Commun Nonlinear Sci Numer Simulat 88 (2020) 105256

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

journal homepage: www.elsevier.com/locate/cnsns

Research paper

Recent nature-Inspired algorithms for medical image segmentation based on tsallis statistics

G.A. Wachs-Lopes[a], R.M. Santos[a], N.T. Saito[b], P.S. Rodrigues[a,b,*]

[a] Department of Computer Science, Inacian Educational Foundation, Sao Bernardo, SP, Brazil
[b] Department of Electrical Engineering, Inacian Educational Foundation, Sao Bernardo, SP, Brazil

ARTICLE INFO

Article history:
Received 7 July 2019
Revised 3 March 2020
Accepted 9 March 2020
Available online 21 March 2020

Keywords:
Tsallis entropy
Non-Extensive theory
Bio-Inspired algorithms
Medical image segmentation
Optimization

ABSTRACT

Recently, many algorithms have emerged inspired by the biological behavior of animal life to deal with complicated applications such as combinatorial optimization. One of the most critical discussions involving these algorithms is concerning their objective functions. Also, recently, many works have demonstrated the efficiency of Tsallis non-extensive statistics in several applications. However, this formalism has not yet been tested in most recent bio-inspired algorithms as an evaluation function. Thus, this paper presents a study of seven of the most promising bio-inspired algorithms recently proposed (a maximum one decade), from this entropy applied to the multi-thresholding segmentation of medical images. The results show the range of values of q, the so-called non-extensivity parameter of the Tsallis entropy, for which the algorithms tested here have their best performance. It is also demonstrated that the Firefly algorithm (FFA) is the one that obtained the best performance in terms of segmentation, and Grey Wolf Optimizer (GWO) presents the fastest convergence.

1. Introduction

In recent decades, several computational applications with high temporal complexity have been faced through the use of optimization algorithms inspired by the animal behavior of several species, the so-called bio-inspired algorithms. Among the best known are the PSO [1], simulated annealing [2], ant colony [3], taboo search [4], fire-fly [5] and cuckoo search [6]. However, due to their great practical success, more algorithms with proven efficiency have been proposed in the last decade than in the entire history of the optimization area. Recent examples can be seen as Krill Herd [7], Grey Wolf [8], Elephant Herding [9], Whale [10] and Grasshopper [11], to name a few.

One of the areas that has benefited from that development is the computer vision area, specifically for applications involving segmentation of images based on several thresholds, or multi-thresholding, where the goal is to find a partition of one-dimensional space of luminances in d-thresholds, resulting in a final process known as multi-segmentation. However, for values of $d > 1$, brute-force solutions with optimal results are computationally costly. On the other hand, for the same dimensions $d > 1$, bio-inspired optimization algorithms achieve similar results to those achieved by the brute-force approaches (which test all possible combinations of the search space) but with computational time much acceptable, which makes them more attractive.

---

[*] Corresponding author.
E-mail address: psergio@fei.edu.br (P.S. Rodrigues).

https://doi.org/10.1016/j.cnsns.2020.105256
1007-5704/

In practice, among the applications that have most demanded this type of approach are applications of medical field, where the volume of images is generally not feasible for manual segmentation or even for the use of brute-force solutions. In addition, medical imaging presents another factor that requires fully automated segmentation: such images are generally best evaluated with the use of multi-segmentation ($d > 1$), which generally separates regions of interest of an image for later independent and correlated analyzes. An example are brain regions, generally observed under Computer Tomography (CT), fMRI (functional Magnetic Resonance Image) or histologically evaluated through microscopic acquisition. In these images, it is common to observe several regions of interest coexisting in the same scene, such as tumoral regions, bones, arterial, muscular regions or the vast area known as glia, which for many decades was supposed to function only to support the physical brain structure.

Despite the major scientific development based on brain imaging, a more accurate recognition and assessment of such regions still requires a lot of work and advancement in terms of computational technology, especially image segmentation. A recent paper from neuroscience area [12] comments that, according to the Federative Community for Anatomical Terminology (FCAT), 97% of the subcortical regions still fall into the denomination called unknown earth. On the other hand, a basic calculation indicates that, at the current rate of histological mapping, the completion of the work will occur only in 1200 years! [12].

