimized by PSOBAGA are shown in Table 3.

Table 3. Hyperparameters Optimized by PSOBAGA

Hyperparameter	Range
Learning Rate	0.0001–0.01
Batch Size	8–32
Dropout Rate	0.1–0.5
Weight Decay	0.00001–0.001

Fitness values are recomputed after BA updates to ensure valid state propagation before GA operations, as detailed in Algorithm 1.

Algorithm 1.
Hybrid PSO–BAT–GA Optimization for IARUMV2 Hyperparameters
(Full-Width, BA→Fitness→GA with Elitism).

Require: IARUMV2 network, dataset, PSO parameters (NP, c_1, c_2, w, Max_{iter}), GA parameters ($mutation\ rate$, $crossover\ rate$, $elite_{count}$), Bat Algorithm parameters ($fmin$, $fmax$, A, r, α, γ).

Ensure: Optimal hyperparameters $Gbest$ for IARUMV2

Initialize each particle's position x_j and velocity v_j randomly within bounds
Evaluate initial fitness $f(x_j)$ for all particles
Set personal best $Pbestj$ and global best $Gbest$
for $iter = 1$ to Max_{iter} do
  for $j = 1$ to NP do
    $fitnessj \leftarrow evaluate_{fitness}(x_j)$
    if $fitnessj > Pbestj$ then
      $Pbestj \leftarrow fitnessj$
      $Pbest_{pos}[j] \leftarrow x_j$
    end if
    if $fitnessj > Gbest$ then
      $Gbest \leftarrow fitnessj$
      $Gbest_{pos} \leftarrow x_j$
    end if
  end for
  if $iter > stagnation_{patience}$ and $|Curr_{Gbest} - Gbest| < stagnation_{threshold}$ then
    for $j = 1$ to NP do
      Generate frequency $fj \in [fmin, fmax]$
      $v_j \leftarrow v_j + (x_j - Gbest) \cdot f_j$
      $x_j \leftarrow x_j + v_j$
      if $rand() > r_j$ then
        Perform local random walk around $Gbest$
      end if
      $A_j \leftarrow \alpha \cdot A_j$
      $r_j \leftarrow r_{j0} \cdot [1 - \exp(-\gamma \cdot iter)]$
    end for
    for $j = 1$ to NP do
      $fitnessj \leftarrow evaluate_{fitness}(x_j)$

29: if $fitnessj > Pbestj$ then
30: $Pbestj \leftarrow fitnessj$
31: $Pbest_{pos}[j] \leftarrow x_j$
32: end if
33: if $fitnessj > Gbest$ then
34: $Gbest \leftarrow fitnessj$
35: $Gbest_{pos} \leftarrow x_{j36}$: end if
37: end for
38: else
39: for $j = 1$ to NP do
40: Update v_j and x_j using PSO formulas
41: end for
42: end if
43: if mod($iter$, 5) == 0 then
44: Sort particles by fitness
45: Preserve top $elite_{count}$ particles
46: $new_{particles} \leftarrow [\,]$
47: for $i = 1$ to $(NP - elite_{count})/2$ do
48: $parent1, parent2 \leftarrow select_{parents}$ (non-elite particles)
49: if $rand() < crossover_{rate}$ then
50: $offspring1, offspring2 \leftarrow crossover$ ($parent1, parent2$)
51: else
52: $offspring1, offspring2 \leftarrow parent1, parent2$
55: $offspring1 \leftarrow$ mutate($offspring1$)
56: $offspring2 \leftarrow$ mutate($offspring2$)
57: end if
58: Add $offspring1, offspring2$ to $new_{particles}$
59: end for
60: Replace non-elite particles with $new_{particles}$
61: end if
62: $Gbest_{prev} \leftarrow Gbest$

Let the normalized *fitness* of particle i at iteration t be

$$p_{i,t} = \frac{fi(t)}{\sum_{j=1}^{N} fi(t)}, \qquad H_t = -\sum_{i=1}^{N} p_{i,t} \log p_{i,t} \tag{9}$$

where H_t denotes the population entropy. High entropy implies high population diversity (exploration), whereas low entropy indicates convergence (exploitation). PSOBAGA employs this entropy-driven rule:

$$\begin{cases} PSO\ step: & if\ H_t > H_{th}, \\ BA\ perturbation: & if\ |H_t - H_{t-1}| < \varepsilon, \\ GA\ refinement: & if\ H_t \le H_{th}, \end{cases} \tag{10}$$

where H_{th} and ε are threshold constants that regulates phase switching. Equation (10) creates an adaptive control law. It permits the optimizer to self-regulate its exploration-exploitation balance, depending on the diversity dynamics of the population rather than relying on static iteration schedules.

This formulation presents a new optimization principle as follows: (1) An entropy-controlled adaptive hybridization, in which the PSO core exploits promising regions. (2) BA reintroduces exploratory variance when stagnation is detected. (3) GA refines the solution space through recombination once diversity falls below a critical entropy level.

The resultant optimizer acts as a feedback-controlled dynamical system as designated in equation (11):

$$x_{i,\,t+1} = F_{PSO}(x_{i,\,t}) + \alpha_t F_{BA}(x_{i,t}) + \beta_t F_{GA}(x_{i,\,t}), \tag{11}$$

where $F_{PSO}(\cdot)$, $F_{BA}(\cdot)$, and $F_{GA}(\cdot)$ denote the update operators corresponding to Particle Swarm Optimization, Bat Algorithm, and Genetic Algorithm, respectively. The adaptive coefficients α_t and β_t regulate the contributions of the BA and GA components based on the population entropy H_t. It continuously adjusts the contribution of each component based on population diversity. This establishes a novel optimization strategy that generalizes prior PSO-BA and PSO-GA hybrids into an entropy-based multi-phase model.

