



OPEN Comprehensive evaluation of optimization algorithms for medical image segmentation

Nijad A. Al-Najdawi^{1✉}, Ali F. Al-Shawabkeh², Sara Tedmori³, Ibrahim I. Ikhries⁴ & Osama Dorgham¹

Medical image segmentation algorithms play a crucial role in assisting healthcare professionals with disease identification, research, and diagnosis. Numerous digital image segmentation methods have been developed, with multilevel thresholding techniques consistently outperforming others in terms of evaluation metrics. The standard algorithms include classical statistical methods, such as the Otsu and Kapur methods, which yield highly accurate results. However, when applied to multilevel thresholding, these methods incur significant computational costs, presenting an optimization challenge. In this work, a set of well-known optimization algorithms is integrated with Otsu's method to assess their effectiveness in reducing computational demands while preserving optimal segmentation quality. Experiments are conducted on publicly available datasets, including the TCIA dataset, particularly the COVID-19-AR collection. This work evaluates the performance of each optimization algorithm in combination with Otsu's method, highlighting those that achieve substantial reductions in computational cost and convergence time while maintaining a competitive level of segmentation quality.

Keywords Medical images, Segmentation, Optimization algorithms, Multi-level thresholding, Computational efficiency

Medical imaging has become one of the most useful diagnostic and treatment planning tools in modern healthcare. Among the wide number of available techniques, image segmentation is considered one of the most important techniques, as it is a well-known task of partitioning an image into distinct regions according to predefined criteria. Accurate segmentation would greatly help in the identification of organs, tissues, or other abnormalities in medical scans like MRI, CT, and ultrasound images. Within the last few decades, many methodologies have been developed for medical image segmentation based on edge-, region growing-, and multi-level thresholding-based approaches¹.

Multilevel thresholding methods have exhibited much promise because they can efficiently deal with complex medical images with heterogeneous distribution of intensities. In the process of multilevel thresholding, the image is divided into more segments, each having a different threshold. Although such algorithms generally produce optimal results, they suffer from the following drawback: their computational load increases exponentially with the increase in the number of thresholds. This makes the procedure computationally expensive, especially when extended to high-resolution medical images, therefore this field has seen interest grow in optimizing these methods².

The most used technique of thresholding is the Otsu method, whereby the best threshold value is obtained automatically to minimize variance within classes of gray levels. However, for multilevel thresholding, the Otsu method is computationally super expensive. Otsu's method divides the image histogram into two clusters, with a threshold determined by minimizing the weighted variance within each class known as the Otsu model and denoted as $\sigma_w^2(t)$, where $\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t)$, $w_1(t)$, $w_2(t)$ are the probabilities of the two classes divided by a threshold (t) , where $(0 \leq t \leq 255)$. Otsu's method offers two options for determining the threshold. The first option is to minimize the within-class variance $\sigma_w^2(t)$ and the second is to maximize

¹Department of Computer Science, Prince Abdullah bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Al-Salt 19117, Jordan. ²Department of Scientific Basic Sciences, Faculty of Engineering Technology, Al-Balqa Applied University, Amman 11134, Jordan. ³Department of Computer Science, King Hussein School for Computing Sciences, Princess Sumaya University for Technology, Amman 11941, Jordan. ⁴Department of Mechanical Engineering, Faculty of Engineering Technology, Al-Balqa Applied University, Amman 11134, Jordan. ✉email: n.al-najdawi@bau.edu.jo

the between-class variance using $\sigma_b^2(t) = w_1(t)w_2(t)[\mu_1(t) - \mu_2(t)]^2$, where μ_i is a mean of class i . The cluster probability function is used to calculate the probability P for each pixel value in the two separated clusters

$$C_1, C_2 \text{ as } w_1(t) = \sum_{i=1}^t P(i) \text{ and } w_2(t) = \sum_{i=t+1}^I P(i), \text{ respectively}^3.$$

Digital images can be represented by an intensity function $f(x, y)$, which assigns gray-level values to pixels. Here, n denotes the total number of pixels, and i represents the number of pixels at a specific gray level. The probability of gray-level i occurrence is $P(i) = n_i/n$. The pixel intensity values for C_1 and C_2 are within the range of $[1, t]$ and $[t + 1, I]$, respectively, where I is the maximum intensity value

(i.e., 255). The means for C_1, C_2 , denoted by $\mu_1(t), \mu_2(t)$, are obtained by $\mu_1(t) = \sum_{i=1}^t iP(i)/w_1(t)$ and $\mu_2(t) = \sum_{i=t+1}^I iP(i)/w_2(t)$, respectively. Therefore the (σ_1^2, σ_2^2) values can be obtained by

$$\sigma_1^2(t) = \sum_{i=1}^t [i - \mu_1(t)]^2 P(i)/w_1(t), \text{ and } \sigma_2^2(t) = \sum_{i=t+1}^I [i - \mu_2(t)]^2 P(i)/w_2(t), \text{ respectively}^3.$$

Otsu's method determines the threshold(s) by maximizing the between-class variance. The core equations and variables involved are summarized as follows:

w_1, w_2 : Probabilities of two classes divided by threshold t .

μ_1, μ_2 : Class means.

σ_b^2 : Between-class variance: $w_1w_2(\mu_1 - \mu_2)^2$.

σ_w^2 : Within-class variance.

The Otsu algorithm selects the optimal threshold by maximizing between-class variance, assuming a Gaussian intensity distribution. However, for right-skewed histograms, this assumption may fail. To overcome this⁴, proposed a lognormal-based model that adjusts the variance calculation using lognormal mean and variance, improving performance on asymmetrical images.

Recently, the multilevel thresholding method relied on optimization algorithms in order to have the best performance. For that, optimization algorithms propose the most accurate thresholds while conducting substantially fewer computations. These algorithms are usually designed to minimize the objective function that represents the error in segmentation but with quick convergence to an optimal solution. In this respect, several algorithms were proposed with different strengths and weaknesses⁵.

Segmentation of medical images, therefore, by applying optimization algorithms, is of great importance in the real world of medical practices. While medical imaging technologies improve, image data volumes and complexities continue to rise. It is usually expected that high-resolution scans from modalities like MRI or CT would require expert segmentation techniques for the proper delineation of important anatomical structures⁶. Where this is the case, optimization algorithms may reduce computational time considerably without sacrificing the accuracy of the calculation. Clinically, this would be very valuable since diagnosis would be fast but with a high degree of accuracy. In addition, optimization algorithms enable more coherent and reproducible segmentation results, which is an important development factor for systems aimed at reducing inter-operator variability in medical diagnosis. Available segmentation methods, either using automatic or semi-automatic optimization algorithms, help the radiologist and the medical practitioner make more informed decisions that shall improve patient outcomes².

The aim of this research is to assess the effectiveness of several well-known optimization algorithms in the context of medical image segmentation. This evaluation includes a range of methods, such as evolutionary algorithms, swarm intelligence techniques, and recent heuristic approaches, which have shown potential for achieving high-quality segmentation with reduced computational complexity. Consequently, this study focuses on key challenges related to minimizing computation time and enhancing segmentation accuracy, providing insights into which algorithms best balance these objectives in medical imaging applications.

Literature review

Optimization algorithms for image segmentation can be broadly classified into nature-inspired, deterministic methods and human-inspired algorithms. An investigation in the field of medical imaging, supported by an intensive study of recent literature, reveals that image segmentation methods are the effective approaches that summarize and simplify the representation and analysis of the medical images. In⁷, the authors have developed an algorithm to increase the effectiveness of image segmentation using multilevel thresholding based on the Kapur and Otsu techniques. The algorithm is based on differential evolutionary (DE) and bird mating optimization (BMO) algorithms. In⁶, the Harris Hawks Optimization technique is combined with Otsu's method to effectively reduce the computational cost while achieving optimal outcomes. The proposed approach is evaluated on publicly available imaging datasets, including chest images that represent a rural COVID-19-positive (COVID-19-AR) population. Based on various performance metrics, the proposed approach demonstrates a substantial reduction in computational cost and convergence time while maintaining a quality level highly competitive with the traditional Otsu method at the same threshold values.

In², the authors emphasized that accurate segmentation of computed tomography (CT) scan volumes is a critical step in radiomic analysis and in developing advanced surgical planning techniques. To address these issues, the authors proposed an automated deep learning (DL) segmentation framework for CT images called U-Net CT Segmentation (U-NetCTS), which integrates the DL U-Net model with CT images for automatic segmentation⁸. Experimental results show that the U-NetCTS framework can accurately segment different regions of interest in various CT DICOM images. Zhang et al.⁹ proposed a hybrid model combining convolutional neural networks and transformers, achieving significant improvements in spatial representation and boundary

accuracy. Additionally, Strika et al.¹⁰ conducted a broad review of how deep learning and large language models can help bridge healthcare access disparities, particularly in under-resourced regions where lightweight or automated segmentation methods are critically needed.

To segment grayscale images using multi-level thresholding, the authors of¹¹ suggested a multi-objective metaheuristic based on a multi-verse optimization method. The idea behind the suggested strategy is to maximize the Kapur and Otsu goal functions to find an approximate Pareto-optimal set. Two popular picture segmentation techniques based on bilevel, and multi-level thresholding are represented by both Kapur's and Otsu's methods. All these techniques, nevertheless, have their own peculiarities and restrictions. While several multi-objective approaches have investigated the advantages of jointly using Kapur and Otsu's methods, only a small number of meta-heuristic approaches have been put out in the literature for the independent optimization of these objective functions with respect to accuracy. In¹², the authors proposed an algorithm to determine a global thresholding value for a specific image. To identify the optimal threshold, they employed a Differential Evolution algorithm combined with the OTSU method and a trained neural network for future applications.

The authors of¹³ proposed a multilevel thresholding algorithm using a meta-heuristic Krill Herd Optimization algorithm for solving the image segmentation problem. Optimum threshold values obtained by maximizing Kapur's or Otsu's objective function using the Krill Herd Optimization technique. In¹⁴, human mental search algorithm has been proposed by authors for image segmentation using a multilevel thresholding algorithm. For this purpose, Kapur and Otsu criteria have been used. This method has been compared with many other existing algorithms. The results obtained showed that the proposed method was efficient in multilevel image thresholding.

A metaheuristic algorithm was proposed by the authors to search for the optimal thresholds. In¹⁵, a multilevel cross entropy is rendered using a modified grasshopper optimization algorithm for this purpose. Besides, Levy flight algorithm is utilized to modify the original grasshopper optimization algorithm and balance its exploration and exploitation. The proposed method was compared to other thresholding methods and showed that the proposed segmentation approach has fewer iterations and higher segmentation accuracy. In¹⁶, two methods have been added to the original Whale optimization algorithm: WOA, a random replacement strategy that increases the algorithm's speed of convergence, and a double adaptive weight strategy that seeks to balance the optimization process's exploitative behaviors in later stages with its exploration searching trend in the early stages. The authors have combined these two approaches and found that doing so increases the overall search ability as well as the convergence speed. Typical benchmark examples include unimodal, multi-modal, fixed multi-modal functions, and three well-known engineering design challenges that are used for in-depth examination and investigation of the benefits of the suggested technique.

In literature, only a few image segmentation techniques consider Many Objectives-Multilevel functions, which look for the best option to lessen conflict between the many objectives. These approaches do, however, have certain drawbacks, which cause them to perform worse as the number of objectives rises because there are more non-dominated solutions. As a result, by considering seven objective functions, authors of¹⁷ have suggested an alternate image segmentation technique utilizing many-objective optimization algorithms.