However, the use of multi-thresholding for medical image partitioning is mainly justified as they are composed of regions with unpredictable topologies, meaning that they are usually images without a human-interpretable cognitive structure, as opposed to natural scenes that can be separated in clusters of regions according to rules of human visual recognition system. In addition, thresholding algorithms are usually easy to implement and execute.

Considering then the bio-inspired algorithms applied to multi-segmentation images, the rules for the use of evaluation functions of these algorithms, which guide their results, have also been the issue of scientific discussion for several decades. For multi-segmentation of medical images specifically, traditional entropies such as Shannon [13], Kapur [14] or Otsu method [15] have been frequently used as kernels of these functions.

However, more recently, some authors have demonstrated that the non-extensive Tsallis entropy is a promise alternative. This new kind of entropy generalizes the traditional Boltzman-Gibbs-Shannon entropy though by including the so-called non-extensivity parameter q, which allows a topological fitting on the probability distribution of the physical system embedding this process.

Since its proposal in the field of statistical mechanics, many scientific researchers have demonstrated the positive effect of Tsallis entropy in several areas of applications [16–21]. However, the foundations for its effects are not yet conclusive. As it is also an open topic of discussion the relation of q values with the applications that have generated more favorable results. In these cases, for most works, the best q values are usually achieved through empirical tests which are tedious and time consuming. Consequently, the works proposing multi-segmentation image applications based on bio-inspired algorithms have not yet shown a relationship between this type of algorithm and the Tsallis entropy as its evaluation function.

Thus, the demonstration of the behavior of bio-inspired algorithms applied to tasks of multi-segmentation for medical images, under the variation of the values of the q parameter of the non-extensive Tsallis entropy used as a kernel of these algorithms, is a demand that impacts not only in the understanding of the involved physical phenomena in this type of image, but also in the induction of new and better computer systems able to accelerate the process of study, analysis and diagnosis (when is the case) of applications involving medical imaging.

Thus, considering the promising results of Tsallis entropy in the medical field and the success of bio-inspired algorithms for the process of multi-segmentation of images, and also considering that no thorough study has been carried out under this context, this paper presents a study of the effects of the non-extensive Tsallis entropy used as an evaluation function in several important and recent bio-inspired algorithms for multi-segmentation of medical images.

The remainder of this paper is the following. The next section presents a review of the main bio-inhaled algorithms in about a decade. Then the context of the tsallis entropy is presented. The details of the bio-inspired algorithms compared and studied here under the context of medical imaging are presented below. Finally, sections explaining our methodology, experimental results and conclusions are presented.

2. Related works

The general idea of combinatorial algorithms is the search for random but coordinated solutions under an iterative process so that some solutions approach the ideal solution in a search space. This ideal solution, according to an evaluation function, is expected to be the closest to a global minimum, not necessarily exact, but with acceptable computational time. In the case of multi-threshold-histograms, the search space is generally linear and small (usually $d < 4$).

The most well-known traditional algorithms that fit this requirement and have already been tested for multi-threshold-histograms are based on Particle Swarm Optimization (PSO) [22], inspired by the social behavior of specimens such as fish and Bird in nature, the so-called swarm intelligence. These algorithms are very similar to those based on genetic algorithms, however they differ because they do not have mutation and cross-over and are therefore simpler. In contrast, PSO-based algorithms are computed based on random solutions that converge to global minimum or maximum. Thus, several variations of the PSO were successfully proposed. Among the best known are Firefly [23], which mimics the collective flight of fireflies following the luminescence of leaders. The article [24] has shown that this algorithm, with non-extensive entropy-based evaluation function, overcomes the cross-entropy and other forms of entropy when applied to natural image bases.

With the increasing popularity of bio-inspired algorithms to deal with combinatorial problems, new research has been constantly proposed. Crustacean-based algorithms, for example, which mimics the live in flocks in constant search for food, have inspired the emergence of Krill Heard Algorithm (KHA), proposed in [25] and generating new works like those in [26–28]. Similarly, algorithms based on the strategy of bird-parasites, such as Cuckoo Search (CS), that lay their eggs in nests of host species were proposed in [6], also generating new works such as [29–38]. Also, algorithms based on the social behavior of Elephant Herds (EH) were proposed in [39], inducing further works on [40,41] and [42].