3.3. OBJECTIVE MODEL AND OPTIMIZING PARAMETERS

The PSOBAGA improves the accuracy and reduces the loss. The major objective is specified in equation (12).

$$f(x) = x.\,Acc - y.\,Loss \tag{12}$$

where $f(x)$ denotes the fitness function, Acc represents segmentation accuracy, $Loss$ denotes the combined training loss, and xand yare weighting coefficients controlling the relative contributions of accuracy and loss. To give equal importance to both loss and accuracy, these weights are initialized to 0.5.

The loss is computed by integrating *Binary Cross Entropy* (*BCE*) and *Dice Loss* (*dl*), according to equation (13):

$$Loss = Binary\ Croos\ Entropy + Dice\ Loss \tag{13}$$

The accuracy metric is calculated as shown in equation (14):

$$cc = \frac{TP + TN}{TP + FP + FN + TN} \tag{14}$$

The *Dice Loss* (*dl*) is given as shown in equation (15):

$$dl = 1 - \frac{2 * TP}{2 * TP + FP + FN} \tag{15}$$

Combining accuracy and loss as part of the fitness function for hyperparameter optimization in semantic segmentation helps optimize for a balance between high accuracy and minimal loss, which often helps obtain a method with better generalization. The metric combined loss is calculated as loss by adding *BCE* to *dl*, as *dl* works well in capturing objects' overall shape and structure, helping enhance segmentation quality for regions that may be misclassified due to imbalances. *BCE* helps accurately classify pixels at the object boundary, while *Dice Loss* promotes better overlap among predicted and ground truth masks, resulting in a smoother and more coherent segmentation.

4. SEGMENTATION PERFORMANCE AND MODEL BEHAVIOR

4.1. EXPERIMENTAL DATASETS

Dataset 1: The Massachusetts Buildings dataset is available on Kaggle. It contains high-resolution satellite images of 151. These are divided into 137 training images, 10 testing images, and 4 validation images. Images were divided into 224×224 patches, resulting in 474 usable samples after outlier removal. Data augmentation such as Gaussian blur, noise addition, and brightness adjustment was applied to improve generalization.

Dataset 2: The WHU dataset [Kaggle] contains 204 images with masks (512×512), resized to 224×224. Similar augmentation techniques, including Gaussian noise, brightness adjustment, and random transformations, were applied to simulate diverse real-world conditions.

Furthermore, batch normalization, residual learning, attention mechanisms, and cross-validation were utilized to enhance robustness and ensure stable training performance.

4.2. EXPERIMENTAL SET UP

The experiments were conducted on the Kaggle platform using an NVIDIA T4 GPU and an Intel® Xeon® CPU operating at 2.20 GHz. The proposed model was implemented using the TensorFlow and Keras frameworks and trained with the Adam optimizer. For Dataset 1, the model was trained for 350 epochs with a batch size of 16 and an initial learning rate of 0.001, while Dataset 2 employed 3-fold cross-validation with 60 epochs per fold using the same training configuration. The IARUMV2 is trained for 350 epochs with a batch size of 16, and the initial learning rate is 0.001. After the training phase, the model is tested with the validation set images, and the results of sample images of Dataset 1 are depicted in Fig. 3, and the results of sample images of dataset 2 are depicted in Fig. 4.

4.3. QUANTITATIVE ANALYSIS

The performance of the proposed IARUMV2 model is quantitatively evaluated using precision, recall, accuracy, F1-score, and Intersection over Union (IoU). Comparative results on Dataset 1 and Dataset 2 are presented in Tables 4 and 5, respectively.



(a) (b) (c)

Fig. 3. Sample segmentation outcomes on Dataset 1, **a)** Actual image, **b)** True mask, **c)** Segmentation mask



(a) (b) (c)

Fig. 4. Sample segmentation outcomes on Dataset 2, **a)** Actual image, **b)** True mask, **c)** Segmentation mask

Table 4. Analysis of Segmentation Metrics for Multiple Models on Dataset 1

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	IOU (%)
U-Net	93.75	81.13	79.41	80.10	67.54
MATU-Net [14]	93.82	89.17	81.22	85.01	83.22
TCNet [15]	94.85	85.17	86.82	84.29	76.21
U-Net-AstPPD [30]	92.64	-	73.70	-	92.60
Self-attention in reconstruction bias U-Net [31]	-	83.34	86.15	84.72	73.49
Integrating C NN- Transformer and U-Net [32]	92.56	-	-	91.48	-
Segnet [33]	91.56	66.06	82.75	73.47	58.07
Seg-U-Net [34]	92.73	85.74	81.74	83.61	-
Multi Scale Feature Fusion Attention Network [35]	-	81.12	82.96	82.03	69.53
SegFormer[36]	-	82.49	73.44	77.70	66.09
DeepLabV3 [37]		85.4	80.7	83.0	70.9
HRFormer [37]	-	87.3	82.5	84.8	73.6
IARUMV2 (ours)	**96.67**	**92.26**	**89.15**	**90.09**	**83.04**

Note:"–" indicates that the corresponding performance metric was not reported in the original reference paper and is therefore unavailable for comparison in Table 4.