In¹⁸, the authors proposed Context-based semantic segmentation. The Spatial Contextual Module makes use of this information to reveal the spatial contextual dependency between pixels. The authors of¹⁹, performed a unification of n-level thresholding and fractional order Darwinian particle swarm optimization with detailed consideration. It was found a competitive approach amongst the various particle swarm optimization for mammogram image segmentation. The authors of²⁰, demonstrated under which circumstances the optimal weighting in terms of weighted cross-entropy exists, that would maximize both Dice score and Jaccard index at test time. It has been observed that the two scores approximate each other, but it had not been the case when using a weighted Hamming similarity. The authors claim that results have been evaluated on 6 medical segmentation tasks and confirm that metric sensitive losses do outperform cross-entropy based loss functions when evaluation with the two scores is done. Moreover, a variety of cutting-edge techniques for lung nodule detection have been assessed²¹⁻²³. These works are based on factors such as validation dataset types, nodule sizes, total cases involved, nodule types, features extracted from conventional feature-based classifiers, sensitivity, and FPs/scans.

Recent studies have introduced enhanced models that integrate optimization techniques with machine learning in healthcare applications. For example, the EO-LWAMCNet model proposed in²⁴ combines a lightweight convolutional neural network with an Evolutionary Optimization (EO) strategy to enable real-time disease prediction using sensor data within an IoT framework. This model demonstrated high accuracy on chronic liver and brain disease datasets (94.8% and 95%, respectively), emphasizing the clinical value of early and remote diagnosis through optimized learning systems. Similarly²⁵, presents a comparative evaluation of machine learning algorithms—such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN)—for interpreting medical images and patient data. The study highlights the role of algorithm selection and data characteristics in diagnostic accuracy, underscoring the relevance of comparative benchmarking when applying AI in medical contexts.

In parallel to traditional and nature-inspired optimization methods, recent advances in medical image segmentation have explored deep learning-based approaches that focus on model efficiency, boundary precision, and privacy preservation. Notable examples include DA-TransUNet, which integrates dual attention and transformer modules for enhanced spatial representation²⁶; FKD-Med, which leverages federated learning and knowledge distillation for lightweight, privacy-aware segmentation²⁷; and methods like mutual inclusion mechanisms²⁸ and decoder-focused distillation, which improve boundary delineation and computational efficiency. While these approaches fall outside the direct scope of this study, they offer valuable perspectives on evolving segmentation strategies.

In Ryalat et al.⁶, the authors applied the Harris Hawks Optimization (HHO) algorithm in combination with Otsu's method for COVID-19 chest CT image segmentation. While the authors in⁶ focused on demonstrating the effectiveness of HHO in reducing computational time while maintaining segmentation accuracy, the present study significantly extends that work in both scope and depth. Specifically, this manuscript conducts a comparative evaluation of 18 optimization algorithms—spanning evolutionary, swarm-based, and nature-inspired methods—on the same task. In addition to testing a broader set of algorithms, this study provides a comprehensive benchmarking of segmentation performance using multiple metrics (e.g., Dice, Jaccard, F1-score, sensitivity, robustness), including memory usage and convergence analysis, and proposes an algorithm selection guide tailored to real-time vs. high-accuracy clinical scenarios. These contributions position this work as an extensive comparative study that offers broader insights beyond those reported in⁶.

While deep learning models—particularly architectures like U-Net and its variants—have become dominant in medical image segmentation tasks, this study intentionally focuses on traditional threshold-based techniques optimized using metaheuristic algorithms. This choice is driven by the need for lightweight, interpretable, and hardware-efficient segmentation approaches, particularly in clinical environments with limited computing resources or in edge-based diagnostic tools. Optimization-based methods offer high accuracy for tasks where intensity-based separation is viable, and they can be implemented without requiring extensive labeled datasets or GPU acceleration. Moreover, this work complements deep learning literature by offering benchmark insights into which optimization strategies are most effective under fixed-resource conditions. Future work may explore hybrid models that incorporate deep features into the optimization process, bridging the strengths of both paradigms.

Medical images dataset

There are several medical imaging datasets that can be employed for testing optimization algorithms and these datasets are publicly available (as shown in Table 1). In this respect, they also span a wide area of applications in medical imaging tasks and, hence, constitute very well-qualified test beds to benchmark various optimization algorithms when viewed from the perspective of medical image segmentation and analysis. The list includes:

The papers in^{38–41}, and⁴² provide recent evaluations of optimization algorithms in medical image segmentation. The work in³⁸ focused on advancements and challenges in deep learning approaches, while the work in³⁹ examined quantum-enabled algorithms for medical image processing. In⁴⁰ The authors discussed multi-modality-based image fusion methods. In⁴¹, the authors provided a comprehensive review for nature-inspired optimization algorithms specifically for segmentation. The authors of⁴² explored traditional, deep learning, and hybrid approaches. Together, these studies offer a comprehensive review of recent optimization methods in the field.

It is apparently evident in the literature that all the wide varieties of available optimization algorithms can be employed to efficiently minimize the required computing cost while keeping optimal results when coupled with Otsu's and Kapur algorithms. Therefore, this research evaluates the effectiveness of popular optimization algorithms for medical image segmentation, with the aim to identify algorithms that optimize segmentation quality while minimizing computation time, highlighting those that best balance these objectives in medical imaging.

Optimization algorithms	Definitions	Applications	Ref.
ISIC	International Skin Imaging Collaboration Dataset	Digital skin imaging for skin cancer	29
BraTS	Brain Tumor Image Segmentation Dataset	Multimodal brain MRI scans	30
DDSM	Digital Database for Screening Mammography	Breast cancer screening	31
ACRIN-FMISO Dataset	The Cancer Imaging Archive	PET/CT images	32
CHUM	(Cardiac Atlas Project) Dataset	heart MRI and CT	33
DRIVE	Digital Retinal Images for Vessel Extraction Dataset	Retinal images for blood vessel segmentation	34
MICCAI	Multimodal Brain Tumor Image Segmentation Benchmark	Multimodal MRI scans for brain tumor segmentation	30
CIA	Cancer Imaging	Chest imaging/ COVID-19-positive population.	35
TCIA Dataset	Datasets of cancer imaging studies	PET scans	36
LIDC-IDRI	The Lung Image Database Consortium and Image Database Resource Initiative Dataset	lung CT images for the detection and diagnosis of lung nodules	37

Table 1. Medical imaging datasets that can be employed for testing optimization algorithms.

Proposed work

The segmentation algorithms of medical images assist medical experts to locate, study, and diagnose specific diseases. Several methodologies have been addressed in the past decades regarding the segmentation of digital images. However, most of these approaches have been outperformed by multilevel thresholding methods with respect to the results of evaluation techniques. The benchmark algorithms for autonomous image thresholding are classical statistical methods like the Otsu and Kapur methods. These algorithms provide the best results but at high computational cost, which is an optimization problem when multilevel thresholding is required.

Well-known optimization algorithms

This research implements, tests, and validates several optimization algorithms that have been used in literature to efficiently reduce the computing cost required for maintaining optimal results. This research evaluates 18 optimization algorithms (see Table 2), categorized into three widely accepted classes based on their underlying design principles: Nature-Inspired algorithms, Deterministic methods, and human-inspired algorithms.

A further classification of the well-known optimization algorithms has been utilized in this research by categorizing these algorithms based on their underlying methodologies, such as Nature-inspired, Deterministic Methods and human-inspired approaches, highlighting the variety of optimization techniques considered. This classification serves as a foundation for selecting and comparing the performance of different algorithms within the study's experimental framework.

Experiments and results

Methodology

In this research, the selected optimization algorithms for medical image segmentation have been evaluated based on the TCIA dataset³⁵. To assess the efficiency of the proposed approach, an imaging dataset publicly available from the Cancer Imaging Archive³⁵ was utilized. This dataset includes CT images with a resolution of 762×762 pixels, pixel spacing of 1.08×1.08 mm and 0.98×0.98 mm, and a slice thickness of 3.14 mm. The images feature chest scans with clinical information for a rural COVID-19-positive population (COVID-19-AR), covering approximately 105 patients. For this research, a subset of 18 patient records were randomly selected to test and evaluate the approach.

The experiment includes implementing and testing 18 different optimization algorithms. Each algorithm is exploited in undertaking multilevel thresholding-based image segmentation. The aim of the experiment is therefore how best each algorithm can optimize the process of segmentation taking into consideration the computational resources and the segmentation accuracy.

Figure 1 summarizes the methodology employed in this research, where preprocessing includes histogram equalization for contrast correction, min-max scaling for normalization, and Gaussian or median filtering for noise reduction (as highlighted in⁶⁰. Such steps can suppress noise and improve contrast, potentially boosting segmentation accuracy. This is followed by segmentation using multilevel thresholding, where threshold values are optimized using metaheuristic algorithms. Post-processing involves morphological operations to refine the segmented masks. The resulting segmented image is then evaluated using the objective function. Robustness is assessed by analyzing the consistency of segmentation performance across images of varying quality.

Each optimization algorithm in this study was applied to maximize the between-class variance defined by Otsu's method for multilevel thresholding (we followed the standard practice of maximizing the between-class variance for computational convenience and clarity). Specifically, the Otsu objective function computes a fitness value based on a given set of threshold values, where higher between-class variance indicates better segmentation. The optimization algorithms (e.g., PSO, HHO, DE) operate by evolving a population of candidate threshold sets across iterations. Each candidate (or agent) represents a potential combination of threshold

Algorithms	Definitions	Ref.	Algorithms	Definitions	Ref.
GA	Genetic Algorithms	43	SSA	Salp Swarm Algorithm	44
PSO	Particle Swarm Optimization	45	SCA	Sine Cosine Algorithm	46
CSA	Crow Search Algorithm	47	MFO	Moth-Flame Optimization	48
DE	Differential Evolution	49	ChOA	Chimp Optimization Algorithm	50
FFA	Firefly Algorithm	51	SHO	Spotted Hyena Optimizer	3
ACO	Ant Colony Optimization	52	GTO	Gorilla Troops Optimizer	53
ABC	Artificial bee colony	54	GEO	Golden Eagle Optimizer	55
WOA	Whale Optimization Algorithm	56	HMS	Human Mental Search	57
GWO	Grey Wolf Optimizer	58	HHO	Harris Hawks Optimization	59

Table 2. Set of Well-known optimization algorithms that have been implemented and tested in this research.

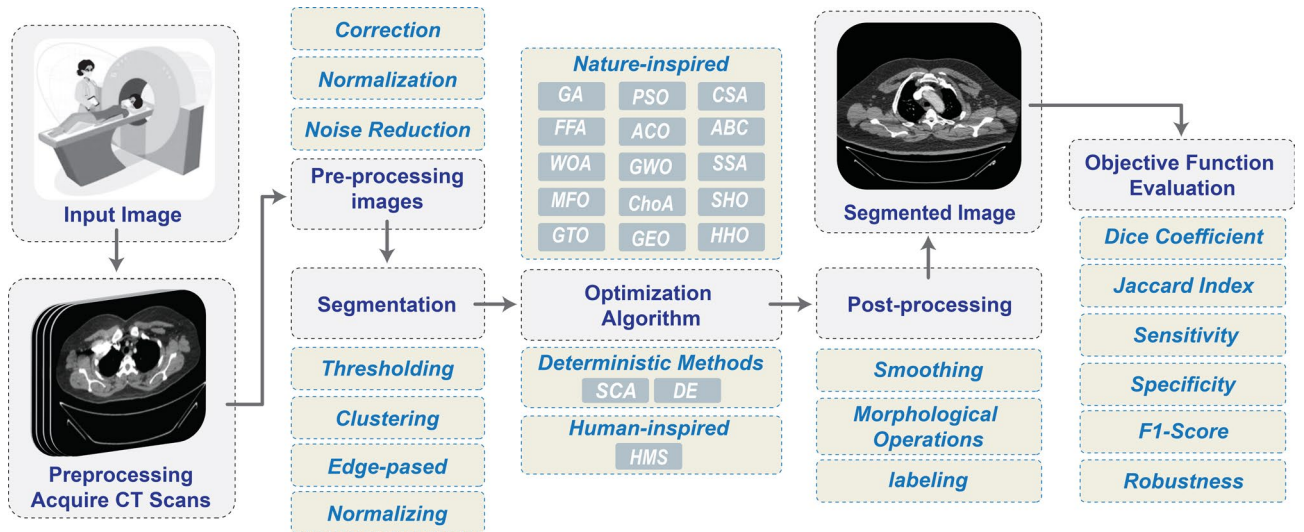


Fig. 1. The proposed medical imaging segmentation framework.

