An Ensemble of Deep Learning Models Using a Genetic Algorithm for Improving Bird Image Segmentation

B. S. Chandrashekar
Department of Studies in Computer Science, University of Mysore, Mysuru, Karnataka, India
coderchandru@gmail.com (corresponding author)

H. S. Nagendraswamy
Department of Studies in Computer Science, University of Mysore, Mysuru, Karnataka, India
hsnswamy@compsci.uni-mysore.ac.in

M. P. Pavan Kumar
Information Science and Engineering Department, Jawaharlal Nehru National College of Engineering, Shivamogga, Karnataka, India
pavankumarmp@jnnce.ac.in

Received: 7 January 2026 | Revised: 10 February 2026 and 24 February 2026 | Accepted: 6 March 2026

 DOI: https://doi.org/10.48084/etasr.17394

ABSTRACT

Bird image segmentation and classification from natural-scene images captured by a device are important for ecological research focused on wildlife conservation and monitoring. Manually segmenting a bird image from a complex background is tedious, time-consuming, and process-intensive. Therefore, an automatic segmentation of bird images from the complex scene image, exploiting image processing techniques, has been extensively studied. Although many efforts have been made to propose segmentation techniques in a different context, the techniques cannot be generalized and may not perform well in some cases. Besides, the task of segmentation is very challenging due to occlusion, morphology, and lighting conditions. Therefore, there is a scope to fine-tune and enhance the segmentation ability of algorithms. The present study proposes a novel method of ensembling deep learning models using Genetic Algorithm (GA) for improving bird image segmentation in a complex background. The proposed methodology employs the segmentation results obtained by the individual deep learning models at the pixel level for ensembling. Five popular deep learning models, namely Unet, PSP-net, Link-net, Feature Pyramid Network (FPN), and Deeplabv3+, are used in the study. The CUB-200-2011 benchmark dataset was employed to conduct experiments, and the efficacy of the proposed approach was evaluated using the Intersection over Union (IoU) metric. The segmentation results obtained were very promising, achieving the highest IoU compared to contemporary segmentation approaches.

Keywords-segmentation; birds; genetic algorithm; chromosome; pixel-wise combination

I. INTRODUCTION

An image pixel segmentation method is used to precisely select the foreground pixels (region of interest) and differentiate them from background pixels. In bird image segmentation, only bird pixels are selected among all the pixels of the image. In general, segmentation of the region of interest is an important task in computer vision applications. In particular, the efficiency and efficacy of any classification technique can be enhanced when an image with only the region of interest is submitted as an input.

Bird image segmentation plays a significant role in bird species classification, which is used for conservation of endangered and extinct bird species, monitoring biodiversity, wildlife management, and protection of the ecosystem. Since manual segmentation of bird images from complex backgrounds is process-intensive, there is a need for automatic segmentation techniques. Also, technology-enabled solutions to ecosystem protection and management involve smart cameras, robotics, and drone technology. These technologies require automation of several tasks, and bird image segmentation is one such task. The literature on segmentation techniques in general and bird image segmentation in particular reveals that although several approaches have been proposed for the segmentation of regions of interest in various contexts, there is still scope to address the limitations of the proposed

approaches. This gap motivated the exploration of combining the results obtained from popular deep learning models in a novel way to enhance overall performance.

Genetic Algorithms (GAs) are often used in solving complex optimization problems in many areas of image-level analysis. Image segmentation is considered an optimization problem, where effective solutions are obtained using parameter tuning and optimization of segmentation variables. The proposed study uses deep learning models, as developments in machine learning have shown noticeable improvements in the results when analyzed with traditional machine learning-based approaches.

The key contributions in the present study are:

- The present study summarizes the achievement of individual deep models for the segmentation of bird images from a complex background. The benchmark bird image dataset used by contemporary researchers in this context was utilized for assessing the segmentation performance of the models.
- Introduction of a novel method of combining the segmentation output obtained from deep learning models to enhance the segmentation performance by using the GA.
- The study employs different joining rules such as Min, Max, Mean, and Median.
- The final segmentation approach is based on the best combination of deep learning models and a combining rule.
- The proposed ensemble approach is applied to the benchmark bird image dataset, and the results are compared with those of other related studies.

II. LITERATURE REVIEW

Authors in [1] proposed optimal segmentation of image datasets using a GA with different colour spaces. They developed an automated algorithm for selecting segmentation parameters and techniques. The study was conducted on the LSM dataset, where the Probabilistic Rand Index (PRI) metric was used to compare different techniques. The average PRI across all images was 0.9378 for Otsu, 0.9701 for PCA, and 0.9925 for GA11. Authors in [2] proposed an ensemble of GA-based explainer with automated image segmentation for melanoma detection. First, the sparsity of the chromosome was identified heuristically, followed by the application of multiple GAs in parallel. Different GAs, based on the number of super pixels in the input image, resulted in chromosomes of varying lengths. All GAs used consensus and seniority voting to obtain the final segmentation result. The melanoma detection dataset was used for evaluation, achieving an explanation error of 0.18 for the proposed census voting approach. Authors in [3] employed a GA–based approach for incorporating priors into medical image segmentation. The approach integrates learned information, such as shape, properties, and positional priors, into a single framework for automated segmentation. The methodology was evaluated on prostate segmentation using MRI and pelvic computed tomography images. The Dice Similarity Coefficient (DSC) values were 0.45 for 2D segmentation and 0.69 for 3D segmentation. The limitations of these methodologies include high computational cost and no guarantee of optimality.

Authors in [4] proposed a brain stroke segmentation method for MRI images using a contextual transformer. The approach employs spatial feature mining coupled with global anatomical processing, further enhanced through feature fusion for segmentation. The core idea is to "generalize out and focus in," meaning that the model generalizes external anatomical structures while focusing on stroke-specific features. Experiments were conducted on the ATLAS v2.0 dataset. The proposed model achieved a Dice score of 0.731, outperforming state-of-the-art methods. However, the methodology is complex, not truly global in context modeling, and introduces hybrid architectural overhead. Authors in [5] developed a combined nondeterministic and deterministic segmentation model for a multi-scene medical image study. A learnable Gaussian Mixture Model (GMM) was designed for nondeterministic prediction, enabling efficient uncertainty modeling and the generation of diverse segmentation masks. The weighted mean of the GMM outputs was used as supplementary input for deterministic models, improving their operation. Additionally, a Monte Carlo technique was applied to approximate the KL divergence of the GMM. The proposed model achieved a Dice score of 92.77% on the LIDC-IDRI dataset. Limitations of the approach include sensitivity to initialization, difficulty in selecting the number of mixture components, the curse of dimensionality, and high computational complexity.

In [6], a sparse diffusion technique for multi-annotator medical image segmentation was proposed. A novel sparse diffusion framework was introduced to effectively handle multi-rater segmentation. A background-changing spatial sparse module improves inference by focusing computational resources on critical image regions. The introduced method achieved a Soft Dice score of 97.02% on the QUBIQ kidney dataset. Limitations of the model include a quality–robustness trade-off, hardware inefficiency, compounding approximation errors, and limited compatibility with advanced techniques. Authors in [7] introduced a high-performance hand–object segmentation framework with fine-grained spatial features. A dual-attention module and an augmented feature pyramid path were designed to capture multi-scale hand information. A triplet attention module further enhanced feature representation, and a boundary refinement module was incorporated to improve boundary precision. The proposed method achieved a mask average precision of 94.6% on the VISOR dataset. However, the model has high computational and memory overhead, redundant attention mechanisms, high sensitivity to noisy backgrounds, limited global context capture, and optimization instability.