Among the recent bio-inspired algorithms that have most attracted the attention of researchers are the Grey Wolf Optimizer (GWO) [8,43], which mimics the leadership hierarchy and hunting mechanism of grey wolfs in nature; the Whale Optimization Algorithm (WOH) [44], that mimics the social behavior of humpback whales, and more recently the Grasshopper Optimization Algorithm (GOA) [45], that mimics the behavior of grasshopper swarm.

Although some of these algorithms have been compared from the point of view of medical images with evaluation function based on the Tsallis entropy, there is not still in the literature a deep study of this performance under this point of view. On the other hand, swarm-based algorithms can be divided into two groups based on the convergence strategy: those that converge in ordered search spaces and those that converge in random search spaces. An example of ordered search spaces is the search for multi-threshold-histograms, and an example of random spaces is the search by ideal sub-parameterization for a system application.

Thus, this paper presents the study of well-known bio-inspired algorithms proposed about approximately a decade ago (CSL, EHO, KHA and FFA), besides three new algorithms (GWO, GOA and WOA) that, although they are also very recent proposals (less than 5 years ago), already have a large number of interests in the scientific literature. However, they have not yet been deeply tested from the point of view of Tsallis's entropy in the specific area of multi-thresholding-histograms applied to medical images. These algorithms, that will be evaluated in this work, are described in more details in the following section.

3. Tsallis non-extensive entropy

Inspired by the multifractal theory, the non-extensive Tsallis entropy, presented in Eq. (1), was proposed in [46] under the context of non-extensive statistical mechanics for better understanding problems from physics of solid state to information theory [16–21]. These works also demonstrate that the Tsallis entropy is a generalization of the Boltzman-Guibs-Shannon (BGS) entropy and can thus be adapted to various physical systems.

$$S_q(A) = \frac{1}{1-q}\left(1 - \sum_i x_i^q\right) \tag{1}$$

where A is the physical system, and q is the so called the non-extensivity parameter. When $q < 1.0$, the system is called sub-extensive, when $q > 1.0$ the system is called super-extensive, and when $q = 1$ the system is reduced to the traditional BGS entropy.

In the area of image processing, the work of [47] showed the idea of Tsallis entropy applied for histogram binarization, and in [48,49] this idea was extended for multi-threshold-histogram in the contex of image segmentation.

The Tsallis entropy applied to image processing is facilitated due to one of its main properties, known as additivity principle. This principle says that, given two independent physical systems, say A and B, with corresponding entropies S_A and S_B respectively, the entropy of the whole system (composed by A and B) can be given by the Eq. (2).

$$S_q(A,B) = S_q(A) + S_q(B) + (1-q).S_q(A).S_q(B) \tag{2}$$

Eq. (2) is applied to the case of only a single threshold $t = 1$ for a unique partition. But in the case of multi-thresholding $t > 1$, this process becomes recursive, given by Eq. (3) for $A, B, \ldots, Z$ physical systems.

$$S_q(A,B,C) = S_q(S_q(A,B),C) = S_q(A, S_q(B,C)) \tag{3}$$

This idea is demonstrated as a pseudocode of Algorithm (1) in the next section.

Finally, in the multi-segmentation-histogram process with bio-inspired algorithms, such as those studied in this paper, Eq. (3) is used as an evaluation function of these algorithms. In this case, consider a solution of d thresholds for any bio-inspired algorithm, given by $X = \{[t_1, t_2, \ldots, t_d]\}$. The X solution will be ideal when X maximizes the Eq. (3). Thus, the recursive solution initially splits the threshold space X into two sub-sets $X_1 = \{[t_1, t_2, \ldots, t_i\}$ and $X_2 = \{[t_{i+1}, t_{i+2}, \ldots, t_d]\}$, recursively computing the value of Eq. (3) until the final result is reached. This process is slightly different from that presented in the works of [24,50], where the solution is non-recursive but extended. However, we note that the solution proposed by [24,50], although presenting satisfactory results in acceptable computational time, are not in accordance with the original multi-additive theory proposed by Tsallis in [46]. Thus, as far as we know, it is the first time that the original equation is used to solve the evaluation function of bio-inspired algorithms. The further discussion of this comparison is, however, out of the scope of our paper.