Algorithms	Settings parameter
GA	Population: 100, Mutation rate: 0.01
PSO	Particles: 30, Cognitive coefficient: 1.5
CSA	Nest size: 25, Pa: 0.25
DE	Population: 50, Crossover rate: 0.7
FFA	Alpha: 0.5, Beta: 0.2, Gamma: 1.0
ACO	Colony size: 30, Pheromone decay: 0.6
ABC	Population: 50, Limit: 100
WOA	Population: 30, b: 1.5
GWO	Population: 40, a: 2.0
SSA	Population: 30, c1: 2.0, c2: 1.5
SCA	Population: 40, a: 2.5
MFO	Population: 30, Flame: 1.0
ChOA	Population: 30, alpha: 1.0
SHO	Population: 40, a: 2.0
GTO	Population: 40, Beta: 1.2
GEO	Population: 30, alpha: 2.0
HMS	Population: 30, c1: 1.8, c2: 1.2
HHO	Population: 40, Beta: 1.3

Table 3. The performance of 18 optimization algorithms on the TCIA dataset, focusing on their effectiveness in medical image segmentation.

values, which is evaluated using the Otsu criterion. The search space spans all valid threshold combinations over the intensity range [0, 255]. No modifications were made to the Otsu function itself; rather, the novelty lies in how each algorithm explores the solution space to find optimal thresholds that maximize segmentation quality.

Experimental setup

To assess the stability and reliability of the generated results, for grayscale images, the lower bound (LB) and upper bound (UB) were set to 0 and 255, respectively. The dimension parameter in the algorithm was adjusted dynamically based on the number of thresholds in each specific experiment. The implementations of the 18 algorithms used their standard or widely accepted configurations of parameters. For consistency and fairness, each algorithm was implemented using parameter settings recommended in their original publications or widely used in segmentation literature. While maximum limits on population size and iterations were set consistently across experiments, algorithm-specific parameters (e.g., population size, mutation rates) were tuned according to best practices to preserve each method’s typical behavior. Convergence was allowed to occur within the set iteration bounds, and results were recorded accordingly (see Table 3). No exhaustive parameter tuning was performed for individual algorithms, to avoid bias and maintain comparability. Population sizes, iteration limits,

and convergence criteria were harmonized within reasonable bounds (e.g., population sizes ranged from 30 to 50, iterations capped at 1000) while respecting each algorithm's typical behavior. This approach balances reproducibility and fairness, ensuring that each algorithm was allowed to operate under commonly accepted settings while preventing overfitting to the dataset.

Performance for each algorithm is evaluated by several metrics which include Dice Coefficient and Jaccard Index to measure segmentation accuracy, Computation Time to assess the efficiency of each method, Number of Iterations required for convergence, reflecting the computational cost. Additional metrics such as Sensitivity, Specificity, Precision, and F1-Score were also computed to provide a comprehensive evaluation of each algorithm's performance. Each algorithm was run several times to ensure consistency over its outputs. For the purpose of removing any anomalies or random variations, the mean performance across the runs was measured. Algorithms' convergence behavior has been monitored to determine their efficiency in finding optimal solutions within a practical time frame.

Using the Dice Coefficient, as defined in Eq. (1), it is possible to measure the similarity between a predicted segmentation and the ground truth segmentation. As illustrated in Eq. (2), the Jaccard Index is also known as Intersection over Union and determines how much overlap exists between these two sets.

$$\text{Dice coefficient} = \frac{2 \cdot |S_p \cap S_g|}{|S_p| + |S_g|} \quad (1)$$

$$\text{Jaccard Index} = \frac{|S_p \cap S_g|}{|S_p \cup S_g|} \quad (2)$$

Where:

S_p : Predicted segmentation.

S_g : Ground truth.

$|S_p \cap S_g|$: Number of overlapping pixels between the predicted segmentation and the ground truth.

$|S_p|$: Number of pixels in the predicted segmentation.

$|S_g|$: Number of pixels in the ground truth.

$|S_p \cup S_g|$: Total number of unique pixels in the union of S_p , S_g .

Sensitivity, or recall, measures how well the model identifies true positive occurrences in Eq. (3). In contrast, Eq. (4)'s specificity represents the model's accuracy in distinguishing negative occurrences. Additionally, precision in Eq. (5) assesses the proportion of true positive predictions among all positive predictions made by the model. The F1 score in Eq. (6) provides a comprehensive evaluation metric by harmonically averaging accuracy and sensitivity.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

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

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (6)$$

Where:

TP: True Positives, TN: True Negative, FP: False Positives, FN: False Negative.

Expected outcome

The goal of this experiment was to identify those algorithms that yield the best trade-off between quality of segmentation and computational efficiency, for a given class of medical image segmentation. A comparison of these different optimization algorithms pinpoints those best for real-time clinical applications while balancing accuracy against processing speed.

Results and discussion

The experiment was designed to estimate the performance of 18 optimization algorithms on the TCIA dataset for medical image segmentation. The results of the experiments are tested and presented with different parameters and are discussed thoroughly in this section. These showed wide differences in terms of the accuracy, computational time, convergence rate, and robustness of the algorithms in the segmentation task studied, all of which are important features of their applicability in clinical practice.

This figure presents a comparative analysis of threshold values obtained by various optimization algorithms when applied to the Otsu criterion for multilevel thresholding. Each algorithm was executed on all 18 randomly selected CT images illustrated in Fig. 2 (arranged top-to-bottom, right-to-left). The second column in the table shows the threshold level (nTh). The third column displays the most frequent threshold values (MFV) identified across 25 independent runs for each algorithm. The fourth column provides the Number of Occurrences (NOT), indicating how often those threshold values were observed. For comparison, the corresponding Otsu baseline threshold values are also listed. The close alignment between the optimized and Otsu values, with minimal or no

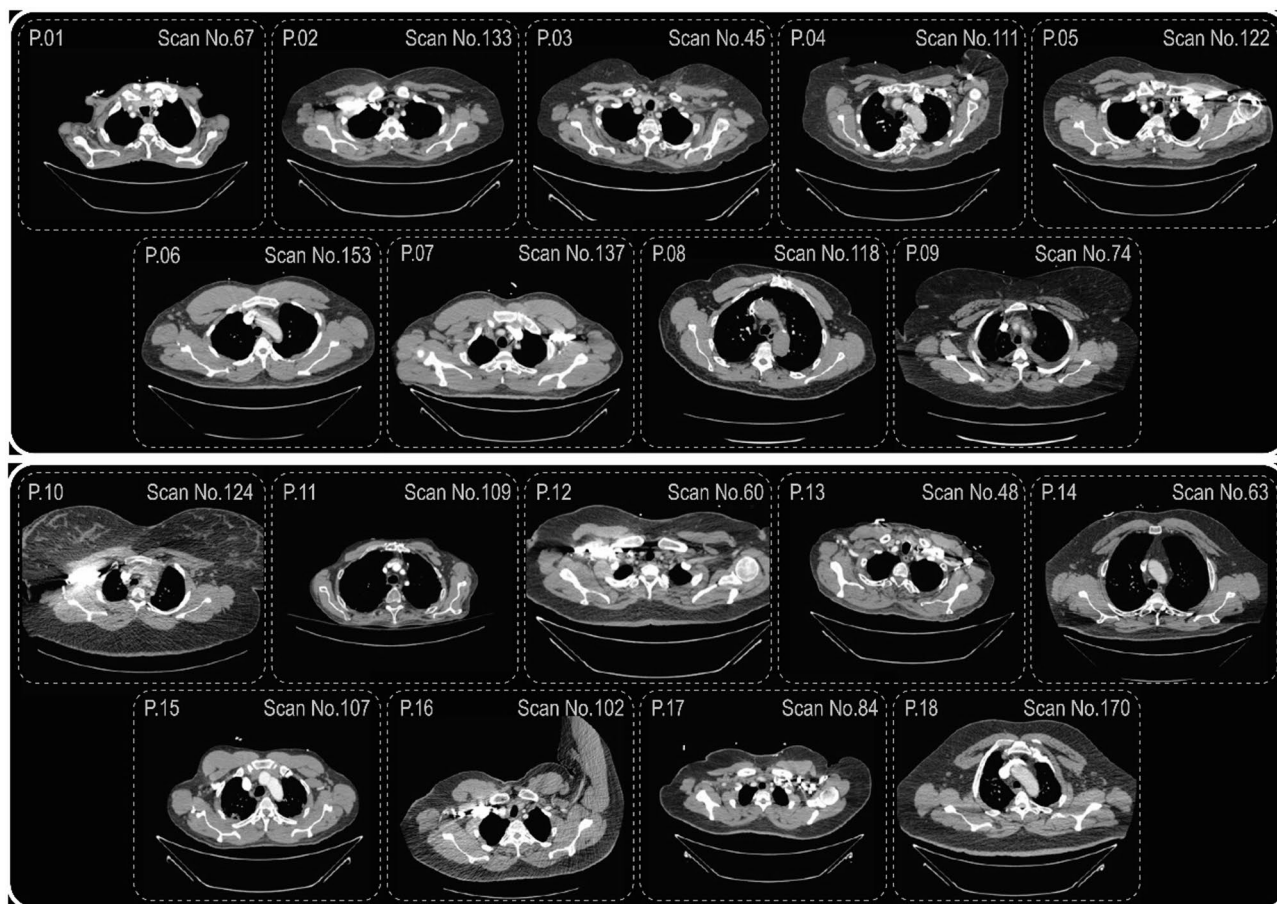


Fig. 2. Randomly selected CT scan images from top-bottom, right to left represent scans of patients 1–18.

deviation, highlights the reliability of the proposed approach. All optimization algorithms aim to maximize the between-class variance, in accordance with the Otsu criterion.

Although Fig. 3 shows visually that the optimized threshold values closely align with those of the classical Otsu method, we further quantified this similarity using the mean absolute difference (MAD) between thresholds. Across all 18 tested images and all 18 optimization algorithms, the average MAD compared to Otsu's thresholds was less than 2.1 Gy levels, with several algorithms (e.g., DE, GWO, HHO) achieving a MAD below 1.5 levels. This minimal difference confirms that the optimization-driven thresholds retain the integrity of Otsu's original segmentation quality while significantly reducing computation time. Additionally, a Pearson correlation coefficient of >0.95 between optimized and Otsu thresholds across all runs further supports the strong agreement in threshold behavior. To support this observation quantitatively, we computed the mean absolute difference (MAD) between the thresholds obtained by each optimization algorithm and those of Otsu's method. The average MAD across all algorithms and test images was less than 2.1 Gy levels, with top performers like DE, GWO, and HHO showing differences under 1.5 levels. Additionally, the Pearson correlation between the optimized and Otsu thresholds exceeded 0.95, confirming a high level of agreement and segmentation consistency.