Authors in [8] proposed a Swin Transformer–based maritime object instance-based segmentation framework with dual attention and a multi-scale architecture. The framework includes a dual-attention module that reduces background noise and improves salient features. A bottom-up path aggregation module was also introduced to further improve segmentation accuracy. The model achieved an average precision of 82.71% on the MOISD dataset. Limitations include high computational

overhead for high-resolution images, restricted global context due to window-based attention, inefficiency in capturing fine details, high memory usage, poor suitability for sequential tasks, and sensitivity to hyperparameters.

Authors in [9] proposed a semi-supervised improved version of the SwinUNet model for PET/CT lung cancer lesion segmentation. The model employs a dual encoder, which includes a Convolutional Neural Network (CNN)-based residual network and a Swin Transformer to obtain rich features. A convolutional block attention module enhances feature representation, while a down-sampling attention module improves segmentation accuracy. A semantic and feature infusion module fuses multi-level features to enhance segmentation quality, and final segmentation is performed using SwinUNet. Experiments were conducted on the Adenocarcinoma dataset, and the proposed model achieved an Intersection over Union (IoU) of 91.33 ± 1.07. Some of the limitations of the model include loss of global context due to window-based attention, spatial detail loss from patch merging, high memory and computational costs, large data requirements for strong generalization, and slow inference.

Authors in [10] introduced a multitask learning network with surrounding regression for ultrasound image segmentation. The framework performs primary segmentation alongside auxiliary boundary regression. A process-based prediction refinement mechanism combines complementary outputs from both modules, and a prediction-related attention loss emphasizes challenging regions. The proposed method achieved Dice scores of 0.919 on DDIT, 0.802 on BUSI, 0.807 on TN3K, 0.791 on UDIAT, and 0.768 on TG3K datasets. The limitations of the study include task imbalance and conflicting gradients, increased training and tuning complexity, and additional computational overhead.

Authors in [11] proposed a Lightweight Multi-Attention (LMA) network to address the computational complexity of brain tumor segmentation. The model integrates an LMA mechanism that assigns weights to feature channels across multiple feature maps. It also incorporates lightweight shared dilated convolutions with different dilation rates to apprehend multi-scale information. The model achieved a Dice score of 91.35% on the BraTS 2023 Glioma dataset with only 0.516 million parameters. However, the model suffers from redundant or overlapping attention maps, a high risk of overfitting, training instability, diminishing performance returns, and sensitivity to attention-related hyperparameters.

Authors in [12] proposed a precise centerline-based method for retinal vessel segmentation in OCTA images. A boundary-peeling approach based on the Euler characteristic was applied to extract vascular centerlines, and an innovative loss function was designed to train segmentation models. The method achieved Dice scores of 0.9162 and 0.8668 on the ROSE and OCTA-500 datasets, respectively. The limitations of the study include sensitivity to image quality, error propagation from boundary peeling, and limited generalization across datasets. Authors in [13] introduced a transformer-related framework for cell counting and nuclei segmentation in microscopic images. The approach integrates a transformer-related encoder with a U-Net architecture to enhance nuclei segmentation. The proposed method was evaluated on the 2018 Data Science Bowl dataset, achieving an IoU of 86.22%. The proposed method has several limitations, including quadratic computational complexity, high data requirements, and the need for fixed-length inputs.

Authors in [14] proposed an encoder architecture named LinkNet for efficient semantic segmentation of diverse objects. The architecture enables effective feature learning without significantly increasing the parameter count. The model was implemented on the CamVid and Cityscapes datasets, achieving Mean Intersection over Union (mIoU) of 55.8% and 58.6%, respectively. Limitations include limited global context modeling, weak boundary delineation, and the absence of built-in attention mechanisms. Authors in [15] introduced an encoder and decoder-based Atrous spatial convolution method for semantic image segmentation. The encoder extracts multi-scale contextual information using various dilation rates and valuable fields of view, while the decoder restores spatial details to produce sharper boundaries. A modified Xception network was used as the backbone to enhance segmentation performance. The model achieved mIoU of 98.0% on the PASCAL VOC 2021 dataset and 82.1% on the Cityscapes dataset. Limitations include blurred boundaries, lack of attention mechanisms, reliance on dense labels, and weak part discrimination for fine-grained tasks.

Authors in [16] proposed a Feature Pyramid Network (FPN) for multi-scale object segmentation. The method constructs a multi-scale pyramid hierarchy using deep CNNs, where high-level semantic features are propagated across scales through a top-down pathway with lateral connections. This approach significantly improves feature extraction and achieves an average recall improvement of 8.3 points on the COCO dataset. The limitations of the study include a weak semantic–spatial trade-off, limited cross-scale interaction, and the absence of attention mechanisms. Authors in [17] presented U-Net, a CNN architecture for biomedical image segmentation. The model consists of a contracting path to capture contextual information and an expanding path to preserve spatial details. The architecture outperformed sliding-window convolutional networks and achieved mIoUs of 77.56% on the DIC-HeLa dataset and 92.3% on the PhC-U373 in the ISBI Cell Tracking Challenge. Limitations include limited global context modeling, weak boundary precision, scale sensitivity, poor generalization with small datasets, and fixed grid representations.

Authors in [18] introduced a Pyramid Scene Parsing Network (PSPNet), incorporating a Pyramid Pooling Module (PPM) to combine contextual information from multiple spatial regions. This design enhances pixel-level predictions for scene-parsing tasks. The model achieved mIoUs of 85.4% on the PASCAL VOC 2012 dataset and 80.2% on the Cityscapes dataset. Limitations include coarse spatial resolution, fixed-scale context modeling, and limited long-range dependency capture.

Authors in [20] proposed a colour image segmentation approach based on the fusion of Differential Evolution (DE) and a GA. Clustering-based segmentation was performed using colour and texture features to obtain prototype cluster centers.

A power-law descriptor was employed to estimate texture features for effective clustering. The GA efficiently handled binary variables, while DE was well-suited for optimizing real-valued variables, thereby enabling optimized image segmentation. Experiments conducted on images from the Berkeley Segmentation Database reported a Rand index of 0.69. However, the proposed methodology suffers from high computational overhead and does not guarantee convergence to a globally optimal solution.

Authors in [21] introduced an MRI image segmentation method based on kernel spatial Fuzzy C-Means (FCM) clustering. The method employs a conditional spatial kernel FCM algorithm, in which an auxiliary variable is associated with each pixel to incorporate spatial information. Additionally, a kernel-induced distance measure replaces the conventional Euclidean distance to enhance clustering performance. The proposed approach achieved a DSC of 0.8594 on the MRI image dataset. Despite its improved robustness, the method faces limitations in processing multi-modal MRI data and exhibits limited edge and boundary preservation capability. Other studies have presented MRI image segmentation using grey-level-based genetic K-means clustering. In this approach, cluster centroids are initialized using a GA, which alleviates the local optimum issue associated with random centroid initialization in conventional K-means. As a result, pixel classification is improved through GA-enhanced K-means clustering. The proposed method achieved a maximum segmentation accuracy of 93% on the MRI image dataset. Nevertheless, the approach remains sensitive to intensity inhomogeneity and incurs a high computational cost due to genetic optimization.