4. Optimization algorithms

This section describes the seven optimization algorithms studied in this paper: Firefly Algorithm (FFA); Cuckoo Search via Lévy Flights (CSL); Krill Herd Algorithm (KHA); Grey Wolf Optimizer (GWO); Elephant Herding Optimization (EHO); Whale Optimization Algorithm (WOA); and Grasshopper Optimization Algorithm (GOA).

All algorithms studied here were originally proposed for general optimization problems and were further adapted for image segmentation using region classification or multi-threshold-histograms. In the present paper, all them are studied only from the point of view of multi-threshold-histograms with evaluation function based on non-extensive Tsallis entropy. The objective here is to compare not only the convergence time but also the correlation of the non-extensivity parameter q.

4.1. Firefly algorithm (FFA)

The Firefly Algorithm was developed by Xin-She Yang in 2007 [51] inspired on bioluminescence behaviour of the fireflies. The true functions of signaling system are still being debated, however, two fundamental functions are know: 1) Attract mating partner and 2) Attract potential prey.

The algorithm follows three idealized rules:

- All Fireflies are unisex;
- Attractiveness is proportional to the their brightness, thus for any two flashing fireflies, the less brighter one will move toward the brighter one. If there no brighter one than a particular firefly, it will move randomly.
- The brightness of a firefly is affected or determined by the landscape of the objective function.

There are two important issues in this algorithm: 1) The variation of light intensity and 2) The formulation of the attractiveness.

The mathematical model proposed for the light intensity is Eq. (4):

$$I(r) = I_0 e^{-\gamma r^2} \tag{4}$$

where I_0 is the original light intensity, γ is the light absorption coefficient and r is the distance between two fireflies.

For the attractiveness, the mathematical model is represented by Eq. (5):

$$\beta(r) = \beta_0 e^{-\gamma r^m}, \qquad (m \geq 1) \tag{5}$$

where β_0 is the attractiveness at $r = 0$.

The movement of a firefly i is attracted to another more attractive firefly j is determined by Eq. (6):

$$x_i = x_i + \beta_0 e^{-\gamma_{ij}^2} (x_j - x_i) + \alpha \epsilon_i \tag{6}$$

where the second term is due to the attraction, the third term is the randomization with α being the randomization parameter, and ϵ_i is a vector of random numbers drawn from a Gaussian distribution. For most implementation $\beta_0 = 1$, $\alpha \in [0, 1]$ and γ varies from 0.1 to 10.

Recent works demonstrate the application in segmentation, He and Huang in 2017 [52], Suresh in 2016 [53] and [24].

Since this algorithm has two internal loops when evaluating its n-size population, and an external loop for Max Number of Generations (T). Thus, the complexity in the extreme case is of the order of $O(N^2T)$, where N is the number of agents. This is the case for most meta-heuristic algorithms and is also the case for all seven algorithms studied here. If the value of N is relatively high, it is still possible to use an internal loop to rank fireflies based on their attractiveness or brightness using some sorting algorithm. In this case, the complexity of firefly becomes of the order of $O(NT\log(N))$.

4.2. Cuckoo Search via Lévy Flights *(CSL)*

Proposed by Yang and Deb in 2009 [6], the Cuckoo Search is based on mimicking the parasitic behavior of cuckoo combined with the Lévy flight mechanism. This algorithm follow three idealized rules:

- Each cuckoo lays one egg at a time, and dump its egg in randomly chosen nest;
- The best nest will carry over to the next generations;
- The number of host nests is fixed, and the cuckoo's egg is discovered by the host bird with a probability $p_a \in [0, 1]$;

The nests discovered by the host are replaced by new nests, according Eq. (7):

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \oplus \text{Lévy}(\lambda) \tag{7}$$

where $\alpha > 0$ is the step size, in most case can use $\alpha = 1$, the first term in the above equation is a Markov chain, the second term is the transition probability and the product $\oplus$ means entrywise multiplications.

The Lévy flight provides a random walk while the random step length is draw from a Lévy distribution, according to Eq. (8):

$$\text{Lévy} \quad u = t^{-\lambda}, \qquad \text{where } (1 < \lambda \leq 3), \tag{8}$$

which has an infinite variance with an infinite mean (Mantegna's algorithm).

Recent works demonstrate the application of CSL in image segmentation [54] and [31].