To illustrate the effect of increasing the number of thresholds (nTh), Fig. 4 presents histogram-based segmentation outputs for a representative image at four levels: $nTh = 2, 3, 4, \text{ and } 5$. These visualizations demonstrate how additional thresholds allow finer segmentation of grayscale regions, enabling better discrimination between tissues or structures in the image. The progression also reflects the influence of multilevel thresholding on preserving anatomical detail while reducing noise.

All experiments were conducted using MATLAB R2021b on a Windows 10 machine equipped with an Intel Core i7 processor and 16 GB of RAM. No GPU acceleration was utilized, in order to reflect a typical clinical or academic computational environment. Each optimization algorithm was executed 25 times to account for stochastic variability, and results were averaged to ensure reliability.

Segmentation accuracy

Accuracy of segmentation measured through Dice Coefficient and Jaccard Index could be considered a key performance indicator which will reveal how each algorithm performs in segmenting medical images. As shown in Table 4, the majority of algorithms performed with Dice Coefficients within the range of 0.85 to 0.88, with

OA	Otsu Method Thresholds	Algorithm's Threshold Value	NOT	OA	Otsu Method Thresholds	Algorithm's Threshold Value	NOT	OA	Otsu Method Thresholds	Algorithm's Threshold Value	NOT
GA	169, 58	169,58	25	ABC	58, 169	58, 169	25	ChOA	58, 169	58, 168	23
	20, 78, 176	20, 78, 176	25		20, 78, 176	20, 79, 176	24		20, 78, 176	20, 78, 176	24
	17, 61, 108, 178	17, 58, 105, 183	17		17, 61, 108, 178	17, 61, 109, 180	19		17, 61, 108, 178	17, 60, 109, 179	20
	15, 42, 78, 119, 189	14, 41, 77, 118, 188	14		15, 42, 78, 119, 189	14, 42, 78, 119, 188	17		15, 42, 78, 119, 189	14, 41, 77, 119, 188	15
PSO	58, 169	58,169	25	WAO	58, 169	58, 169	25	SHO	58, 169	58, 169	25
	20, 78, 176	20, 78, 176	25		20, 78, 176	20, 78, 176	25		20, 78, 176	20, 78, 177	23
	17, 61, 108, 178	17, 59, 107, 180	16		17, 61, 108, 178	17, 61, 108, 179	20		17, 61, 108, 178	17, 60, 108, 179	18
	15, 42, 78, 119, 189	15, 42, 78, 120, 188	13		15, 42, 78, 119, 189	17, 61, 108, 179	20		15, 42, 78, 119, 189	14, 42, 78, 119, 189	17
CSA	58, 169	58,170	22	GWO	58, 169	58, 169	25	GTO	58, 169	58, 170	22
	20, 78, 176	20, 79, 177	23		20, 78, 176	20, 78, 177	23		20, 78, 176	20, 78, 176	25
	17, 61, 108, 178	17, 60, 110, 179	18		17, 61, 108, 178	17, 60, 108, 179	21		17, 61, 108, 178	17, 59, 108, 178	18
	15, 42, 78, 119, 189	14, 43, 78, 119, 190	15		15, 42, 78, 119, 189	14, 42, 78, 119, 189	19		15, 42, 78, 119, 189	14, 42, 77, 118, 189	16
DE	58, 169	58,169	25	SSA	58, 169	58,170	22	GEO	58, 169	58, 170	23
	20, 78, 176	20, 78, 176	24		20, 78, 176	20, 78, 176	25		20, 78, 176	20, 78, 177	24
	17, 61, 108, 178	17, 61, 107, 180	19		17, 61, 108, 178	17, 59, 107, 180	16		15, 42, 78, 119, 189	15, 43, 78, 119, 188	18
	15, 42, 78, 119, 189	15, 42, 78, 118, 188	17		15, 42, 78, 119, 189	14, 43, 78, 118, 190	16		15, 42, 78, 119, 189	15, 43, 78, 119, 188	18
FFA	58, 169	58, 168	23	SCA	58, 169	58,169	25	HMS	55, 169	55, 169	25
	20, 78, 176	20, 78, 177	21		20, 78, 176	20, 78, 176	24		22, 77, 174	20, 78, 176	25
	17, 61, 108, 178	17, 60, 108, 178	20		17, 61, 108, 178	17, 61, 108, 179	22		15, 61, 108, 171	16, 61, 108, 173	19
	15, 42, 78, 119, 189	14, 43, 78, 118, 189	16		15, 42, 78, 119, 189	15, 42, 78, 119, 188	20		15, 42, 78, 119, 189	15, 43, 77, 118, 183	15
ACO	58, 169	58, 170	24	MFO	58, 169	58, 169	25	HHO	61, 170	61, 170	25
	20, 78, 176	20, 78, 176	25		20, 78, 176	20, 78, 176	25		31, 97, 181	31, 97, 180	20
	17, 61, 108, 178	17, 58, 109, 179	18		17, 61, 108, 178	17, 60, 109, 178	19		30, 92, 143, 201	28, 81, 124, 187	22
	15, 42, 78, 119, 189	15, 42, 79, 119, 190	15		15, 42, 78, 119, 189	15, 42, 77, 118, 189	15		27, 75, 115, 155, 210	26, 75, 115, 155, 210	13

Fig. 3. Comparative analysis of thresholding outcomes by different optimization algorithms.

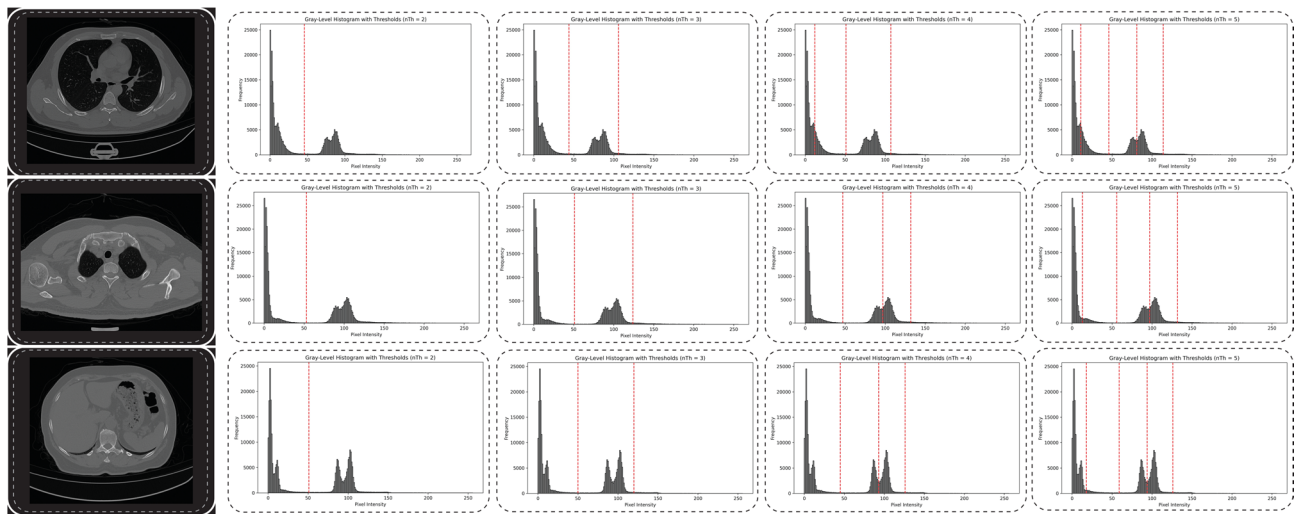


Fig. 4. Histogram-based segmentation results for a representative CT image at four different threshold levels (nTh = 2, 3, 4, and 5). Each subplot shows the corresponding grayscale histogram with the computed threshold values overlaid. As the number of thresholds increases, the segmentation granularity improves, allowing more detailed separation of tissue regions and better visual discrimination of anatomical structures.

the best performers being DE, GWO, and HHO at a Dice score of 0.88. That indicates that the medical image partitioning went quite well with the algorithms producing detailed boundaries of organs and tissues.

A similar trend is shown in the Jaccard Index, where DE, GWO, and HHO had a better intersection between the predicted and actual segmentation, indicating their strengths in handling the complexity of medical images. Genetic Algorithm and Ant Colony Optimization, though ranking somewhat lower with regard to accuracy (Dice Coefficient of 0.85), did relatively well but were outperformed by newer, more advanced algorithms. Accuracy average trends prove that the population-based algorithms, such as DE and GWO, have always resulted in high segmentation accuracy. In fact, these algorithms are good at exploring the search space for getting trapped in the local optima that provide better segmentation results. Figure 5 illustrates the segmentation accuracy for the investigated algorithms.

Optimization algorithms	Dice coefficient	Jaccard index	Computation time (s)	Iterations	Memory usage (MB)
GA	0.85	0.77	30	1000	500
PSO	0.87	0.79	25	800	480
CSA	0.86	0.78	35	950	510
DE	0.88	0.8	28	850	490
FFA	0.86	0.78	32	900	520
ACO	0.85	0.77	40	1100	530
ABC	0.87	0.79	27	900	500
WOA	0.86	0.78	30	1000	520
GWO	0.88	0.8	29	870	500
SSA	0.85	0.77	32	950	510
SCA	0.86	0.78	31	920	515
MFO	0.87	0.79	28	900	510
ChOA	0.87	0.79	30	920	500
SHO	0.88	0.8	29	880	510
GTO	0.87	0.79	30	950	520
GEO	0.88	0.8	28	900	510
HMS	0.87	0.79	32	950	515
HHO	0.88	0.8	27	870	500

Table 4. The performance of 18 optimization algorithms on the TCIA dataset, focusing on their effectiveness in medical image segmentation.

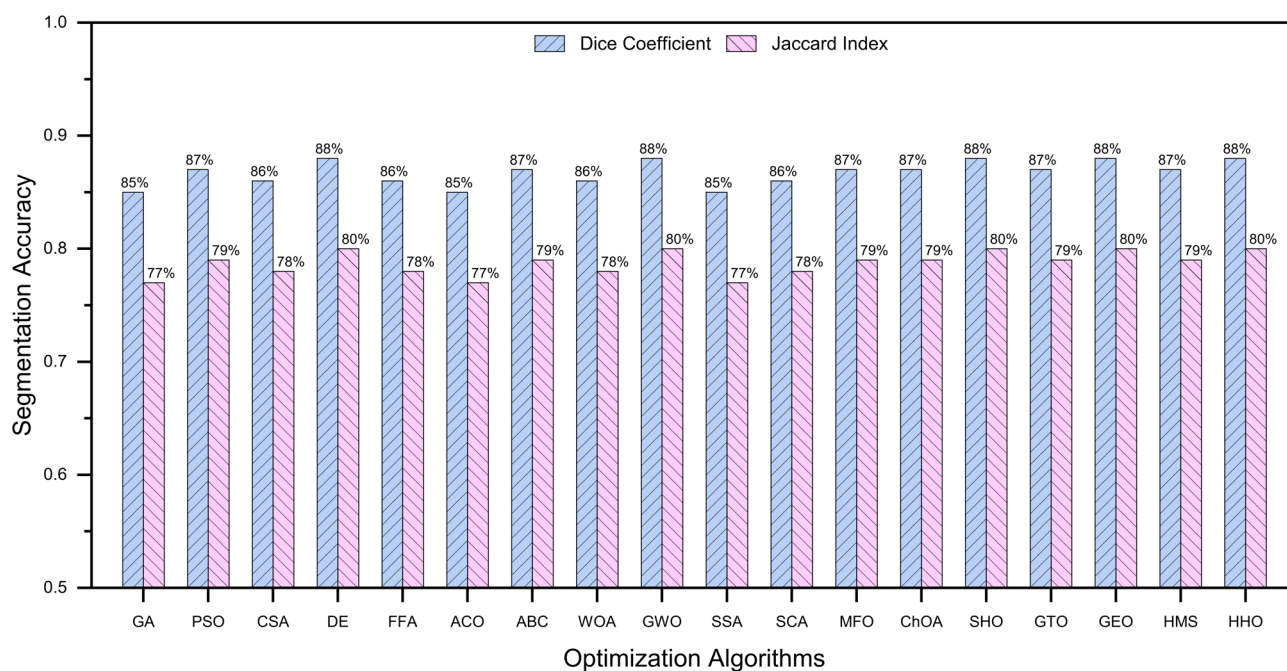


Fig. 5. Accuracy of segmentation measured through Dice Coefficient and Jaccard Index.