A method for detailed preservation and adaptive colour correction of underwater images, followed by segmentation, has been proposed. The approach addresses wavelength-dependent attenuation in the red and blue colour channels using an adaptive fusion strategy. The method achieved a perceptual colour quality index of 1.00 on shipwreck images. Segmentation performance was evaluated using geodesic active contours, and the qualitative results demonstrated that the proposed method outperformed existing techniques. However, the approach is highly sensitive to varying water conditions and involves high computational complexity, limiting its applicability in real-time underwater scenarios.

However, the ensemble of segmentation techniques remains relatively underexplored in the existing literature. Based on this research gap, an ensemble of segmentation techniques using GA is explored in the present work, where the GA explores the best combination of segmentation techniques using the right combining rule.

III. PROPOSED METHODOLOGY

The proposed methodology involves two major steps. In the first step, the training dataset consisting of images captured in nature and the corresponding manually segmented ground truth (mask or region of interest) is used to train the selected deep learning models. After adequate training, the performance of the models is evaluated, and the masks generated by each model are considered to produce the best mask. In the second step, the GA is used as an optimization methodology to produce the best segmentation mask by ensembling the masks generated by the individual models. Through the evolution process, the GA identifies the best performing models for various combining rules. The models and the combining rule, which have produced the best segmentation mask, are considered the final ensemble, as illustrated in Figure 1. Once the right deep learning models and the combining rule for ensembling are identified, any untrained natural scene image with a complex background can be submitted to the proposed model to segment the area of interest (segmented bird image), as shown in Figure 2.

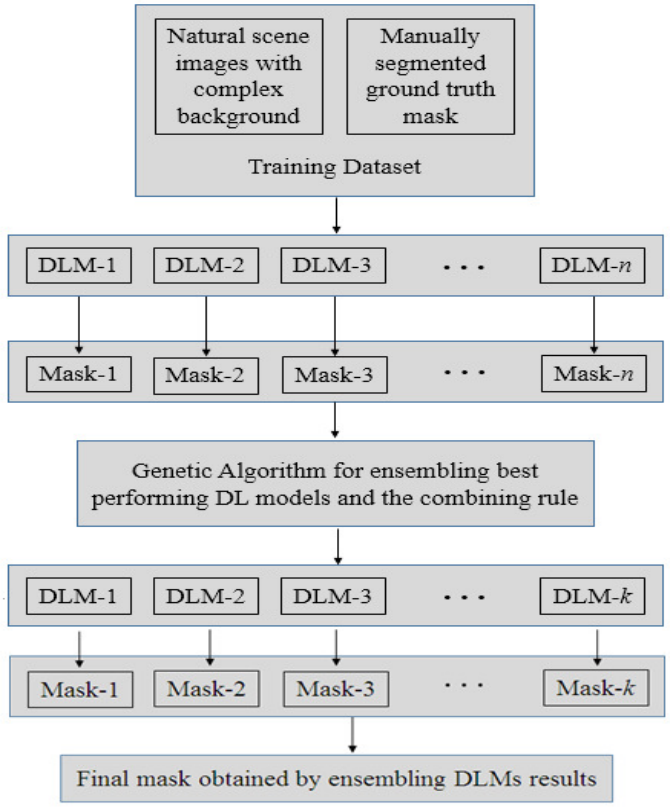

Fig. 1. Architecture of the proposed GA-based deep learning models.

A. *Deep Learning Models*

The developments in the field of image processing have significantly changed the way images are processed to build any machine learning model. Image segmentation is regarded as one of the important tasks in many image processing applications. Several deep learning models were studied to address the task of image segmentation more effectively. These models can learn complex patterns directly from the images, enabling an accurate description of objects or regions of interest within images. All these models have encoder-decoder-based architectures in common. Generally, encoders have a CNN as a backbone for extracting features and a decoder for up-sampling the tensor to produce the mask. Deep learning models were employed to develop the proposed ensembled image segmentation model. These 5 segmentation models were chosen for their reported performance in mage segmentation.

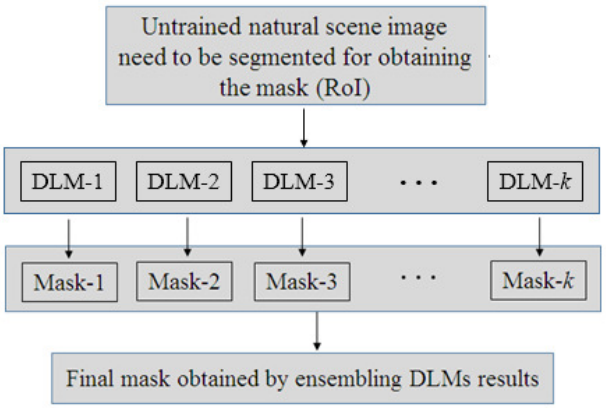

Fig. 2. Segmentation of the region of interest (bird image) from untrained natural scene images.

1) U-Net

It is a U-shaped symmetric encoder and decoder-based CNN mainly introduced for medical image segmentation [17]. The encoder (contracting path) component is designed to extract features, and a decoder (expansive path) component is used to reconstruct detailed segmentation masks. The skip connections feature introduced in the model establishes a link between the encoder and decoder modules and enables the model to preserve fine-grained spatial information. The model is found to perform well even with a small amount of labeled data.

2) Pyramid Scene Parsing Network

It is a Deep learning model proposed in [18] for image segmentation. The model typically consists of an encoder module, PPM, and a decoder module. The encoder module is a standard CNN, like ResNet, which acts as the backbone to extract rich feature maps from the input image. The PPM is used to apply pooling operations at different scales (e.g., 1×1, 2×2, 3×3) to the feature map to capture context from different regions and scales. The pooled feature maps are then upsampled to the same size and concatenated to form a final context representation. The decoder module uses convolutional layers and upsampling to generate a pixel-wise segmentation map.

3) LinkNet

It is an encoder-decoder-based neural network architecture proposed in [14] for semantic segmentation tasks. The encoder module uses CNNs to progressively decrease the spatial resolution of the given image and enhance the number of channels to extract complex and high-level features. The decoder module is used to increase the spatial resolution of feature maps obtained from the encoder module using deconvolution operations and produce the final output segmentation map. Utilization of skip connections to improve parameter efficiency is one of the distinctive features of the LinkNet model. The downsampling operations performed in the encoder result in the loss of granulated spatial details. The residual skip connections circumvent the features extracted from the resultant levels of the encoder to the decoder and preserve high-resolution spatial information useful for accurate pixel-level segmentation.

4) Feature Pyramid Network

It is a neural network architecture proposed in [16] and predominantly used for the detection and segmentation of objects with varying scales. The limitation of traditional CNNs to handle scale variation is addressed by building feature pyramids in FPNs. At each level of the feature pyramid, semantic information captured from high-level features is combined with fine-grained details captured from low-level features to produce feature maps that contain both semantic and spatial information. Thus, FPNs are used to extract robust multi-scale features for pixel-level segmentation tasks.