The computational complexity of the nature-inspired CS algorithm is given by the initial generation of cuckoo nests of size P, which produces the complexity of the order of $O(PN)$, where N is the dimension of the decision variables. This algorithm calculates a sorting strategy, therefore, of the order of $O(N\log(N))$. Thus, the total complexity becomes about $O(N\log(N)) + O(PN)$ [6].

4.3. Krill herd algorithm (KHA)

First proposed by Gandomi and Alavi in [25], the Krill Herd Algorithm (KHA) has three essential actions for determining the time-dependent position of an individual, they are: movement induced by the presence of other individual; foraging activity and random diffusion. These three actions are denoted in the following lagrangian model for an N dimensional decision space represented by Eq. (9):

$$\frac{dX_i}{dt} = N_i + F_i + D_i \tag{9}$$

where N_i is the motion induced by the other individuals, F_i is the foraging motion and D_i is the physical diffusion of the *ith* krill individual.

The mathematical model for each individual motion is proposed by Eq. (10):

$$N_i^{new} = N^{max}\alpha_i + \omega_n N_i^{old} \tag{10}$$

where $\alpha_i = \alpha_i^{local} + \alpha_i^{target}$, $N^{\max}$ is the maximum induced speed, ω_n is the inertia weight of the motion induced in the range [0, 1], N_i^{old} is the last motion induced, α_i^{local} is the local effect provided by the neighbors and α_i^{target} is the target direction effect provided by the best krill individual.

The mathematical model for foraging motion is proposed by Eq. (11):

$$F_i = V_f\beta_i + \omega_f F_i^{old} \tag{11}$$

where $\beta_i = \beta_i^{food} + \beta_i^{best}$, V_f is the foraging speed, ω_f is the inertia weight of the foraging motion in the range [0, 1], F_i^{old} is the last foraging motion, β_i^{food} is the food attractive and β_i^{best} is the effect of the best fitness of the *ith* krill.

The mathematical model for physical diffusion is proposed by Eq. (12):

$$D_i = D^{max}\delta \tag{12}$$

where $d^{\max}$ is the maximum diffusion speed and δ is the random directional vector and its arrays are random value between -1 and 1.

Recent works demonstrating the KHA application in image segmentation can be achieved in [7] and [55].

The computational complexity of the KH algorithm is based on the number of krills, N, within each iteration t, with the maximum number of iterations T. The dynamics of this algorithm depend on the distance between krills, called sensing distance d_{s_i}, being calculated with the complexity order $O(N)$. Thus, the total computational time of the KHA algorithm is of the order of $O(TN^2)$.

4.4. Grey wolf optimizer (GWO)

The Grey Wolf Optimizer (GWO), developed by Mirjalili et. al in [8,43], simulates the leadership hierarchy applying four types of grey wolfs: alpha, beta, delta and omega, where alpha is the leader, beta is the second in command, delta is the third and omega are the lowest in the hierarchy. The hunting behavior is implemented by three steps: searching, encircling and attacking.

The mathematical for encircling the prey is proposed in Eqs. (13) and (14):

$$\vec{D} = \left|\vec{C}\cdot\vec{X}_p(t) - \vec{X}(t)\right| \tag{13}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}\cdot\vec{D} \tag{14}$$

where t is the iteraction, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}$ is the position vector and $\vec{X}_p$ is the best solution.

The vectors $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = 2\vec{a}\cdot\vec{r}_1 - \vec{a} \tag{15}$$

$$\vec{C} = 2\cdot\vec{r}_2 \tag{16}$$

where $\vec{a}$ is linearly decreased from 2 to 0 over the course of iterations and $\vec{r}_1$ and $\vec{r}_2$ is a random vector in [0,1].

In order to simulate the hunting behavior, it is suppose that alpha (the best candidate solution), beta and delta have better knowledge about the potential location of prey. The Eqs. (17), (18) and (19) are proposed in this regard:

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right|,\\ \vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right|,\\ \vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right| \tag{17}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha),\\ \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta),\\ \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta), \tag{18}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{19}$$

Attacking the prey represents the exploitation phase, when the grey wolfs attack the prey. When the pray stops moving, in the algorithm it is represented by decreasing the value of vector $\vec{a}$; that way the value of $\vec{A}$ decreasing as well. When $\vec{A} < 1$ the wolfs are forced to attack the prey, in other hand, when $\vec{A} > 1$ the wolfs go looking for another prey (exploration phase).