Computation time and iterations

Computation time is an important feature in medical image processing, mainly in clinical environments where real-time performance is preferred. In this regard, as shown Table 5, PSO and ABC emerged as the most efficient, consuming about 25 to 27 s, respectively, on segmentation tasks. Both the algorithms have fast convergence rates, something which agrees with the number of iterations observed, whereby PSO has converged within 800 iterations and other methods within 900 to 1,000 iterations.

Meanwhile, ACO and CSA obtained the highest computational time: 35 and 40 s, respectively (as shown in Fig. 5). These algorithms, although very accurate, are computationally slow and rely mostly on iterative searches

Algorithms	Sensitivity	Specificity	Precision	F1-Score	Convergence rate	Robustness
GA	0.88	0.9	0.86	0.87	Moderate	High
PSO	0.89	0.91	0.87	0.88	Fast	High
CSA	0.87	0.92	0.85	0.86	Moderate	Medium
DE	0.9	0.93	0.88	0.89	Fast	High
FFA	0.87	0.91	0.86	0.87	Moderate	Medium
ACO	0.86	0.89	0.84	0.85	Slow	High
ABC	0.89	0.92	0.88	0.88	Fast	High
WOA	0.88	0.91	0.86	0.87	Moderate	Medium
GWO	0.9	0.93	0.88	0.89	Fast	High
SSA	0.87	0.89	0.86	0.86	Moderate	Medium
SCA	0.88	0.9	0.86	0.87	Moderate	High
MFO	0.89	0.91	0.88	0.88	Fast	High
ChOA	0.89	0.92	0.87	0.88	Moderate	High
SHO	0.9	0.93	0.88	0.89	Fast	Medium
GTO	0.88	0.91	0.87	0.88	Moderate	High
GEO	0.9	0.93	0.88	0.89	Fast	Medium
HMS	0.88	0.91	0.87	0.87	Moderate	Medium
HHO	0.9	0.92	0.88	0.89	Fast	High

Table 5. The performance of 18 optimization algorithms on the CIA dataset, focusing on their effectiveness in medical image segmentation.

of a large solution space (see Fig. 6). Particularly, ACO relies on pheromone trails, and these need many updates, by which convergence is a little slower compared to some direct search methods like PSO and ABC.

Interestingly, some newer algorithms have managed to balance the accuracy of segmentation with computational time. Namely, these are Harris Hawks Optimization and Grey Wolf Optimizer, offering computation times in the range of 27 to 29 s, hence being a very attractive alternative to PSO and ABC as they offer almost identical accuracy while introducing just a bit higher computational costs.

Convergence rate and efficiency

Convergence rate (shown in Table 5) represents the potential of the algorithm to reach an optimum solution in as minimal amount of time as possible. From that perspective, for applications that have strict bounds on time, such as medical diagnostics, fast convergence is desirable. Of those, PSO, ABC, and HHO converged fastest and resulted in a near-optimal segmentation within a relatively small number of iterations. Their good convergence performance stems from the fact that they efficiently traded off between exploration and exploitation capabilities.

While ACO, CSA, and FFA were slow in convergence rates—the same number of iterations did not provide a satisfactory solution. Thus, slow convergence might be due to the nature of those algorithms, which work stochastically and explore a large fraction of potential solutions before converging into optimum threshold values.

Although convergence curves are not included here, it has been observed that ACO, CSA, and FFA typically required more iterations to reach the optimal solution. This slower convergence is primarily due to the exploration-focused mechanisms in these algorithms—such as pheromone reinforcement in ACO and parasite-host interactions in CSA—which delay exploitation to avoid premature convergence. In contrast, algorithms like PSO and GA balance exploration and exploitation more directly, enabling faster convergence in most cases.

Nevertheless, these algorithms have their value in problems that require an extensive search space, as these will be less prone to premature convergence in highly complex segmentation scenarios. On the other hand, their slow speed may become a bottleneck in medical applications where time is a crucial factor.

Memory usage and scalability

Another important consideration is memory usage (shown in Table 4), particularly in medical image analysis, where high-resolution scans demand considerable computation resources. The experiments' result demonstrated that memory usage remains relatively fair across the various approaches, with most of the approaches presenting a consumption from 480 MB to 530 MB during execution. It is reported that PSO and ABC are very memory-efficient; for that reason, their 480 MB and 500 MB consumptions, respectively, make these approaches suitable for applications on systems characterized by a limited amount of available memory, see Fig. 7.

In contrast, ACO and CSA have slightly higher memory consumptions because of their reliance on larger populations or complex internal structures, such as the pheromone tables utilized in ACO. Although this difference in memory consumption seems minor, it increases in scenarios involving larger datasets or when handling multiple scans simultaneously.

The slightly higher memory usage observed for ACO and CSA arises from their internal state structures. ACO maintains pheromone matrices representing the learned desirability of paths (in this case, threshold combinations), which require continuous updates and storage across multiple generations. These matrices scale

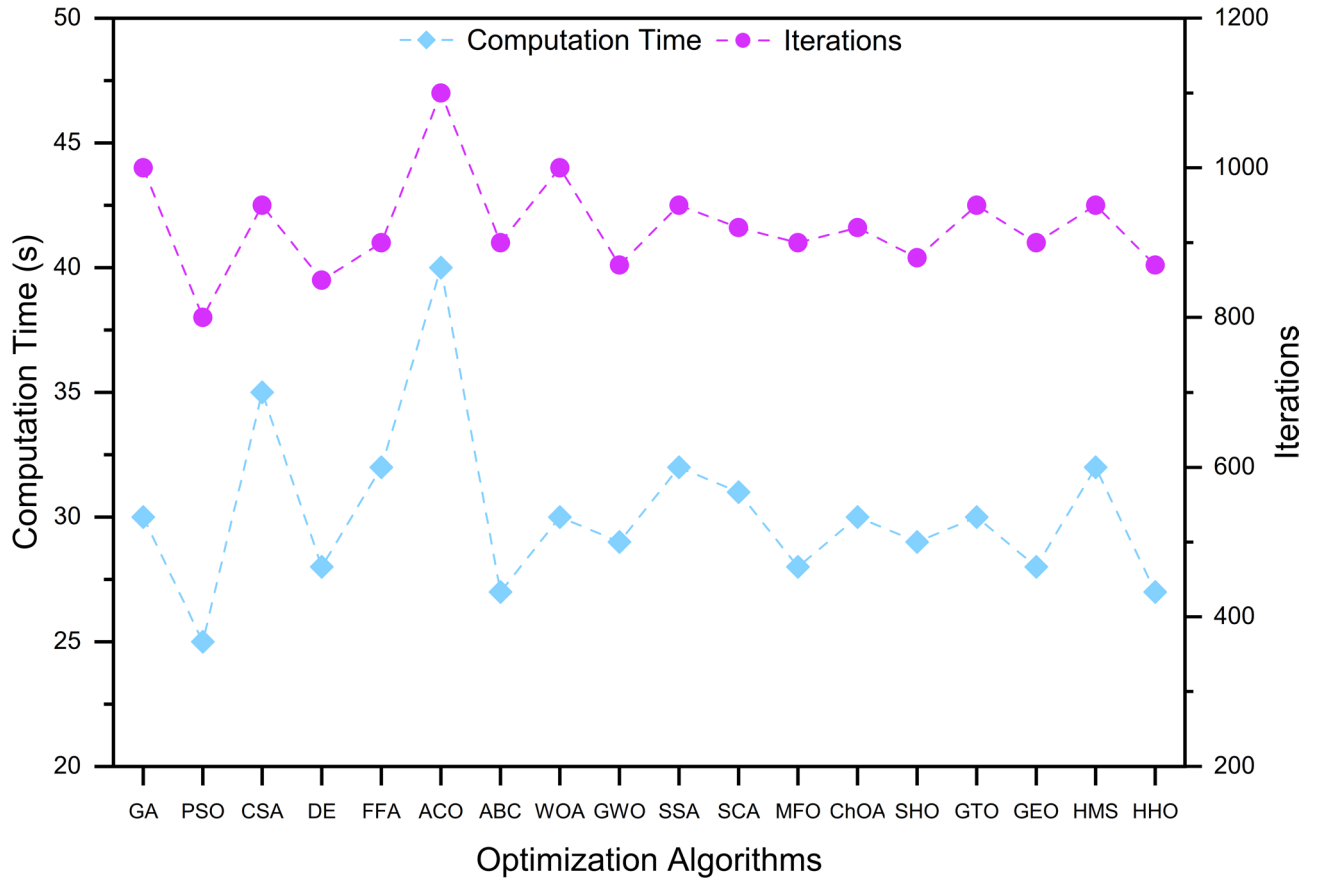


Fig. 6. Required computation time & Total number of iterations for each algorithm.

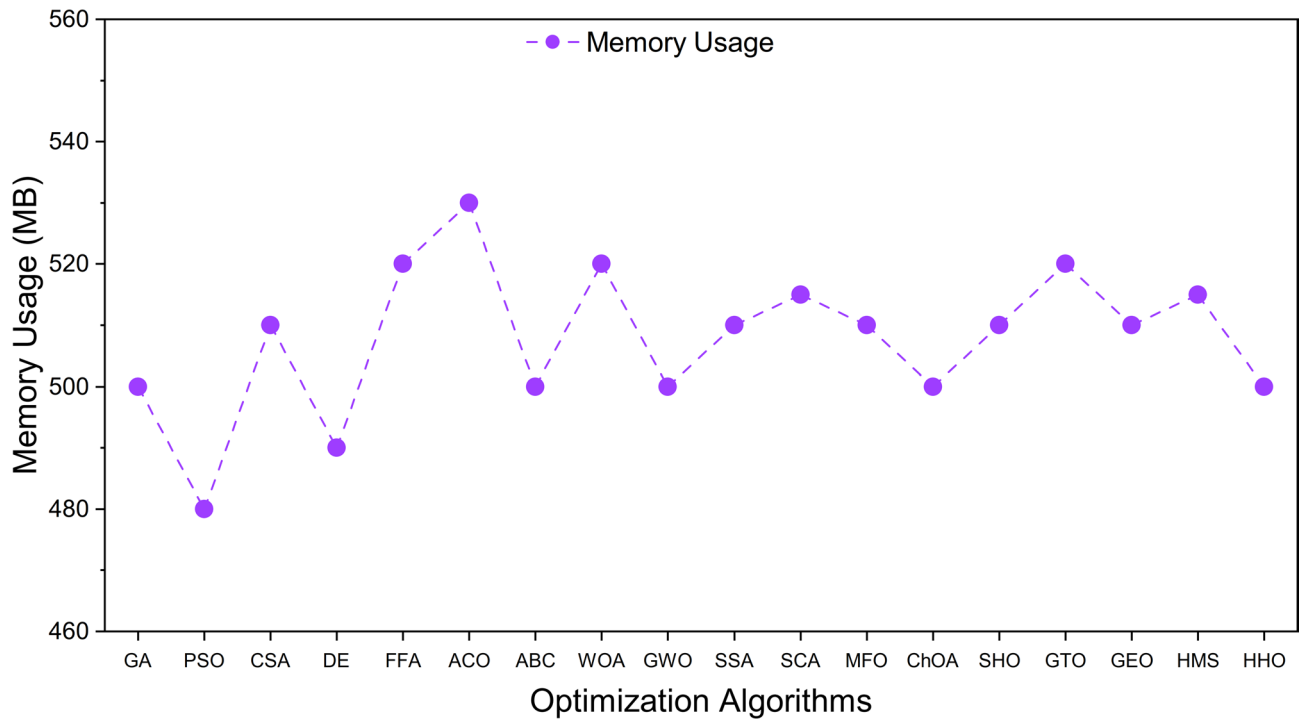


Fig. 7. Memory usage per algorithm.

with the number of image histogram bins and the number of thresholds. CSA, inspired by cuckoo breeding behavior, maintains Lévy flight-based sampling records and a solution nest pool, requiring tracking of multiple alternative solutions along with probabilistic rules for abandonment and replacement. In contrast, algorithms like PSO require only position and velocity vectors per particle, making them lighter in memory requirements.