5) Deeplabv3+

It is a neural network model proposed in [15] for semantic segmentation. It captures image features at multiple scales by using Atrous convolution and Atrous Spatial Pyramid Pooling (ASPP) for accurate pixel-level image labeling. Resnet-like architecture was used as a backbone to extract features initially. These original features are then processed using Atrous convolution to regulate the receptive field and feature resolution. ASPP is employed to combine these multi-scale features to comprehend object context at different levels before generating the final semantic segmentation map.

6) GA for Building Ensembled Segmentation Model

The GA is an optimization technique effectively used to find optimal or near-optimal solutions to complex problems. The method is enthused by the principles of natural selection and genetics. In the proposed approach, a GA-based approach is employed to build an ensembled segmentation model by identifying the best performing, optimal number of base segmentation models and the right combining rule. The GA, as an optimization technique, combines the strengths of both evolutionary computation and ensemble learning to improve accuracy and generalization. During the evolutionary process, the GA judiciously optimizes the significance of the individual models for the ensemble rather than relying on manual tuning. The objective is to overcome the limitations of a single model and accomplish higher performance by exploiting the merits of many models for the task of segmenting the region of interest from the complex image background in general and bird image segmentation in particular.

B. The Basic Concepts of GA Used in the Proposed Work

1) Population

It is defined as a collection of potential solutions to the problem. In the proposed work, all possible combinations of the base segmentation models with all possible combining rules are considered as an initial population. Let N denote individual candidate solutions in the population. From these, a K number of potential solutions is randomly selected for further evaluation.

2) Chromosome

It is defined as the representation of a single solution within the population. It can be encoded as a string of binary digits. In this method, each chromosome is represented using a string of

7 bits. The first 2 bits describe the encoding of the rules, and the remaining 5 bits describe the inclusion or exclusion of each of the base segmentation models considered in the present work. Figure 3 shows an example chromosome representation, and Tables I and II present the encoding of combining rules and encoding bits for each segmentation model used in the study, respectively.

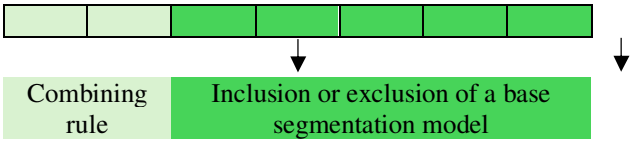

Fig. 3. Representation of a chromosome structure.

TABLE I. ENCODING OF COMBINING RULES IN THE CHROMOSOME

Combining rule	Encoding	Ensembling method
Mean	00	Average quantity of the corresponding pixels from the masks obtained by the base segmentation models
Median	01	Median quantity of the corresponding pixels from the masks obtained by the base segmentation models
Min	10	Minimum quantity of the corresponding pixels from the masks obtained by the base segmentation models
Max	11	Maximum quantity of the corresponding pixels from the masks obtained by the base segmentation models

TABLE II. ENCODING OF THE BASE SEGMENTATION MODEL IN THE CHROMOSOME

Deep learning models	Encoding bit
Deeplabv3+	1st
FPN	2nd
LinkNet	3rd
PSPNet	4th
U-Net	5th

3) Gene

Gene is the fundamental unit of a chromosome, representing a specific parameter or characteristic of the solution. Figure 4 provides an example of a gene representing a specific parameter, which indicates that the Median combining rule is used with Deeplabv3+, PSPNet, and U-Net deep models as base segmentation techniques.

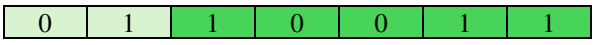

Fig. 4. Representation of a gene.

4) Fitness Function

It is a function that evaluates the quality or "fitness" of each individual in the population. A higher fitness score indicates a better solution. The fitness function is used to guide the selection process. In the proposed approach, the performance of the ensembled models in terms of segmentation accuracy, measured using the mIoU metric, is considered the fitness function.

5) Selection

It is the methodology of choosing individuals from the current population to create the next generation. Fitter individuals have a higher probability of being selected, mimicking "survival of the fittest." In the proposed approach, the binary tournament selection method is used to choose the individuals.

a) Binary Tournament Selection

Binary tournament selection is a technique used in the GA to select fitter individuals. Two chromosomes are randomly selected from the chromosome population. Based on the chromosome pattern, mIOU is estimated and assigned as the fitness; between both chromosomes, the one with the higher fitness is chosen.

Algorithm 1: Binary Tournament Selection

Input: Two chromosomes r_1, r_2 randomly selected from the population N

Output: Chromosome R with a higher fitness value

Method:

- let r_1, $r_2 \in P$, where r_1 = rand (N) and r_2 = rand (N)
- $B_1 = pattern\ (r_1)$ and $B_2 = pattern(r_2)$ where B_1 and B_2 are the pattern of the chromosomes
- $fitness1 = F(B_1)$, $fitness2 = F(B_2)$ where $F\ (.)$ returns the fitness for the given pattern.
- R= $\begin{cases} r_2 \ if\ fitness1 < fitness2 \\ r_1 \qquad\qquad otherwise \end{cases}$
- Return R

6) Crossover

It is the genetic operator that combines genetic information from two chromosomes to create new offspring chromosomes. This mimics biological reproduction and allows for investigating new solution combinations. The chromosomes selected by tournament selection are subject to recombination based on the crossover operation. The point of crossover is randomly chosen, and all bits after the point of crossover are swapped between the two chromosomes to produce two new offspring.

For example, let Chr,omosome A = 0000000 and chromosome B = 1111111, if the point of crossover is the 4th bit, then the newly generated offspring are A = 00011111 and B = 11100000. Crossover operation is used to refine solutions.

Algorithm 2: *Crossover*

Input: Population of size N, chromosome size S. chromosome$_1$ r_1, chromosome$_2$ r_2.

Output: Two offspring chromosomes r_{e1}, r_{e2}.

Method:

- Let r_1 and r_2 be the two chromosomes selected based on two binary tournament selection algorithms.

- Let P be the point of crossover such that $P_o = rand\ (0,\ S)$ where $rand()$ is a random function
- $r_{e1}[i] = \begin{cases} r_1[i]\ if\ i < P \\ r_2[i]\ otherwise \end{cases}$ where i =0, …, S−1
- $r_{e2}[i] = \begin{cases} r_2[i]\ if\ i < P \\ r_1[i]\ otherwise \end{cases}$ where i =0…, S−1
- Return r_{e1} and r_{e2}.

7) Mutation

It is a genetic operator that produces small, random changes in a chromosome's genes. This helps prevent premature convergence to suboptimal solutions by exploring new regions of the search space and maintaining diversity within the population. The mutation operator is applied to the fit chromosomes selected by the binary tournament selection process. The mutation operator compares each bit of the chromosome with a random number generator in the range [0, 1]. If the generated random number is less than the specified threshold, bits are flipped. The higher the value of the threshold (mutation rate), the higher the chances of bits getting flipped.