The IARUMV2 architecture demonstrates superior performance over existing segmentation models, achieving consistently higher precision, recall, and accuracy while maintaining competitive F1-score and IoU values.

Notably, IARUMV2 attains accuracies of 96.67% and 98.26% on Dataset 1 and Dataset 2, respectively, after 350 training epochs, confirming its robustness and effectiveness for high-resolution image segmentation.

Table 5. Analysis of Segmentation Metrics for multiple models on Dataset 2

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	IOU (%)
U-Net	97.27	91.86	92.52	92.19	85.51
MATU-Net [14]	97.06	95.05	92.23	93.62	92.14
TCNet [15]	98.80	95.15	95.55	93.95	91.16
Self-attention in reconstruction bias U- Net [31]	-	93.25	95.56	94.40	89.39
Segnet [33]	92.68	92.48	91.55	92.00	85.20
Multi-Scale Feature Fusion Attention Network [35]	-	95.00	95.20	95.10	90.66
SegFormer [36]	-	88.17	88.31	88.23	79.05
DeepLabV3 [37]	-	93.0	92.2	92.6	86.2
HRFormer [37]	-	94.7	94.2	94.4	89.5
Swin-UNet [38]	-	84.0	81.0	82.0	81.0
Mask2Former [38]	-	91.0	88.0	89.0	89.0
SAM based segmentation [38]	-	73.0	70.0	71.0	68.0
IARUMV2 (ours)	**98.26**	**96.12**	**96.40**	**96.10**	**93.19**

Note: "–" indicates that the corresponding performance metric was not reported in the original reference paper and is therefore unavailable for comparison in table 5.

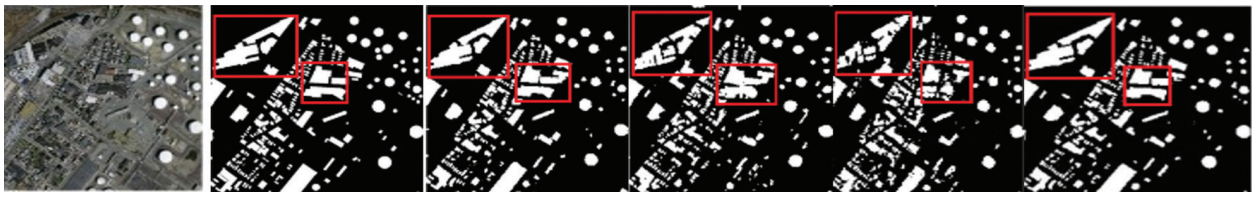
4.4. QUALITATIVE VISUAL COMPARISON OF SEGMENTATION RESULTS

A sample photo of the WHU Building and Massachusetts Building datasets are used to do the qualitative analysis of the models. The prediction masks derived from the DeepLabV3+, SegFormer, Swin-UNet and IARUMV2 models are displayed in Figs. 5 and 6. Significant variations across the figures are shown in red.

An analysis of the figures reveals that DeepLabV3+ successfully predicts most buildings but struggles to accurately delineate their boundaries due to the decoder's limited receptive field. The SegFormer model can identify smaller buildings, but because it lacks sophisticated learning features, it is incomplete when it comes to large buildings. The Swin-UNet model can anticipate the majority of buildings, but it struggles with irregular and dispersed structures because it lacks modules to learn long-range relationships. On the other hand, the introduced IARUMV2 is able to detect small objects along with all the buildings, showing high performance. Even at the edges the segmentation results remain sharp and accurate. With respect to this aspect, the proposed model shows superior performance compared to other approaches.

4.1. ABLATION STUDY

An ablation study is performed to evaluate the contribution of individual modules within the proposed IARUMV2 architecture, with results summarized in Tables 7 and 8 for Dataset 1 and Dataset 2, respectively.



(a) (b) (c) (d) (e) (f)

Fig. 5. Comparative Analysis of IARUMV2 and State-of-the-Art Models in Massachusetts Building Segmentation. **(a)** Actual Image, **(b)** True mask **(c)** DeepLabV3+ **(d)** SegFormer **(e)** Swin-UNet **(f)** IARUMV2



(a) (b) (c) (d) (e) (f)

Fig. 6. Comparative Analysis of IARUMV2 and State-of-the-Art Models in WHU Building Segmentation. **a)** Actual Image, **b)** True mask **c)**, DeepLabV3+ **d)**, SegFormer **e)**, Swin-UNet **f)** IARUMV2

Table 7. Segmentation Performance Comparison based on Module Combinations for Dataset 1

Method	Accuracy (%)	Loss (%)	Precision (%)	Recall (%)	F1- score (%)	IOU (%)
U-Net	93.73	19.88	81.13	79.41	80.10	67.54
Unet+MobileNetV2	93.86	19.87	82.34	79.95	80.40	67.59
Attention gate+UMV2	93.69	20.22	80.42	79.74	79.82	67.26
Attention gate+UMV2+SE	94.40	16.78	83.56	83.25	83.35	72.22
Inception+UMV2+SE	95.51	11.23	89.94	86.99	88.43	79.53
IARUMV 2	**96.67**	**9.71**	**91.36**	**88.85**	**90.08**	**82.14**
IARUMV2+ SE	93.23	22.31	78.80	78.00	78.11	64.96