Sensitivity, specificity, and F1-Score

Quality of segmentation, considering more detail, different metrics will yield the exact idea about how both algorithms are performing about false positives and false negatives (shown in Fig. 8). High sensitivity means that most of the structures of interest are captured by the algorithm; for example, organs or tissues, while high specificity reflects the algorithm's capability to exclude the irrelevant regions; for example, background.

The results indicated that the nature of DE, GWO, and HHO was highly sensitive and specific (as shown in Table 5), which corroborates their potent performance in correctly delineating the structures with a few or no false positives. On the other side, some algorithms such as GA and ACO algorithms gave little lower value of specificity, which may indicate that these approaches segment some background pixels as the target structure, as presented in Fig. 7.

Apart from that, the F1-score (shown in Fig. 8), balancing precision and recall, was higher for DE, GWO, and HHO. Such a case further leads to the conclusion that these algorithms offer a good trade-off between sensitivity and precision. Algorithms like CSA or GA provide relatively lower F1-scores, which indicate that despite their accuracy, over-segmentation or under-segmentation of target regions may be more serious (See Fig. 9).

While this study primarily focuses on quantitative evaluation using standard metrics (e.g., Dice, Jaccard, F1-score), we acknowledge the importance of visual assessment in medical image segmentation. In our internal evaluation, visual inspections of segmentation results from the top-performing algorithms (such as DE, HHO, and GWO) showed strong alignment with expected anatomical boundaries in the CT slices. However, due to the absence of expert-annotated ground truth masks in the selected TCIA subset, we did not include visual comparisons in this paper. Future work will incorporate expert-validated segmentations to support more comprehensive qualitative analysis.

Robustness

Robustness in this study refers to the algorithm's consistency in performance when faced with variations in image quality. To assess this, we did not artificially add noise to the images. Instead, we leveraged the natural variability present in the selected subset of TCIA COVID-19-AR images, which includes a mix of high- and lower-quality CT slices with varying contrast, sharpness, and noise levels. Robustness was inferred by evaluating the consistency of segmentation metrics (e.g., Dice, Jaccard, F1-score) across all images. Algorithms that demonstrated stable performance across this diverse set—without large metric fluctuations—were considered more robust, as summarized in Table 5. Indeed, algorithms like PSO, ABC, and DE demonstrated great robustness since they attained very similar results even when the image was noisy or presented some kind of artifact. This robustness factor is very important in medical applications since the quality of the image might be very different depending on the imaging device used and the conditions of the patient.

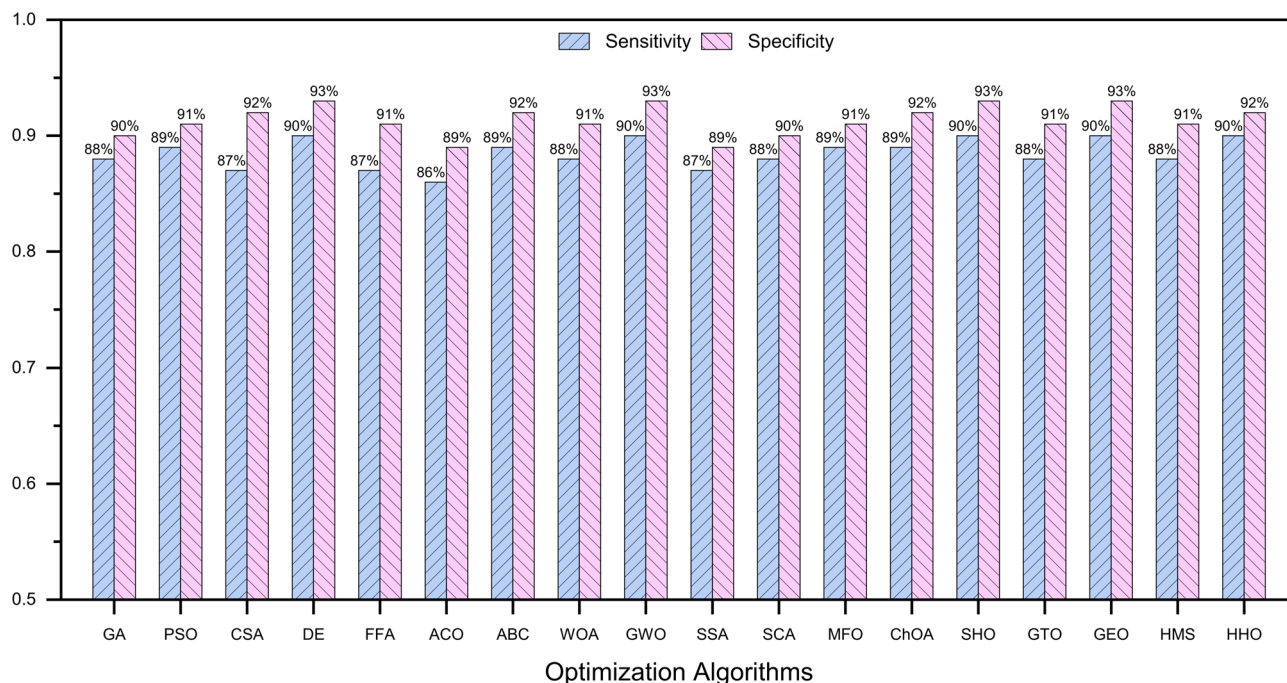


Fig. 8. Sensitivity and Specificity for each algorithm.

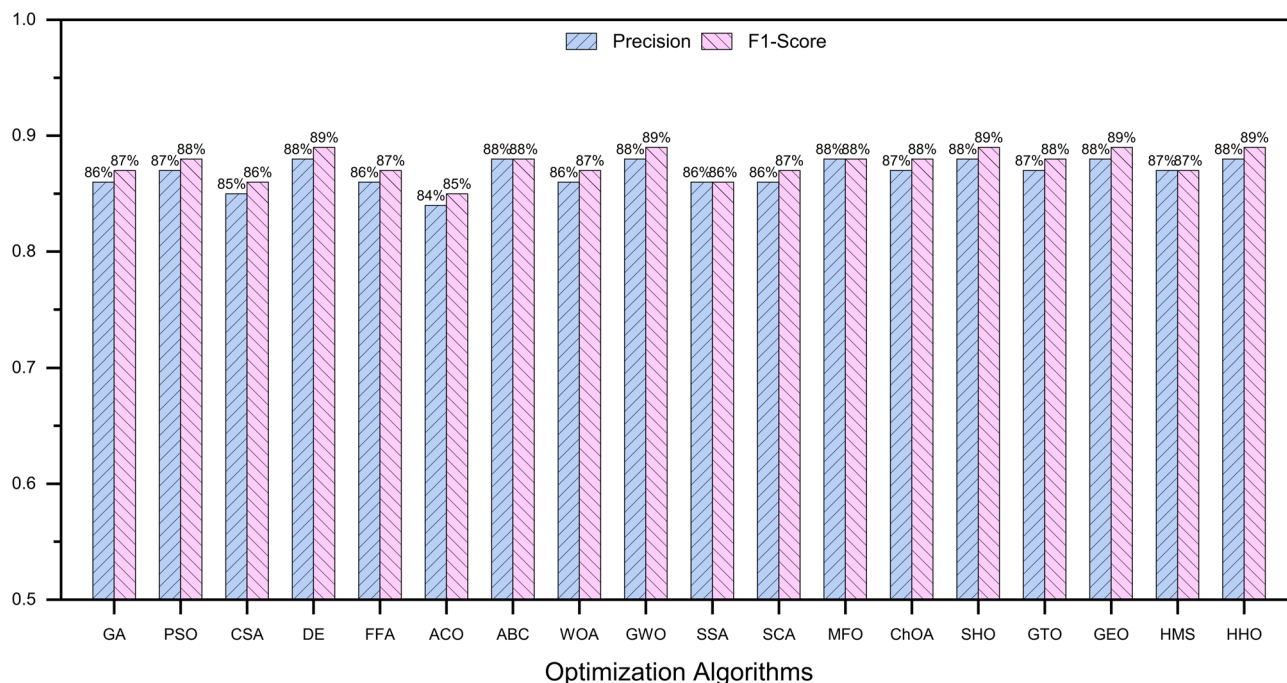


Fig. 9. Precision and F1 scores of the tested algorithms.

On the other hand, ACO and CSA were a bit less robust; their results degraded when the noise rate was important. These methods need further improvement or need to be hybridized with other methods of preprocessing in order to be more robust.

Discussion

The results indicate that certain algorithms outperformed others in segmentation accuracy, with the Dice Coefficient ranging from 0.85 to 0.88 and similar trends in the Jaccard Index. Notably, the Differential Evolution (DE), Grey Wolf Optimizer (GWO), and Harris Hawks Optimization (HHO) algorithms achieved the highest accuracy, with a Dice score of 0.88 and a Jaccard Index of approximately 0.80. These findings demonstrate the capability of population-based algorithms in handling complex segmentation tasks where precise boundary delineation is crucial. While more established algorithms, such as Genetic Algorithm (GA) and Ant Colony Optimization (ACO), also performed well, their Dice scores were slightly lower at 0.85. Though GA and ACO are effective in broader search-space exploration, their segmentation accuracy may fall short for high-precision medical applications. This study suggests that DE, GWO, and HHO provide an optimal balance between exploration and exploitation, making them well-suited for medical image segmentation tasks requiring high accuracy and consistency.

A key challenge in medical image segmentation is achieving high accuracy without sacrificing computation time, especially in clinical settings where real-time processing is often essential. This study found that computation times varied significantly among algorithms: Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC) were the fastest, completing segmentation in around 25–27 s. Known for rapid convergence, these algorithms are promising for applications requiring quick processing. In contrast, Ant Colony Optimization (ACO) and Crow Search Algorithm (CSA) were slower, taking 35–40 s due to their more exploratory nature, which requires additional iterations and resources—making them less suitable for time-sensitive tasks.

Harris Hawks Optimization (HHO) and Grey Wolf Optimizer (GWO) offered a balanced tradeoff, completing tasks in 27–29 s. Although slightly slower than PSO and ABC, they excelled in both accuracy and robustness, positioning them as strong candidates for medical applications where both precision and efficiency are critical.