Algorithm 3: *Mutation*
Input: Input chromosome r, mutation rate M, chromosome size S
Output: Chromosome r_e
Method:

- Let r be the chromosome selected based on binary tournament selection algorithms.
- $r_e[i] = \begin{cases} r[i]\ if\ rand([0,1]) > M \\ 1 - r[i] \quad otherwise \end{cases}$ where, i = 0,…, S−1
- *Return* r_e

8) Generations

It is the iterative process of applying selection, crossover, and mutation operations on the chromosomes to evolve the population over time. Each iteration represents a new generation to produce progressively fitter solutions. The newer generation replaces the older generation-based accuracy of segmentation, measured in terms of mIOU. The size of the population (N) increases by two times when offspring from crossover and mutation are added to the existing population. Before passing the population to a newer generation, all chromosomes are sorted in decreasing order of fitness (mIOU), and the top N chromosomes are considered.

Algorithm 4: Production of New Generation
Input: Population of the present generation $P_{present}$ of size p is $\{r_1, r_2, \dots, r_p\}$
Output: Population of the next generation P_{next} of size q is $\{r_1, r_2, \dots, r_q\}$ where $q < p$
Method:

- $m_i \leftarrow mIOU(\text{ri}) \text{ for } i = 0, \dots, p-1$ where m_i is the fitness of each of the chromosomes i and r_i is the i^{th} chromosome of the population.
- $P_{present} \leftarrow$ Sort $\{r_1, r_2, r_3 \dots r_p\}$ is the sorted population in decreasing order of the mIOU i.e. $m_i \geq m_{i+1} \text{ for } i = 0, \dots, p-1$
- $P_{next} \leftarrow \{r_1, r_2, r_3 \dots r_q\}$ is the population of the next generation where $q < p$.
- Return P_{next}

C. Selection of Best Segmentation Models and Combining Rule

The proposed GA is used to identify the best combination of the segmentation models and the best combination rule for the ensembling process with reduced computational complexity. According to the number of base segmentation models (5) and the various combining rules (4) considered for analysis, the computational complexity of the brute force technique would be 4× (25) = 128 possible combinations. Figure 5 presents the overall workflow of the proposed GA.

Algorithm 5: Selection of Best Segmentation Ensemble
Input: (i) Population of size N
(ii) number of points of crossover m_1, mutation rate m_2, fitness metric(mIOU), number of generations n.
Output: Chromosome with the highest fitness
Method:

1. Randomly create a population of chromosomes of size N.
2. Evaluate the mIOU of each of the chromosomes and assign it as the fitness.

Loop start:

1. After picking two chromosomes from the population, apply binary tournament selection to select the best of the two chromosomes.
2. Apply crossover or mutation with equal probabilities
3. *Number = Rand (0, 1)*, where Rand is a random number generator in the given interval.
4. Offspring chromosome= $\begin{cases} crossover(chromosome)\ \ if\ 0.5 < Number \\ mutation(chromosome) \qquad\qquad otherwise \end{cases}$
5. Add the offspring chromosome to the existing *population. (N = N + offspring_chromosome)*
6. Sort all the chromosomes based on the fitness.$(Sort(N))$

7. Select the top N number of chromosomes from the population after sorting them
8. Terminate the loop when the population of the chromosome after n generations

Loop end:

1. Return the pattern of the chromosome with the highest mIOU.$(optimum\ chromosome)$

D. Dataset

The performance of the proposed ensemble segmentation model was evaluated by conducting several experiments on the Caltech UCSD BIRD-200-2011 benchmark dataset [22]. The dataset consists of 200 categories of bird species and a total of 11788 images, with the ground truth segmentation mask. The dataset is divided into training and test data based on the official dataset. The test data are further divided into a validation set and a test set. All validation data were used to obtain the best-performing top 5 chromosomes. The performance of the top 5 chromosomes was tested on the test images. GA was applied to the test set to obtain the best segmenter ensemble. To measure the performance of the model, its predictions were compared with the ground truth of the test set by applying mIoU.

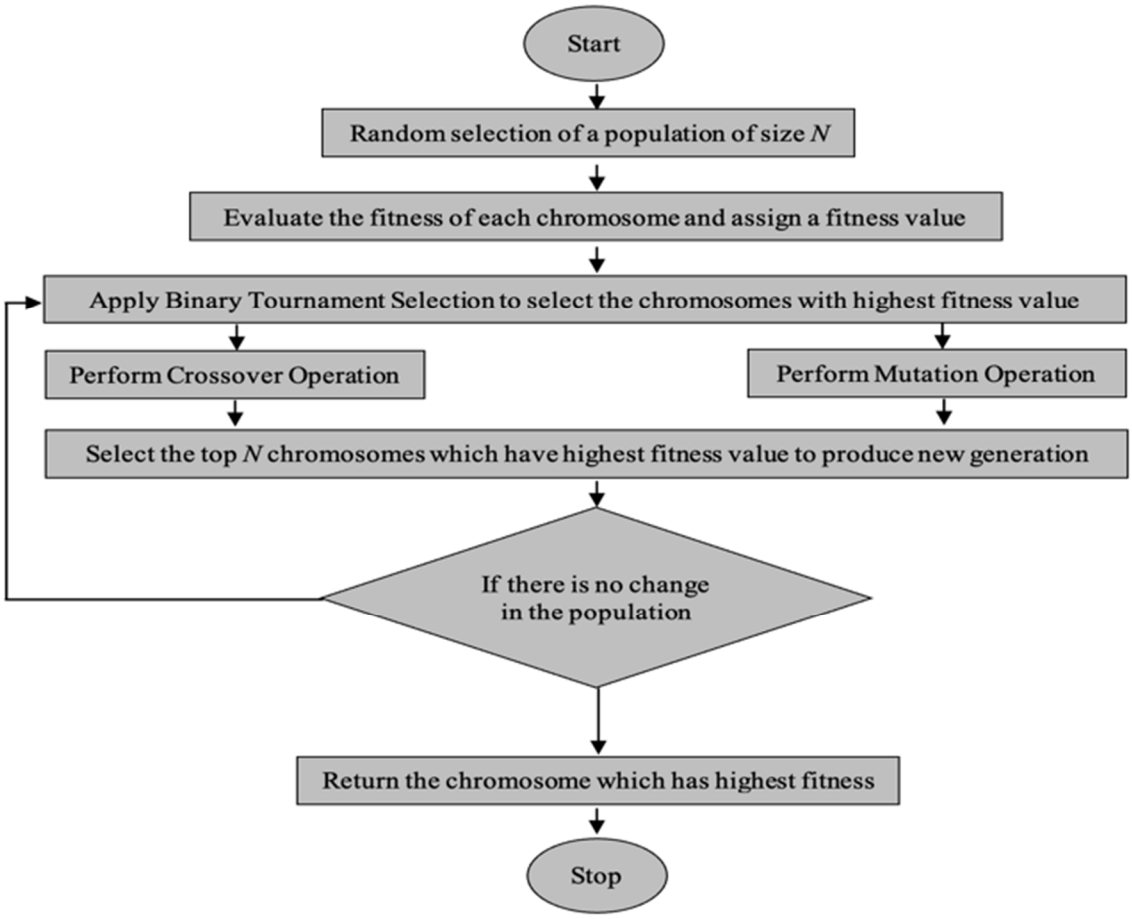

Fig. 5. Workflow of the proposed GA.