Table 8. Segmentation Performance Comparison based on Module Combinations for Dataset 2

Method	Accuracy (%)	Loss (%)	Precision (%)	Recall (%)	F1-score (%)	IOU (%)
U-Net	91.86	11.89	86.31	89.20	87.65	78.30
Unet+MobileNetV2	91.90	11.76	86.45	89.91	87.71	78.38
Attention gate+UMV2	91.93	11.68	85.93	89.93	87.78	78.49
Attention gate+UMV2+SE	92.40	11.01	88.40	88.98	88.62	79.78
Inception+UMV2+SE	93.57	8.68	91.04	90.89	90.93	83.59
IARUMV 2	**98.26± 0.04**	**4.14± 0.002**	**96.11± 0.05**	**96.10± 0.04**	**96.10± 0.04**	**93.39± 0.11**
IARUMV2+ SE	92.54	10.77	88.49	88.93	88.93	79.85

4.6. OPTIMIZATION ALGORITHM COMPARISON

To evaluate hyperparameter optimization effectiveness, GA, PSO, PSO+BA, PSO+GA, and the proposed PSOBAGA are applied to the IARUMV2 model. The objective of PSOBAGA is not to reduce the computational complexity of model training. Instead, it identifies effective hyperparameter configurations that enable the IARUMV2 model to achieve higher validation accuracy and lower validation loss within fewer training epochs. Consequently, the optimized model reaches stable convergence more rapidly than configurations selected through manual tuning. Segmentation performance is assessed using validation accuracy and loss, as reported in Tables 9 and 10, while overall accuracy comparisons are summarized in Table 11. The convergence trends in Figs 7 and 8 further demonstrate that PSOBAGA stabilizes within approximately 50 epochs, whereas competing optimizers require 100-150 epochs, confirming improved computational efficiency without additional architectural overhead. To further interpret the influence of PSO parameter tuning, heatmap visualizations were included to show the variation in validation accuracy and loss across different parameter combinations, thereby complementing the numerical comparison and convergence curves as shown in Fig. 9.

Table 9. Comparison of validation accuracy for various combinations of PSO and GA on IARUMV2

No. of Epochs	GA	PSO	PSO+BA	PSO+GA	PSOBAGA
30	83.00	85.00	86.20	93.96	96.20
40	85.00	88.00	89.67	94.56	96.42
50	88.50	93.00	93.85	95.82	96.83
60	93.50	94.00	94.50	95.97	97.12
100	93.75	93.50	94.72	96.12	97.20
150	93.83	93.72	94.78	96.12	97.20
200	93.85	94.10	94.84	96.20	97.20
250	93.85	94.15	94.88	96.20	97.20
300	93.85	94.15	94.92	96.20	97.20
350	93.85	94.15	94.94	96.20	97.20

Table 10. Comparison of validation loss for various combinations of PSO and GA on IARUMV2

No. of Epochs	GA	PSO	PSO+BA	PSO+GA	PSOBAGA
30	3.98	3.11	2.97	2.91	1.65
40	3.85	2.95	2.33	2.72	1.32
50	2.74	2.37	1.70	2.20	1.23
60	2.92	1.80	1.64	1.56	1.14
100	2.81	1.64	1.56	1.47	1.07
150	2.74	1.52	1.44	1.35	1.04
200	2.58	1.26	1.24	1.17	1.04
250	2.37	1.17	1.11	1.08	1.04
300	2.26	1.17	1.11	1.08	1.04
350	2.24	1.17	1.11	1.08	1.04

Table 11. Comparison of accuracy values for IARUMV2 without PSOBAGA and after implementing PSOBAGA

No. of Epochs	IARUMV2 without PSO-BAGA Optimization (%)	IARUMV2 with PSOBAGA Optimization (%)
30	86.92	96.22
50	90.14	96.83
60	91.46	97.12
100	92.52	97.12
150	93.70	97.20
200	94.68	97.20
250	95.27	97.20
300	96.76	97.20
350	96.78	97.20