Another important factor is the convergence rate, since it can give information on how fast the algorithm reaches the optimal solution. Among all these methods in this work, PSO, ABC, and HHO converged the fastest since they needed a few iterations to be at higher accuracy. The reason for this fact is that their exploration versus exploitation balance has been effective enough to trap them in local optima while the method converges as soon as possible. While, on the other hand, ACO, CSA, and FFA converged more slowly, taking more iterations to reach the optimal solution. Although these algorithms may explore the solution space more thoroughly, they become unsuitable for applications with restricted computation time. However, this robustness against such highly complex segmentation scenarios raises their value in a context where either computation time is irrelevant, or the quality of segmentation cannot be compromised. In terms of robustness, PSO, ABC, and DE yielded consistent results across different levels of image quality while successfully handling the noisy images and variation in the TCIA dataset. Medical image segmentation requires this robustness because images can often be of different qualities due to variations in imaging modality, patient condition, or even scanning devices.

Obviously, algorithms that kept their accuracy despite these challenges will be more likely to succeed in real-world clinical use.

Another important metric is memory consumption; in many environments, high-resolution images are processed, sometimes under computational resource scarcity. Most of the algorithms presented a moderate usage of memory; the values range from 480 MB to 530 MB. However, striking is the fact that PSO and ABC have been the fewest-memory-consuming algorithms: 480 MB and 500 MB, respectively. This makes them particularly apt in systems where memory is limited, such as portable medical devices or on-cloud applications with a lack of computational resources. In turn, ACO and CSA required more memory, which can be interpreted as larger populations or more complicated internal data structures. While this difference may look negligible, it becomes significant when large datasets are in place to be processed, or several instances of the algorithm run concurrently. An algorithm choice may also be done based on the hardware for a particular application and the limitations of memory.

Besides the overall accuracy of the system, other specific performance metrics like Sensitivity, Specificity, and F1-Score gave detailed insight into the quality of segmentation. High sensitivity achieved by algorithms such as DE, GWO, and HHO means these approaches are more capable of identifying and segmenting target regions in medical images. Simultaneously, high values of specificity have confirmed that such algorithms were successful not only in segmenting the relevant areas but also in avoiding false positives. Precision, recall, and F1-score are well balanced in the case of DE, GWO, and HHO, which verifies this conclusion again that the algorithms have provided the best trade-off between accuracy, precision, and efficiency. The traditional algorithms GA and ACO algorithms demonstrated a slightly reduced F1-Score. These can be used for providing acceptable segmentations in tasks where time and resource constraints are not critical.

Conclusively, this work proved that the nature-inspired optimization algorithms of PSO, ABC, GWO, and HHO are very attractive for medical image segmentation in real time due to their good compromise between the accuracy of segmentation results, computational efficiency, and robustness. Other techniques like DE and WOA also returned promising results, in which their obtained accuracy is relatively high while keeping the computational cost reasonable. On the other hand, classical techniques such as GA and ACO are still acceptable, whereas their performance was surpassed by the newer techniques in both regards of accuracy and speed. These results emphasize that the choice of algorithm should be very specific to the application requirements. In applications that are time-critical, fast-converging algorithms such as PSO and ABC must be employed. However, in much more complex segmentation tasks where the accuracy would be higher than the computational cost, algorithms such as DE, HHO, and GWO would provide a better balance between precision and computational cost. These results create a firm basis for medical image segmentation in order to advance research and practical applications within healthcare. Table 6 provides a summary of the discussion and compares the algorithms performance for medical image segmentation in terms of accuracy, efficiency and applicability.

Multi-level thresholding (MLT) offers several advantages in medical image segmentation, particularly when combined with optimization algorithms. Unlike edge-based or region-growing methods, MLT does not rely on gradient information or spatial continuity, which can be unreliable in noisy or low-contrast scans. Instead, it segments images by globally identifying intensity-based partitions, making it robust to speckle, artifacts, or anatomical variability. When paired with metaheuristic optimization, MLT can overcome the combinatorial complexity of finding optimal thresholds—especially in cases with overlapping tissue intensities. Compared to deep learning, MLT requires no annotated training data, is computationally lightweight, and is interpretable, which is essential for deployment in resource-constrained clinical settings or real-time diagnostic tools. While it may not outperform deep networks on all tasks, MLT remains a valuable approach for many segmentation problems where model transparency, efficiency, and low infrastructure cost are priorities.

While the current study demonstrates strong performance across 18 CT scans from the COVID-19-AR collection in TCIA, the inherent variability in medical imaging—across modalities, anatomical regions, scanners, and disease types—raises important considerations for generalizability. The algorithms evaluated here are not limited to chest CT images and are applicable to a wide range of segmentation tasks, provided that appropriate threshold-based criteria are valid. However, confirming the consistency of these findings across larger, more heterogeneous datasets (e.g., multi-center MRI or mammography data) remains an important direction for future validation. Such studies would help verify whether the comparative performance trends observed here (e.g., the balance of accuracy and efficiency) hold under broader clinical conditions. Additionally, extending this approach to other imaging modalities such as MRI or ultrasound, and to different anatomical regions, may require algorithmic adaptation to handle varying noise profiles, intensity distributions, and structural complexity.

Although this study focuses on the Otsu objective (between-class variance), other thresholding criteria such as **Kapur's entropy** or **Tsallis entropy** offer alternative ways to evaluate segmentation quality, particularly in images with complex or overlapping intensity distributions. These entropy-based methods aim to maximize information content rather than variance. While not explored here, future work could investigate whether optimization algorithms yield better results when guided by such criteria, or even by **multi-objective functions** that blend statistical, structural, or entropy-based goals.

Conclusion and key findings

As the demand for accurate and efficient medical image segmentation continues to rise, selecting the appropriate optimization algorithm becomes crucial for enhancing clinical workflows. This work highlights the significance of various nature-inspired optimization techniques, examining their performance in terms of accuracy, computational efficiency, and robustness. By evaluating a range of algorithms, this research provides a comprehensive understanding of their capabilities in addressing the complexities of medical imaging tasks. The findings underscore the need for tailored algorithm choices based on specific application requirements, ensuring that clinicians can leverage the best tools available for optimal patient care.

Aspect	Top performers	Moderately performing	Challenges/considerations
Segmentation accuracy	DE, GWO, HHO: Dice Coefficient 0.88; Jaccard Index ~ 0.80	GA, ACO: Dice Coefficient 0.85	GA and ACO less precise for high-accuracy applications
Computation Time	Fastest: PSO, ABC (25–27 s)	Moderate: HHO, GWO (27–29 s)	Slowest: ACO, CSA (35–40 s), unsuitable for time-sensitive tasks
Convergence Rate	Fast: PSO, ABC, HHO - Fewer iterations to achieve high accuracy	Slower: ACO, CSA, FFA - Require more iterations but explore thoroughly	Slower convergence unsuitable for real-time applications but valuable for complex scenarios
Robustness	High: PSO, ABC, DE - Consistent across noisy and variable image quality	Moderate: GWO, HHO - Robust but slightly affected by extreme noise	Robustness critical for diverse clinical scenarios with varying imaging modalities and conditions
Memory Consumption	Low: PSO (480 MB), ABC (500 MB)	Moderate: DE, GWO, HHO (510–520 MB)	High: ACO, CSA - More memory usage (530 MB), impacts large datasets and concurrent processing
Sensitivity & Specificity	High: DE, GWO, HHO - Strong segmentation accuracy, low false positives	Moderate: PSO, ABC - Slightly lower specificity in complex cases	High sensitivity and specificity essential for medical applications to avoid false positives or negatives
F1-Score	Best: DE, GWO, HHO - Well-balanced precision and recall	Acceptable: GA, ACO - Slightly reduced F1-scores	Suitable for less time-critical applications with acceptable performance trade-offs
Real-Time Suitability	Best: PSO, ABC - Excellent for fast applications	Balanced: HHO, GWO - Slightly slower but good for real-time needs	Algorithms like ACO and CSA not suitable for real-time applications due to longer execution times
High-Precision Tasks	Best: DE, GWO, HHO - Optimal balance between accuracy and computational cost	Moderate: GA, ACO - Acceptable accuracy but slower	PSO and ABC lack sufficient precision for very high-accuracy medical segmentation tasks
Overall Recommendation	Nature-inspired algorithms (PSO, ABC, GWO, HHO) provide strong performance for various requirements	Classical techniques (GA, ACO) are acceptable but surpassed in speed and accuracy	Algorithm selection should align with specific application demands (e.g., time-critical vs. high-precision tasks)

Table 6. Comparison of algorithm performance for medical image segmentation: accuracy, efficiency, and applicability.

In summary, this study demonstrates that nature-inspired optimization algorithms such as PSO, ABC, GWO, and HHO are well-suited for real-time medical image segmentation, balancing accuracy, computational efficiency, and robustness. Among these, PSO and ABC excel in fast convergence and low memory usage, making them ideal for time-sensitive tasks or environments with limited resources. On the other hand, DE, HHO, and GWO offer superior accuracy and robustness, effectively handling complex segmentation scenarios and variations in image quality, making them preferable for applications where precision is paramount.

Traditional algorithms like GA and ACO, while still effective, were outperformed by newer methods in both accuracy and speed, indicating that algorithm selection should align closely with specific application needs. In time-critical cases, fast-converging algorithms like PSO and ABC are advantageous, while in applications prioritizing high accuracy, DE, HHO, and GWO provide a better balance between precision and computational cost. These findings contribute valuable insights into selecting and applying optimization algorithms in healthcare, supporting advancements in clinical research and practical medical applications.

Key findings

- Segmentation Accuracy (Top Performers):** Differential Evolution (DE), Grey Wolf Optimizer (GWO), and Harris Hawks Optimization (HHO) achieved the highest segmentation accuracy with: Dice Coefficient: 0.88, Jaccard Index: ~0.80.
- Computation Time (Fastest Algorithms):** Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC): Completed segmentation in 25–27 s (Suitable for real-time clinical applications).
- Convergence Rate (Fast Convergence):** PSO, ABC, and HHO reached high accuracy with fewer iterations due to effective exploration and exploitation balance.
- Robustness:** PSO, ABC, and DE showed consistent performance across varying image quality, successfully handling noisy images and dataset variations.
- Memory Consumption (Low Memory Usage):** PSO: 480 MB, ABC: 500 MB (Suitable for resource-constrained environments (e.g., portable medical devices, cloud applications)).
- Sensitivity and Specificity:** High values for DE, GWO, and HHO indicate strong target region segmentation and low false-positive rates.

7. **F1-Score:** DE, GWO, and HHO offered the best trade-off between precision, recall, and accuracy.
8. **Overall Recommendations:**
 - **Real-Time Applications:** PSO and ABC for their speed and efficiency.
 - **High-Precision Applications:** DE, HHO, and GWO due to their balance of accuracy and computational cost.
 - **Versatility:** Nature-inspired algorithms like PSO, ABC, GWO, and HHO offer strong potential for advancing medical image segmentation in healthcare.

Limitations and future work

While this study presents a comprehensive comparative analysis of optimization algorithms for multilevel thresholding-based segmentation, several limitations should be acknowledged. First, the approach was tested on a relatively small set of CT images focused on a single pathology (COVID-19), limiting generalizability across other imaging modalities (e.g., MRI, ultrasound) and anatomical regions. Additionally, the method relies on intensity-based separation, which may be less effective in cases of low-contrast or overlapping tissue distributions where spatial context is essential.

From a computational perspective, while most algorithms performed efficiently on standard hardware, real-time deployment in embedded systems or clinical devices may require further optimization or algorithm simplification. Moreover, the absence of ground truth annotations prevented detailed comparison with expert segmentations.

As future work, we plan to (i) expand the evaluation to include diverse datasets with expert-labeled masks, (ii) explore hybrid models that combine metaheuristic optimization with learned features from deep neural networks, and (iii) investigate adaptive thresholding frameworks that integrate spatial and contextual priors to improve segmentation in complex cases. Such directions will help bridge the gap between lightweight optimization methods and high-capacity deep learning systems for broader clinical applicability.