IV. RESULTS AND DISCUSSION

A. Model Parameters

Table III presents the parameters used while training the segmentation models. The Backbone model used for building the proposed model is ResNet50. Table IV presents the parameters used for the backbone model. The binary cross-entropy is utilized as the loss function, since the model was built for one-class segmentation. The Adam optimizer was used with a learning rate of 1×10^{-4}. The model was trained for 20 epochs with early stopping, with a patience of 5. The weights of the ImageNet dataset were used by the backbone during the model training process. The learning curve during the segmentation model training is illustrated in Figure 6. There are 50 layers in the backbone of ResNet50. For Global average pool layers, a pool size of 7×7 with a stride of 2 is used to reduce channels, while for the rest of the places, the stride is 1.

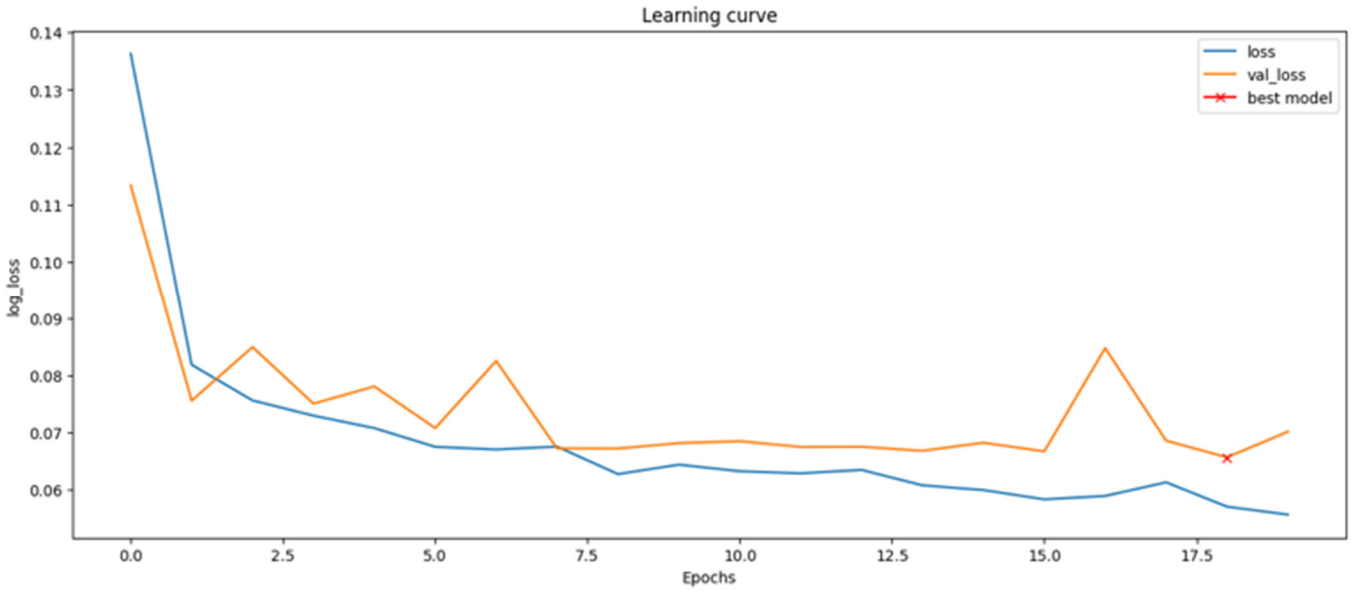

Fig. 6. Learning curve of loss while training the segmentation model.

TABLE III. TRAINING PARAMETERS FOR THE SEGMENTATION MODELS

Parameters	Values/metrics
Optimizer	Adam (1×10^{-4})
Loss	Binary cross_entropy
Metrics	Iou_score
Epochs	20
Callbacks parameters	Early stopping
Early stopping	patience=5, restore_best_weights = True
Pretrained weights of the backbone	ImageNet
dropout	0.2
Backbone CNN	ResNet50

TABLE IV. PARAMETERS FOR THE BACKBONE MODEL

Metrics	Values
Layers	50
Reduce channel conv	1×1 conv
Restore channel conv	1×1 conv
Feature extraction conv	3×3 conv
Stride (reduce channels)	2
Stride (rest of the places)	1
Poolsize (global average pool layer)	7×7

B. Parameters for the GA

Table V presents the detailed description of the parameters utilized in the experiment. mIOU is used as a metric to measure the fitness of the chromosomes. The binary tournament selection is employed to select the fittest chromosome between two chromosomes selected randomly. The mutation rate and crossover rate were 0.5 and 1.0, respectively. The main reason behind choosing a high mutation rate is to ensure that the optimization algorithm does not get stuck in the local maxima. The chosen population size is 20. The number of generations is considered to be 5, as the chromosomes in the population do not change consistently and saturate. The number of base segmentation models considered is 5. The number of combinations generated is 100 (*Population* × *Number of generations* = 20 × 5 = 100).

TABLE V. PARAMETERS USED IN THE EXPERIMENT

Parameters	Values/methods/metric
Selection of parents	Binary tournament selection
Mutation rate	0.5
Point of crossover	1
Population size	20
Number of generations	5
Number of validation images	2897
Number of test images	2897
Number of base segmentation models	5
Fitness function	mIoU

V. EXPERIMENTAL RESULTS

When models are trained, they use a mask whose values are either 0 or 1. When model inference is generated for the test images, they contain a mask whose values lie in the range [0, 1]. To classify the pixels into foreground (bird) and background, all pixel values that lie between 0 and 1 must be thresholded to convert the fractional values into 0 and 1. The threshold values decide the performance of segmentation, as the right threshold value classifies the bird from the background. Several experiments were conducted to measure segmentation performance in terms of the F1-score. The best thresholding value, which produces accurate segmentation results, was chosen empirically. The F1-score was measured by computing values, such as True Positives (TP), False Positives (FP), and False Negatives (FN), as described in Table IV.

Precision, recall, and F1-score were computed using (5), (6), and (7). The F1-score and mIOU are positively correlated, meaning that the highest F1-score corresponds to the maximum mIOU. So, the threshold value, which results in the highest F1-score, is considered to obtain the best segmentation performance by the model in terms of mIOU. Table VII depicts

the precision, recall, and F1-score for various values of threshold. It is observed that the highest mIOU was obtained for the threshold value of 0.50. Therefore, 0.50 values are considered the best threshold. Thus, in all the experiments, 0.50 is considered the threshold.

$$Precision = \frac{TP}{(TP+FP)} \quad (5)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (6)$$

$$F1 - score = 2 \times \frac{2(Precision \times Recall)}{(Precision+Recall)} \quad (7)$$

TABLE VI. DESCRIPTION OF TP, FP, AND FN IN THE CONTEXT OF BIRD SEGMENTATION

Outcome	Predicted	True	Explanation
TP	Bird	Bird	The model properly identifies the bird pixel.
FP	Bird	Background	The model mistakenly labeled a background pixel as a bird.
FN	Background	Bird	The model missed a bird pixel.