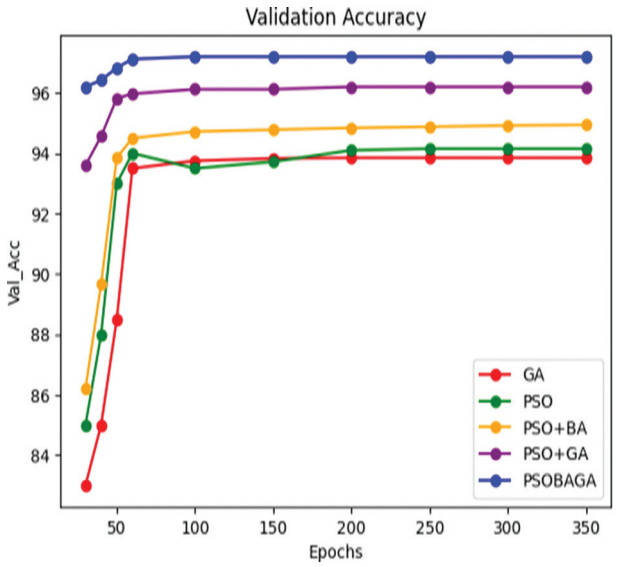


Fig. 7. Comparison of validation accuracy for various combinations of PSO

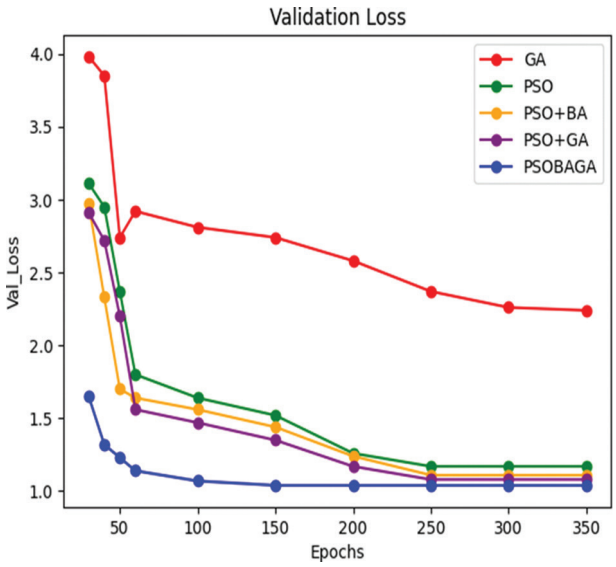


Fig. 8. Comparison of validation loss for various combinations of PSO

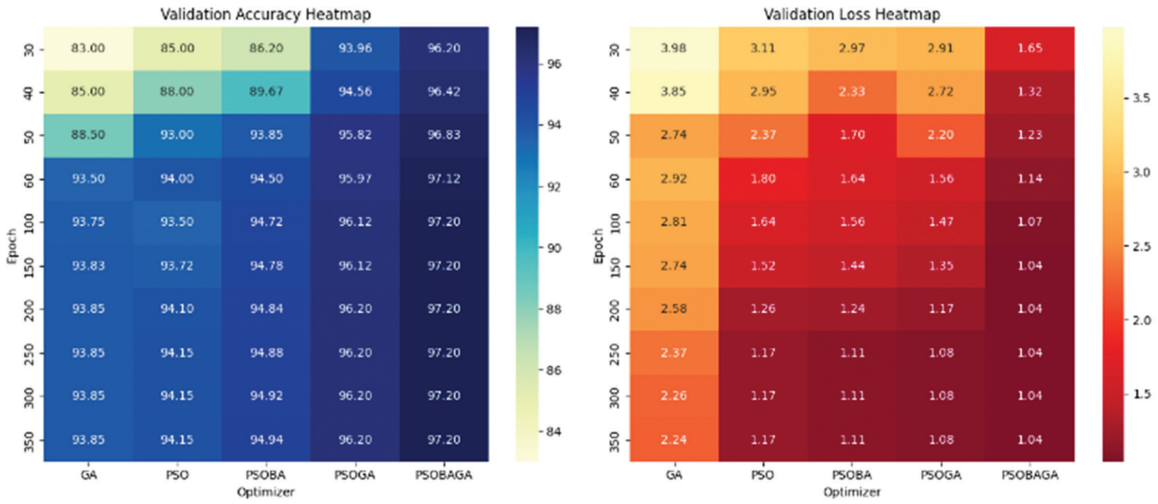


Fig. 9. Validation accuracy and loss heatmap for GA, PSO, PSOBA, PSOGA, and PSOBAGA

4.7. STATISTICAL ROBUSTNESS ANALYSIS

Since PSOBAGA incorporates stochastic optimization mechanisms through Particle Swarm Optimization, Genetic Algorithm, and Bat Algorithm operators, the optimization process may exhibit run-to-run variability. To evaluate the robustness and stability of the proposed optimization framework, the experiments were repeated 30 independent times using different random initializations. The mean and standard deviation of validation accuracy and loss were computed and are reported in Tables 12 and 13. Furthermore, a Wilcoxon signed-rank statistical test was conducted to assess the significance of the observed improvements. The obtained p-values confirm that the performance gains achieved by PSOBAGA are statistically significant (< 0.05), demonstrating the reliability and reproducibility of the proposed optimization strategy shown in Table 14.

Table 12. Statistical Analysis of Validation Accuracy after 30 Independent Runs

Optimizer	Main accuracy	Std. Dev. (%)	Best (%))	Worst (%)
GA	93.72	0.41	94.35	92.96
PSO	94.08	0.36	94.68	93.42
PSO+BA	94.87	0.31	95.29	94.18
PSO+GA	96.15	0.22	96.48	95.72
PSOBAGA	97.18	0.11	97.42	96.95

Table 13. Statistical Analysis of Validation Loss after 30 Independent Runs

Optimizer	Main Loss	Std. Dev. (%)	Best (%)	Worst (%)
GA	2.31	0.18	2.05	2.72
PSO	1.23	0.09	1.11	1.39
PSO+BA	1.15	0.07	1.04	1.28
PSO+GA	1.09	0.05	1.01	1.18
PSOBAGA	1.04	0.02	1.01	1.08

Table 14. Wilcoxon Signed-Rank Statistical Test

Comparison	P-value
PSOBAGA vs GA	0.0008
PSOBAGA vs PSO	0.0016
PSOBAGA vs PSO+BA	0.0031
PSOBAGA vs PSO+GA	0.0094

4.8. Model Compilation

For Dataset 1, preprocessing and data augmentation using Gaussian blur, noise injection, and brightness enhancement are applied, followed by resizing images from 1500×1500 to 224×224. The pre-split Kaggle dataset is trained using the proposed IARUMV2 architecture for 350 epochs with a batch size of 16 and learning rate of 0.001.