Data availability

The datasets generated and/or analysed during the current study are available in the Cancer Imaging Archive (TCIA) repository, <https://www.cancerimagingarchive.net/access-data/>.

Received: 22 April 2025; Accepted: 30 July 2025

Published online: 24 October 2025

References

1. Dorgham, O., Ryalat, M. H. & Naser, M. A. Automatic body segmentation for accelerated rendering of digitally reconstructed radiograph images. *Inf. Med. Unlocked*. **20**, 100375 (2020).
2. Dorgham, O. et al. U-NetCTS: U-Net deep neural network for fully automatic segmentation of 3D CT DICOM volume. *Smart Health*. **26**, 100304 (2022).
3. Dhiman, G. & Kumar, V. Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications. *Adv. Eng. Softw.* **114**, 48–70 (2017).
4. Jumiawi, W. A. H. & El-Zaart, A. Improvement in the Between-Class variance based on lognormal distribution for accurate image segmentation. *Entropy (Basel Switzerl.)* **2022**, 24 (2022).
5. Al-Najdawi, N., Biltawi, M. & Tedmori, S. Mammogram image visual enhancement, mass segmentation and classification. *Appl. Soft Comput.* **35**, 175–185 (2015).
6. Ryalat, M. H. et al. Harris Hawks optimization for COVID-19 diagnosis based on multi-threshold image segmentation. *Neural Comput. Appl.* **35**, 6855–6873 (2023).
7. Ahmadi, M., Kazemi, K., Aarabi, A., Niknam, T. & Helfroush, M. S. Image segmentation using multilevel thresholding based on modified bird mating optimization. *Multimed Tools Appl.* **78**, 23003–23027 (2019).
8. Drogham, O. et al. Dynamic colormap visualization integrated with Harris Hawks optimization for enhanced lung CT segmentation and diagnostic precision. *Cluster Comput.* **28**, 25 (2025).
9. Zhang, Z. et al. A novel deep learning model for medical image segmentation with convolutional neural network and transformer. *Interdisc. Sci. Comput. Life Sci.* **15**, 663–677 (2023).
10. Strika, Z., Petkovic, K., Likic, R. & Batenburg, R. Bridging healthcare gaps: a scoping review on the role of artificial intelligence, deep learning, and large Language models in alleviating problems in medical deserts. *Postgrad. Med. J.* **101**, 4–16 (2024).
11. Elaziz, M. A., Oliva, D., Ewees, A. A. & Xiong, S. Multi-level thresholding-based grey scale image segmentation using multi-objective multi-verse optimizer. *Expert Syst. Appl.* **125**, 112–129 (2019).
12. Sharma, A., Kumar, S. & Singh, S. N. Brain tumor segmentation using DE embedded OTSU method and neural network. *Multidim Syst. Sign Process.* **30**, 1263–1291 (2019).
13. Baby Resma, K. P. & Nair, M. S. Multilevel thresholding for image segmentation using Krill herd optimization algorithm. *J. King Saud Univ. - Comput. Inform. Sci.* **33**, 528–541 (2021).
14. Mousavirad, S. J. & Ebrahimipour-Komleh, H. Human mental search-based multilevel thresholding for image segmentation. *Appl. Soft Comput.* **97**, 105427 (2020).
15. Liang, H., Jia, H., Xing, Z., Ma, J. & Peng, X. Modified grasshopper Algorithm-Based multilevel thresholding for color image segmentation. *IEEE Access*. **7**, 11258–11295 (2019).
16. Chen, H., Yang, C., Heidari, A. A. & Zhao, X. An efficient double adaptive random spare reinforced Whale optimization algorithm. *Expert Syst. Appl.* **154**, 113018 (2020).
17. Elaziz, M. A. & Lu, S. Many-objectives multilevel thresholding image segmentation using knee evolutionary algorithm. *Expert Syst. Appl.* **125**, 305–316 (2019).
18. Li, Z., Sun, Y., Zhang, L., Tang, J. & CTNet Context-Based tandem network for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 9904–9917 (2022).
19. Kumar, A. S., Kumar, A., Bajaj, V. & Singh, G. K. In *2018 International Conference on Communication and Signal Processing (ICCSP)* 160–164 (IEEE, 2018).
20. Eelbode, T. et al. Optimization for medical image segmentation: theory and practice when evaluating with dice score or Jaccard index. *IEEE Trans. Med. Imaging*. **39**, 3679–3690 (2020).
21. Shaukat, F., Raja, G. & Frangi, A. F. Computer-aided detection of lung nodules: a review. *J. Med. Imag.* **6**, 1 (2019).

22. Jiang, H. et al. An automatic detection system of lung nodule based on multigroup Patch-Based deep learning network. *IEEE J. Biomedical Health Inf.* **22**, 1227–1237 (2018).
23. Zhang, G. et al. Automatic nodule detection for lung cancer in CT images: a review. *Comput. Biol. Med.* **103**, 287–300 (2018).
24. Singh, M. P. et al. A healthcare system employing lightweight CNN for disease prediction with artificial intelligence. *TOPHJ* **17**, 104000 (2024).
25. Chandan, R. R. et al. Reviewing the impact of machine learning on disease diagnosis and prognosis: a comprehensive analysis. *TOPAINJ* **17**, e18763863267142 (2024).
26. Sun, G. et al. DA-TransUNet: integrating Spatial and channel dual attention with transformer U-net for medical image segmentation. *Front. Bioeng. Biotechnol.* **12**, 1398237 (2024).
27. Sun, G. et al. FKD-Med: Privacy-Aware, Communication-Optimized medical image segmentation via federated learning and model lightweighting through knowledge distillation. *IEEE Access.* **12**, 33687–33704 (2024).
28. Pan, Y. et al. A mutual inclusion mechanism for precise boundary segmentation in medical images. *Front. Bioeng. Biotechnol.* **12**, 1504249 (2024).
29. Codella, N. C. F. et al. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* 168–172 (IEEE, 2022).
30. Menze, B. H. et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging.* **34**, 1993–2024 (2015).
31. Ahmad, J., Akram, S., Jaffar, A., Rashid, M. & Bhatti, S. M. Breast cancer detection using deep learning: an investigation using the DDSM dataset and a customized AlexNet and support vector machine. *IEEE Access.* **11**, 108386–108397 (2023).
32. Qin, J., Tang, Y. & Wang, B. Regional 18F-fluoromisonidazole PET images generated from multiple advanced MR images using neural networks in glioblastoma. *Medicine* **101**, e29572 (2022).
33. Fonseca, C. G. et al. The cardiac atlas Project—an imaging database for computational modeling and statistical atlases of the heart. *Bioinf. (Oxford England).* **27**, 2288–2295 (2011).
34. Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A. & van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging.* **23**, 501–509 (2004).
35. Desai, S. et al. Chest imaging representing a COVID-19 positive rural U.S. Population. *Sci. Data.* **7**, 414 (2020).
36. Clark, K. et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging.* **26**, 1045–1057 (2013).
37. Armato, S. G. et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* **38**, 915–931 (2011).
38. Rayed, M. E. et al. Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Inf. Med. Unlocked.* **47**, 101504 (2024).
39. Yan, F., Huang, H., Pedrycz, W. & Hirota, K. Review of medical image processing using quantum-enabled algorithms. *Artif. Intell. Rev.* **57**, 773 (2024).
40. Diwakar, M., Singh, P., Ravi, V. & Maurya, A. A Non-Conventional review on Multi-Modality-Based medical image fusion. *Diagn. (Basel Switzerl.)* **2023**, 13 (2023).
41. Houssein, E. H., Mohamed, G. M., Djenouri, Y., Wazery, Y. M. & Ibrahim I. A. Nature inspired optimization algorithms for medical image segmentation: a comprehensive review. *Cluster Comput.* **27**, 14745–14766 (2024).
42. Xu, Y. et al. Advances in medical image segmentation: a comprehensive review of traditional, deep learning and hybrid approaches. *Bioeng. (Basel Switz.)* **2024**, 11 (2024).
43. Taino, D. F. et al. Analysis of cancer in histological images: employing an approach based on genetic algorithm. *Pattern Anal. Applic.* **24**, 483–496 (2021).
44. Mirjalili, S. et al. Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. *Adv. Eng. Softw.* **114**, 163–191 (2017).
45. Kennedy, J. & Eberhart, R. In *Proceedings of ICNN'95 - International Conference on Neural Networks 1942–1948* (IEEE, 1995).
46. Mirjalili, S. S. C. A. A sine cosine algorithm for solving optimization problems. *Knowl. Based Syst.* **96**, 120–133 (2016).
47. Yang, X. S. & Deb, S. In *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)* 210–214 (IEEE, 2011).
48. Mirjalili, S. Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm. *Knowl. Based Syst.* **89**, 228–249 (2015).
49. Storn, R. & Price, K. Differential Evolution – A simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **11**, 341–359 (1997).
50. Khishe, M. & Mosavi, M. R. Chimp optimization algorithm. *Expert Syst. Appl.* **149**, 113338 (2020).
51. Yang, X. S. Firefly algorithm, stochastic test functions and design optimisation. *IJBIC* **2**, 78 (2010).
52. Dorigo, M., Maniezzo, V. & Colnari, A. Ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man. Cybern.* **26**, 29–41 (1996).
53. Abdollahzadeh, B., Soleimani Gharehchopogh, F. & Mirjalili, S. Artificial Gorilla troops optimizer: a new nature-inspired metaheuristic algorithm for global optimization problems. *Int. J. Intell. Syst.* **36**, 5887–5958 (2021).
54. Karaboga, D. & Basturk, B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *J. Glob. Optim.* **39**, 459–471 (2007).
55. Eluri, R. K. & Devarakonda, N. Binary golden eagle optimizer with Time-Varying flight length for feature selection. *Knowl. Based Syst.* **247**, 108771 (2022).
56. Mirjalili, S. & Lewis, A. The Whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016).
57. Mousavirad, S. J. & Ebrahimipour-Komleh, H. Human mental search: a new population-based metaheuristic optimization algorithm. *Appl. Intell.* **47**, 850–887 (2017).
58. Mirjalili, S., Mirjalili, S. M. & Lewis, A. Grey Wolf optimizer. *Adv. Eng. Softw.* **69**, 46–61 (2014).
59. Heidari, A. A. et al. Harris Hawks optimization: algorithm and applications. *Future Generation Comput. Syst.* **97**, 849–872 (2019).
60. Jumaiwi, W. A. H. & El-Zaart, A. In *2022 International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE)* 1–6 (IEEE, 2022).

Acknowledgements

This work was conducted during sabbatical leave supported by Al-Balqa Applied University. The authors express their gratitude to Al-Balqa Applied University, Prince Sultan University and Princess Sumaya University for Technology for their provision of laboratories, materials, and equipment that facilitated the entirety of the experimental work.

Author contributions

N.A., S.T., A.A.: Conceived the core idea of the study, formulated the research objectives, and provided expert supervision throughout the project. Also contributed to the design of the methodology and overall coordination. N.A., S.T., A.A., I.I.: Collected and curated the dataset, handled data preprocessing, and ensured compliance with relevant medical imaging standards and ethical considerations. N.A., S.T., I.I.: Designed and implemented the

image processing algorithms, including feature extraction, segmentation, and enhancement techniques. Led the coding and technical development aspects. N.A, S.T, O.D: Conducted extensive experiments, statistical analysis, and validation of results. Also contributed to the tuning of model parameters and performance evaluation. A.A, I.I, O.D: Conducted a detailed literature review, assisted with result interpretation, and was the lead in drafting the manuscript, including figures and tables.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.A.A.-N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025