TABLE VII. PRECISION, RECALL, AND F1-SCORE FOR VARIOUS THRESHOLD VALUES

Threshold values	Precision	Recall	F1-score	mIOU
0.05	0.8428	0.9366	0.8872	0.7870
0.10	0.9054	0.9083	0.9068	0.8264
0.20	0.9543	0.8679	0.9091	0.8677
0.30	0.9463	0.9397	0.9430	0.8842
0.40	0.9406	0.9515	0.9460	0.8891
0.50	0.9628	0.9296	0.9460	0.8903
0.55	0.9711	0.9164	0.9430	0.8852
0.65	0.9606	0.9232	0.9415	0.8823
0.75	0.9769	0.8870	0.9298	0.8607
0.85	0.9175	0.8860	0.9015	0.8029
0.95	0.9444	0.7418	0.8309	0.6852

In the GA, each of the chromosomes processed is stored in the dictionary, which is used to return the top five chromosomes based on fitness. When the top five chromosomes are retrieved, it is observed that Linknet and FPN are common deep segmentation techniques in all the top 5 chromosomes.

Table VIII presents a performance comparison of the individual deep segmentation models with the ensembled model. The ensembled model has outperformed all individual models based on an mIOU of 0.8903, where all individual models achieved an mIOU less than the ensemble approach. The GA effectiveness for segmentation is highlighted in Table VIII.

TABLE VIII. PERFORMANCE COMPARISON OF THE INDIVIDUAL DEEP SEGMENTATION MODELS WITH THE ENSEMBLED MODEL

Deep segmentation model	mIOU	Ensembled model
Deeplabv3+	0.8771	0.8903
FPN	0.8825	
Linknet	0.8782	
PSPnet	0.8106	
Unet	0.8717	

Table IX presents the encoding string, deep segmentation models, combining rule, and their corresponding mIOU for the top five best-performing chromosomes. Among the top five best-performing chromosomes, the one with the encoding string 0011101 has produced the highest mIOU by including Deeplabv3+, FPN, Linknet, Unet segmentation models, and the mean combining rule. The five best-performing chromosomes outperformed the individual segmentation models, attaining mIOU values of 0.89033, 0.8899, 0.8895, 0.8893, and 0.8886, respectively.

TABLE IX. DETAILS OF THE TOP FIVE BEST-PERFORMING CHROMOSOMES

Top list	Encoding string	Deep segmentation models included	Combining rule	mIOU
1	0011101	Deeplabv3+, FPN, Linknet, Unet	Mean	0.89033
2	1111101	Deeplabv3+, FPN, Linknet, Unet	Max	0.8899
3	1101101	FPN, Linknet, Unet	Max	0.8895
4	0001101	FPN, Linknet, Unet	Mean	0.8893
5	1111100	Deeplabv3+, FPN, Linknet,	Max	0.8886

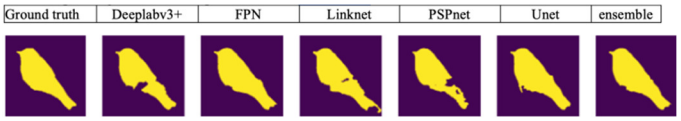

Fig. 7. Sample images of the masks produced by the segmentation models.

Fig. 8. Sample images of the masks produced by the segmentation models.

Figures 7 and 8 show a few examples of segmented masks produced by the proposed model. In Figure 7, it is observed that the mask produced by LinkNet is better than the other models. Similarly, as demonstrated in Figure 8, the FPN and U-Net models produced better results. Based on these examples, the ensemble represents a cumulative mask that performs well and does not rely on any single technique. To demonstrate the superiority of the proposed ensemble method, the segmentation results obtained on the benchmark dataset are compared with the results obtained by the other state-of-the-art methods available in the literature. Table X presents the comparative analysis of the results.

Fig. 9. Sample images of the masks produced in the experiment.

Figure 9 illustrates a few example segmentation masks produced by the individual deep segmentation models and the proposed ensembled deep model. It can be observed from the masks that the outcomes obtained by the proposed ensembled method are closer to the ground truth masks than the individual models.

The details missed by the individual models are covered by the ensembling process and are evident from the images presented in Figure 9. None of the individual masks produces better results for all images in the dataset. This indicates that ensembling the results obtained by the individual models is more effective in obtaining better segmentation. The segmentation masks obtained for all images used in the dataset indicate that the ensemble model using GA has performed well with satisfactory results.

TABLE X. SEGMENTATION PERFORMANCE OF THE PROPOSED METHOD WITH STATE-OF-THE-ART METHODS ON THE CUB-200-2011 DATASET

Method	Backbone	mIOU	Precision	Recall	F1-score
U-Net	ResNet-34	78.2	84.1	81.2	82.6
DeepLabV3+	ResNet-50	81.4	86.7	83.9	85.3
HRNet	HRNetV2-W18	82.1	87.4	85.0	82.1
SegFormer	MiT-B1	83.5	89.2	86.5	87.8
GeloVec [19]	ResNet-34	83.6	92.1	88.6	89.0
Proposed	-	89.033	96.28	92.96	94.6

VI. CONCLUSION

This study presented a novel approach of ensembling deep learning models using the Genetic Algorithm (GA) for improving bird image segmentation in a complex background. Before ensembling, the segmentation capability of individual segmentation models was studied, and their limitations were observed. The results indicate that ensembling effectively addresses the limitations of individual models.

Segmentation results obtained by conducting experiments on the benchmark dataset indicate that there is a significant improvement in overall performance by the ensembling model over individual models. In the proposed GA-based ensembling approach, the genetic operations such as crossover and mutation were applied in parallel, compared to previous studies in which both crossover and mutation were applied one after

another. The GA was able to find the optimal chromosome with a lower search space compared to the brute force methodology. The parameters were chosen empirically by conducting several experiments to fine-tune the GA. The method involves a significant amount of training time with considerably large and diverse images to build the model. The novelty of this work lies in combining pixel-level information obtained from different masks to produce a better segmentation mask. This idea can be further extended for other tasks, such as denoising an image and obtaining better feature maps.

DECLARATION OF COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

The Caltech UCSD BIRD-200-2011 dataset used in this study was collected from [22]. The source code is available at [23]. The file has to be downloaded in the local system and has to be opened with Jupyter notebook.

REFERENCES

[1] J. C. Canales, J. C. Canales, F. García-Lamont, A. Yee-Rendon, J. S. R. Castilla, and L. R. Mazahua, "Optimal Segmentation of Image Datasets by Genetic Algorithms Using Color Spaces," *Expert Systems with Applications*, vol. 238, Mar. 2024, Art. no. 121950, https://doi.org/10.1016/j.eswa.2023.121950.

[2] H. Nematzadeh, J. García-Nieto, I. Navas-Delgado, and J. F. Aldana-Montes, "Ensemble-Based Genetic Algorithm Explainer with Automized Image Segmentation: A Case Study on Melanoma Detection Dataset," *Computers in Biology and Medicine*, vol. 155, Mar. 2023, Art. no. 106613, https://doi.org/10.1016/j.compbiomed.2023.106613.

[3] P. Ghosh, M. Mitchell, J. A. Tanyi, and A. Y. Hung, "Incorporating Priors for Medical Image Segmentation Using a Genetic Algorithm," *Neurocomputing*, vol. 195, pp. 181–194, Jun. 2016, https://doi.org/10.1016/j.neucom.2015.09.123.

[4] M. Nouman, M. Mabrok, M. al-Shatouri, and E. A. Rashed, "Contextual Transformer for Brain Stroke Segmentation in MRI," *Computers and Electrical Engineering*, vol. 129, Jan. 2026, Art. no. 110795, https://doi.org/10.1016/j.compeleceng.2025.110795.