For Dataset 2, images of size 512×512 are resized to 224×224 and normalized to [0-1], with an 80:20 train-validation/test split and 3-fold cross-validation, where each fold is trained for 60 epochs. As suggested in [25], feature channels are doubled at each down-sampling stage until a 14×14 resolution is reached, resulting in 4096 channels at the deepest layer.

The decoder mirrors this structure using inception blocks, up-sampling layers, and attention gates, with

concatenate features fed into the prediction module. Table 15 reports the effective training time required to achieve stable convergence. Although PSOBAGA introduces additional computations during hyperparameter optimization, it identifies effective hyperparameter configurations at an early stage, allowing the model to converge more rapidly. Consequently, PSOBAGA achieves stable performance within approximately 50–60 epochs, whereas competing optimization strategies require 100–350 epochs to attain comparable performance. Therefore, the reduction in training time shown in Table 15 reflects improved convergence efficiency rather than a reduction in optimization complexity.

To further evaluate the computational efficiency of the proposed PSOBAGA-IARUMV2 framework, additional metrics including parameter count, FLOPs, GPU memory consumption, and inference time were analyzed. As shown in Table 16, the proposed model achieves a favorable balance between segmentation performance and computational cost.

Table 15. Comparison of effective training time required to achieve stable convergence for IARUMV2 with and without PSOBAGA optimization

No. of Epochs	Time without PSOBAGA (s)	Time with PSOBAGA (s)
30	17.7	7.1
60	35.2	14.3
90	52.5	21.4
120	69.6	28.5
150	86.6	35.6
210	120.5	49.5
270	154.3	63.2
350	183.2	75.3

Table 16. Computational Efficiency Comparison of Segmentation Models

Model	Parameters (M)	FLOPs (G)	GPU Memory (GB)	Inference Time (ms)
U-Net	31.0	55.2	4.8	28
SegNet	29.4	48.5	4.3	25
Attention U-Net	34.6	62.8	5.4	33
CT-UNet	36.2	67.4	5.8	36
IARUMV2	24.8	42.1	3.9	22
PSOBAGA-IARUMV2	23.8	41.1	3.9	21

4.9. REPRODUCIBILITY AND IMPLEMENTATION DETAILS

To ensure reproducibility, all experiments were conducted using a fixed random seed of 42 for TensorFlow, NumPy, and Python random libraries. The PSOBAGA optimizer employed a population size of 30 particles, maximum iterations of 350, crossover probability of 0.8, mutation probability of 0.05, inertia weight decreasing linearly from 0.9 to 0.4, and acceleration coefficients $c_1 = c_2 = 1.5$.

The hyperparameter search space included learning rate $\in [10^{-5}, 10^{-2}]$, batch size $\in \{8,16,32,64\}$, dropout rate $\in [0.1,0.5]$, and weight decay $\in [10^{-6},10^{-3}]$. Particle positions were initialized uniformly within these ranges, while velocities were initialized randomly within ±10% of the corresponding parameter bounds.

Optimization terminated when either the maximum iteration limit was reached or when the improvement in global best fitness remained below 10^{-4} for 20 consecutive iterations. Training stability was enhanced through batch normalization, gradient clipping with a maximum norm of 1.0, adaptive learning-rate scheduling, and checkpoint-based model recovery.

The computational complexity of PSOBAGA is approximately $O(NTD)$, where N denotes the population size, T denotes the number of optimization iterations, and D denotes the number of hyperparameters being optimized. The proposed framework maintained practical training times while providing improved convergence and segmentation accuracy.

5. CONCLUSION

To improve model accuracy, achieve fast convergence, and enhance the prediction capability in HR-RSI, this work suggested a novel IARUMV2 model. IARUMV2 incorporates MobileNetV2, Inception modules, attention mechanisms, and residual connections into the standard U-Net model. These modules enable the effective and efficient extraction of both contextual and spatial features. The novel algorithm PSOBAGA is applied to the suggested model to fine-tune the hyperparameters. The PSOBAGA combines the PSO, GA, and BA. In PSOBAGA, when stagnation is detected in PSO after a predefined number of iterations, the BA dynamically updates the positions of the particle. Due to this, enhancing the exploratory behavior of the swarm and preventing local optima entrapment. Experimental results show that the integration of IARUMV2 with PSOBAGA suggestively outper-forms conventional models and optimization strategies. The suggested model achieves high segmentation accuracy and convergence speed, indicating its effectiveness in remote sensing image analysis. Although the proposed framework demonstrated strong performance on two benchmark building-segmentation datasets, future work will focus on evaluating its generalization capability across additional remote-sensing datasets and diverse land-use categories, including agricultural, residential, industrial, and forest regions.

Future research will also focus on validating the model on large-scale multi-class remote sensing datasets, cross-domain benchmark datasets, and noisy real-world scenarios to further assess the robustness, scalability and generalization capability of the proposed IARUMV2-PSOBAGA framework.

Even though the proposed framework demonstrates superior performance compared with several established remote-sensing segmentation methods, future work will include comprehensive comparisons with recent transformer-based architectures, such as SegFormer, Swin-UNet, HRFormer, Mask2Former, and foundation-model-based approaches. Such investigations will provide additional insights into the trade-offs between segmentation accuracy, computational complexity, and model efficiency.

CONFLICTS OF INTEREST

The authors confirmed that they have no personal or financial relationships to influence the research explained in this paper.

FUNDING STATEMENT

The authors indicating that there is no external funding or financial support was taken for this manuscript.

ACKNOWLEDGMENTS

Babitha Lokula was answerable for the study and designing the model. The data collection, implementation of the model experimentation, was supported by Kishore P V V. L. V. Narasimha Prasad guided the overall planning and direction.

DATA AVAILABILITY AND ACCESS STATEMENTS:

https://www.kaggle.com/datasets/balraj98/massachuse tts-buildings-dataset

http://gpcv.whu.edu.cn/data/building_dataset.html

are publicly available datasets used in this study for performance evaluation. Source code or trained models be publicly released through a GitHub repository upon publication of this manuscript to facilitate producibility and future research.

DECLARATION OF THE USE OF AI TOOLS

The authors declare that no artificial intelligence (AI) tools were used in the preparation, writing, analysis, or editing of this manuscript

REFERENCES

[1] R. Datla *et al*. "Learning scene-vectors for remote sensing image scene classification", Neurocomputing, Vol. 587, 2024, p. 127679. doi: 10.1016/j.neucom.2024.127679

[2] I. Ulku, E. Akagündüz, "A survey on deep learning-based architectures for semantic segmentation on 2D images", Applied Artificial Intelligence, Vol. 36, No. 1, 2022, p. 2032924. doi: 10.1080/08839514.2022.2032924

[3] T. Zhou, S. Ruan, S. Canu, "A review: Deep learning for medical image segmentation using multimodality fusion", Array, Vol. 3, 2019, p. 100004. doi: 10.1016/j.array.2019.100004

[4] H. Zhang *et al*. "Automated delineation of agricultural field boundaries from sentinel-2 images using recurrent residual U-net", International Journal of Applied Earth Observation and Geoinformation, Vol. 105, 2021, p. 102557. doi: 10.1016/j.jag.2021.102557

[5] B. Provencher, A. Badran, J. Kroll, M. Marsh, "Hyperparameter tuning for deep learning semantic image segmentation of micro computed tomography scanned fiber-reinforced composites", Tomography of Materials and Structures, Vol. 5, 2024, p. 100032. doi: 10.1016/j.tmater.2024.100032

[6] T. Yu, H. Zhu, "Hyper-parameter optimization: A review of algorithms and applications", arXiv:2003.05689, 2020.

[7] D. H. Wolpert, W. G. Macready, "No free lunch theorems for optimization", IEEE Transactions on Evolutionary Computation, Vol. 1, No. 1, 1997, pp. 67-82. doi: 10.1109/4235.585893

[8] M. Clerc, J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space", IEEE Transactions on Evolutionary Computation, Vol. 6, No. 1, 2002, pp. 58-73. doi: 10.1109/4235.985692

[9] O. Oktay *et al*. "Attention u-net: Learning where to look for the pancreas", arXiv:1804.03999, 2018.

[10] B. Lokula, P. V. V. Kishore, LV Narasimha Prasad, "UMV2: deep learning model for semantic segmentation of satellite imagery", International Journal of Computational and Experimental Science and Engineering, 2025.

[11] R. Li, S. Zheng, C. Duan, J. Su, C. Zhang, "Multistage attention resu-net for semantic segmentation of fine-resolution remote sensing images", IEEE Geoscience and Remote Sensing Letters, Vol. 19, 2021, pp. 1-5. doi: 10.1109/LGRS.2021.3063381

[12] P. Lokula, Babitha, P. V. V. Kishore, LV Narasimha Prasad, "A unified deep learning framework for segmentation in remote sensing imagery", Discover Computing, Vol. 28, No. 1, 2025, p. 338. doi: 10.1007/s10791-025-09774-0

[13] J. Lei, X. Liu, H. Yang, Z. Zeng, J. Feng, "Dual hybrid attention mechanism-based U-net for building segmentation in remote sensing images", Applied Sciences, Vol. 14, No. 3, 2024, p. 1293. doi: 10.3390/app14031293

[14] Y. Wang, S. Wang, J. Yuan, A. Dou, Z. Gu, "A deep learning method for building extraction from remote sensing images by fuzzing local and global features", Journal of Sensors, Vol. 2024, No. 1, 2024, p. 5575787. doi: 10.1155/2024/5575787

[15] X. Xiang, W. Gong, S. Li, J. Chen, T. Ren, "TCNet: Multiscale fusion of transformer and CNN for semantic segmentation of remote sensing images", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 17, 2024, pp. 3123-3136. doi: 10.1109/JSTARS.2024.3349625

[16] S. Liu, H. Ye, K. Jin, H. Cheng, "Ct-Unet: Context-transfer-Unet for building segmentation in remote sensing images", Neural Processing Letters, Vol. 53, 2021, pp. 4257-4277. doi: 10.1007/s11063-021-10592-w

[17] Y. Fan *et al*. "A hybrid sparrow search algorithm of the hyperparameter optimization in deep learning", Mathematics, Vol. 10, No. 16, 2022, p. 3019. doi: 10.3390/math10163019

[18] H. Yan, C. Zhong, Y. Wu, L. Zhang, W. Lu, "A hybrid-model optimization algorithm based on the gaussian process and particle swarm optimization", Science, Vol. 4, No. 5, 2023, p. 690. doi: 10.1631/FITEE.2200515

[19] H. Yan, C. Zhong, Y. Wu, L. Zhang, W. Lu, "A hybrid-model optimization algorithm based on the gaussian process and particle swarm optimization for mixed-variable CNN hyperparameter automatic search", Frontiers of Information Technology & Electronic Engineering, Vol. 24, No. 11, 2023, pp. 1557-1573.