[5] H. Zhou, L. Xu, C. Liu, and G. Li, "UDNDSNet: A Unified Deterministic and Non-Deterministic Segmentation Network for Multi-Scene Medical Image Analysis," *Knowledge-Based Systems*, vol. 331, Jan. 2026, Art. no. 114641, https://doi.org/10.1016/j.knosys.2025.114641.

[6] Y. Ji, H. Li, G. Li, S. Liu, and X. Wan, "Sparse Diffusion Models for Multi-Annotator Medical Image Segmentation," *Knowledge-Based Systems*, vol. 330, Nov. 2025, Art. no. 114783, https://doi.org/10.1016/j.knosys.2025.114783.

[7] W. Wang, J. Dang, Y. Wang, and R. Pan, "RefineHOS: A High-Performance Hand–Object Segmentation with Fine-Grained Spatial Features," *Computer Vision and Image Understanding*, vol. 262, Dec. 2025, Art. no. 104548, https://doi.org/10.1016/j.cviu.2025.104548.

[8] H. Yin *et al.*, "Swin Transformer-Based Maritime Objects Instance Segmentation with Dual Attention and Multi-Scale Fusion," *Computer Vision and Image Understanding*, vol. 262, Dec. 2025, Art. no. 104556, https://doi.org/10.1016/j.cviu.2025.104556.

[9] S. Tang *et al.*, "Segmentation of PET/CT Lung Cancer Lesion Images via a Semi-Supervised Improved SwinUNet Model," *Biomedical Signal Processing and Control*, vol. 113, Mar. 2026, Art. no. 108797, https://doi.org/10.1016/j.bspc.2025.108797.

[10] Z. Huang, Q. Yang, and Y. Gao, "A Multi-Task Learning Network with Boundary Regression for Automatic Ultrasound Image Segmentation," *Biomedical Signal Processing and Control*, vol. 113, Mar. 2026, Art. no. 108828, https://doi.org/10.1016/j.bspc.2025.108828.

[11] Z. Chen, Y. Wang, X. Li, and B. Li, "LSMAtt-Net: A Lightweight Brain Tumor Segmentation Approach with Efficiency–Complexity Tradeoff," *Biomedical Signal Processing and Control*, vol. 113, Mar. 2026, Art. no. 108930, https://doi.org/10.1016/j.bspc.2025.108930.

[12] X. Chen, C. Wang, X. Wu, Q. Hu, X. Li, and S. Li, "S3-OCTA: A Precise Centerline-Based Method for Retinal Vessel Segmentation in OCTA Images," *Biomedical Signal Processing and Control*, vol. 113, Mar. 2026, Art. no. 109117, https://doi.org/10.1016/j.bspc.2025.109117.

[13] H. P. Ramya Shree, Minavathi, and M. S. Dinesh, "Transformer Based Framework for Nuclei Segmentation and Cell Counting on Microscopic Images of Multiple Organs," *International Journal of Intelligent Networks*, vol. 6, pp. 283–292, 2025, https://doi.org/10.1016/j.ijin.2025.11.003.

[14] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation," in *IEEE Visual Communications and Image Processing*, St. Petersburg, FL, USA, Dec. 2017, pp. 1–4, https://doi.org/10.1109/VCIP.2017.8305148.

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Computer Vision – ECCV 2018*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer International Publishing, 2018, pp. 833–851.

[16] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, Jul. 2017, pp. 936–944, https://doi.org/10.1109/CVPR.2017.106.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer International Publishing, 2015, pp. 234–241.

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, Jul. 2017, pp. 6230–6239, https://doi.org/10.1109/CVPR.2017.660.

[19] B. Kriuk and M. Yordanov, "GeloVec: Higher Dimensional Geometric Smoothing for Coherent Visual Feature Extraction in Image Segmentation," in *Empowering Novel Geometric Algebra for Graphics and Engineering*, vol. 16510, E. Hitzer, G. Papagiannakis, and H. Hadfield, Eds. Cham, Switzerland: Springer Nature Switzerland, 2026, pp. 25–36.

[20] R. V. V. Krishna and S. Srinivas Kumar, "Hybridizing Differential Evolution with a Genetic Algorithm for Color Image Segmentation," *Engineering, Technology & Applied Science Research*, vol. 6, no. 5, pp. 1182–1186, Oct. 2016, https://doi.org/10.48084/etasr.799.

[21] B. Gharnali and S. Alipour, "MRI Image Segmentation Using Conditional Spatial FCM Based on Kernel-Induced Distance Measure," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2985–2990, Jun. 2018, https://doi.org/10.48084/etasr.1999.

[22] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "Caltech-UCSD Birds-200-2011 (CUB-200-2011)." California Institute of Technology, 2011, [Online]. Available: https://www.vision.caltech.edu/datasets/cub_200_2011/.

[23] B. S. Chandrashekar, "segmentation-using-Genetic-algorithm." [Online]. Available: https://github.com/chandrashekar6/segmentation-using-Genetic-algorithm-/blob/main/drafting_GA_segmentation_final_(1)-3.ipynb.

AUTHORS PROFILE

B S Chandrashekar is pursuing his PhD under the supervision of H S Nagendraswamy in the Department of Studies in Computer Science, University of Mysore, Mysuru. He has received a B.Tech. and M.Tech degree in computer science from the Indian Institute of Information Technology Design and Manufacturing, Kancheepuram in 2021. His research area includes single-event upset, design of fault-tolerant systems, image classification, image segmentation, image denoising, machine learning, data mining, and image processing. He can be contacted at email: coderchandru@gamil.com.

Google Scholar ID: https://scholar.google.com/citations?user=jITBwngAAAAJ&hl=en

Orchid ID: https://orcid.org/0009-0000-4655-4357

Web of Science ResearcherID: MSV-7962-2025

Nagendraswamy H S obtained his M. Sc and PhD degrees from the University of Mysore, India, in 1994 and 2007, respectively. He is currently working as a Senior Professor in the Department of Studies in Computer Science, University of Mysore, Manasagangothri, Mysore, Karnataka, India. His research areas include shape analysis, texture analysis, sign language recognition, precision agriculture, symbolic data analysis, fuzzy theory, biometrics, and video analysis. He has been a reviewer for many international journals, including fuzzy systems, pattern recognition, pattern recognition letters.

https://orcid.org/0009-0001-2614-6169

Dr. Pavan Kumar M P has a PhD from Visveswaraya Technological University, Belagavi. He has more than 15 years of Teaching experience. Currently, he is guiding 5 PhD scholars. He has qualified for the Karnataka State Eligibility Test (KSET) and served as a principal investigator for the Karnataka Govt. VGST KFIST II funding project entitled "Novel Integrated Filter-Based Approaches for Image Abstraction to Investigate the Hidden Research Issues in the Image Processing Domain" for a grant of Rs. 40 lakhs. He has also served as a member of the Board of Examinations at VTU, Belagavi, and is currently serving as a Member of the Board of Studies at VTU, Belagavi. He is an awardee of Rastrapathi Purasakar, received from the President of India.

https://scholar.google.co.in/citations?user=EEKifKYAAAAJ&hl=en

https://orcid.org/0000-0002-9372-9065