[20] M. Serqueira, P. González, E. Bezerra, "A population-based hybrid approach to hyperparameter optimization for neural networks", arXiv:2011.11062, 2020.

[21] U. K. Lilhore *et al*. "A precise model for skin cancer diagnosis using hybrid U-net and improved mobilenet-v3 with hyperparameters optimization", Scientific Reports, Vol. 14, No. 1, 2024, p. 4299. doi: 10.1038/s41598-024-54212-8

[22] M. Açıkkar, Y. Altunkol, "A novel hybrid PSO-and GS-based hyperparameter optimization algorithm for support vector regression", Neural Computing and Applications, Vol. 35, No. 27, 2023, pp. 19961-19977. doi: 10.1007/s00521-023-08805-5

[23] L. Zahedi, F. G. Mohammadi, M. H. Amini, "Hyp-ABC: A novel automated hyper-parameter tuning algorithm using evolutionary optimization", arXiv:2109.05319, 2021.

[24] A. Sagingalieva *et al*. "Hybrid quantum resnet for car classification and its hyperparameter optimization", Quantum Machine Intelligence, Vol. 5, No. 2, 2023, p. 38. doi: 10.1007/s42484-023-00123-2

[25] H. N. Fakhouri, S. Alawadi, F. M. Awaysheh, F. Hamad, "Novel hybrid success history intelligent optimizer with gaussian transformation: Application in CNN hyperparameter tuning", Cluster Computing, Vol. 27, No. 3, 2024, pp. 3717-3739. doi: 10.1007/s10586-023-04161-0

[26] J.-Y. Li *et al*. "Surrogate-assisted hybrid-model estimation of distribution algorithm for mixed-variable hyperparameters optimization in convolutional neural networks", IEEE Transactions on Neural Networks and Learning Systems, Vol. 34, No. 5, 2021, pp. 2338-2352. doi: 10.1109/TNNLS.2021.3106399

[27] S. Mehroliya, S. Tomar, A. Arya, A. Verma, "A novel hybrid GA-PSO algorithm-based optimization of transmission and expansion planning", SN Computer Science, Vol. 20, No. 12, 2025, p. e0338104. doi: 10.1007/s42979-023-02188-z

[28] Z. Shen, W. Liu, S. Xu, "DS-SwinUNet: Redesigning skip connection with double scale attention for land cover semantic segmentation", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 18, 2024, pp. 4382-4395. doi: 10.1109/JSTARS.2024.3496725

[29] P. Drakopoulou *et al*. "Integrating AI for In-Depth Segmentation of Coastal Environments in Remote Sensing Imagery", Remote Sensing, Vol. 18, No. 2, 2026, p. 325. doi: 10.3390/rs18020325

[30] A. Ramalingam, V. Srivastava, S. V. George, S. Alagala, M. L. Manickam, "Building rooftop extraction from aerial imagery using low complexity U-Net variant models", Journal of Spatial Science, Vol. 69, No. 3, 2024, pp. 773-800.

[31] Z. Chen *et al*. "Self-attention in reconstruction bias U-net for semantic segmentation of building rooftops in optical remote sensing images", Remote Sensing, Vol. 13, No. 13, 2021, p. 2524. doi: 10.3390/rs13132524

[32] R. Wang, M. Cai, Z. Xia, Z. Zhou, "Remote sensing image road segmentation method integrating CNN-transformer and U-Net", IEEE Access, Vol. 11, 2023, pp. 144446-144455. doi: 10.1109/ACCESS.2023.3344797

[33] V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 12, 2017, pp. 2481-2495. doi: 10.1109/TPAMI.2016.2644615

[34] A. Abdollahi, B. Pradhan, A. M. Alamri, "An ensemble architecture of deep convolutional SegNet and Unet networks for building semantic segmentation fromhigh-resolution aerial images", Geocarto International, Vol. 37, No. 12, 2022, pp. 3355-3370.

[35] J. Liu, H. Gu, Z. Li, H. Chen, H. Chen, "Multi-scale feature fusion attention network for building extraction in remote sensing images", Electronics, Vol. 13, No. 5, 2024, p. 923. doi: 10.3390/electronics13050923

[36] M. Li *et al*. "Method of Building Detection in Optical Remote Sensing Images Based on SegFormer", Sensors, Vol. 23, No. 3, 2023, p. 1258. doi: 10.3390/s23031258

[37] P. Zhu *et al*. "MSHFormer: A multiscale hybrid transformer network with boundary enhancement for VHR remote sensing image building extraction", IEEE Transactions on Geoscience and Remote Sensing, Vol. 63, 2025. doi: 10.1109/TGRS.2025.3545919

[38] T. Revathi, R. Niranchana, "Hybrid Transformer-Based Framework for Building Detection in High-Resolution Satellite Imagery: Integrating SAM, PVT, and Mask2Former", Procedia Computer Science, Vol. 270, 2025, pp. 2036-2046. doi: 10.1016/j.procs.2025.09.324