Recent works demonstrate the application of GWO in image segmentation [56] and [57].

The wolf pack initialization has $O(N)$ complexity, where N is the number of wolves. For the initialization of parameters a, A, and C, the complexity is $O(3)$ and $30(N)$ to select the leaders. The primary computational step is the main loop. In GWO, individual positions of wolves can be updated in time $O(N)$, and the suitability of the set of wolves can also be calculated in time $O(N)$, assuming that the computational complexity of the benchmark function is $O(1)$. Parameter and leader updates have computational complexity $O(3)$ and $3O(N)$, respectively. As the wolf update process continues until the maximum number of iterations is reached, the computational complexity in the main while loop becomes $O((5N+3) \cdot T)$, where T is the maximum number of iterations. Thus, the total computational complexity in GWO is $O(T \cdot N)$.

4.5. Elephant Herding Optimization (EHO)

Developed by Wang, Deb and Coelho in [9], the Elephant Herding Optimization (EHO) is inspired by the herding behavior of elephant group where different clans live together under a leadership of a matriarch. This behavior is implemented as clans updating operator, and the male elephant will leave the group when they grow up, and this behavior is implemented as a separating operator.

The clan updating operator is represented by the Eq. (20):

$$x_{new,ci,j} = x_{ci,j} + \alpha \times (x_{best,ci} - x_{ci,j}) \times r \tag{20}$$

where $x_{new,\, ci,\, j}$ and $x_{ci,j}$ are newly updated and old position for the elephant j in the clan ci, respectively. $\alpha \in [0, 1]$ is a scale factor of matriarchal influence, $x_{best,ci}$ represents the matriarch and $r \in [0, 1]$.

The matriarch is updated by the following Eq. (21):

$$x_{new,ci,j} = \beta \times x_{center,ci} \tag{21}$$

where $\beta \in [0, 1]$ is a factor that determines the influence of the $x_{center,ci}$ on $x_{new,\, ci,\, j}$.

To compute the clans center for a *d-th* dimension the following Eq. (22) is used:

$$x_{new,ci,d} = \frac{1}{N_{ci}} \times \sum_{j=1}^{N_{ci}} x_{ci,j,d} \tag{22}$$

where $1 \leq d \leq D$ indicates the *d-th* dimension, D is the total dimensions, N_{ci} is the number of elephants and $x_{ci,j,d}$ is the elephant j in the clan ci of dimension d.

The separating operator is presented by the Eq. (23):

$$x_{worst,ci} = x_{min} + (x_{max} - x_{min} + 1) \times rand \tag{23}$$

where $x_{\max}$ and $x_{\min}$ are respectively the upper and lower bound of the elephant individual position, $x_{worst,\, ci}$ is the worst elephant in the clan and the $rand \in [0, 1]$ is the stochastic distribution.

Recent works demonstrating the application of GWO in image segmentation are [58] and [59].

The EHO algorithm performs the following four main steps: classification, updating, separation and evaluation of elephant population. Using the Quick Sort algorithm to rank the N elephants, the order of complexity for this step will then be $O(N\log(N))$. In turn, the update step consists of two levels of cyclic operations, depending on the number of clans and the number of elephants in each clan, thus resulting in $O(N_{clan}N_{(c_i)})$. Also, the separation step performs operations on each clan, reaching a complexity of $O(N_{clan})$. Finally, the population assessment concerns the total number of elephants, so the complexity of this step is $O(N_{c}lanN_{(c_i)})$. Therefore, the total complexity of the EHO algorithm is the joining of the four steps for a total of the maximum number of generations (T): $O(T(N\log N + N_{clan} * (1 + 2N_{ci}))$, where the terms of the sum correspond, respectively, to the four operations of the algorithm.

4.6. Whale optimization algorithm (WOA)

The Whale Optimization Algorithm (WOA), developed by Mirjalili and Lewis in [60], mimics the social behavior of humpback whales, inspired by the bubble-net hunting strategy. The optimization include three operators to simulate the search for prey, encircling prey and bubble-net foraging behavior.

The operator that simulate the encircling prey is represented by Eqs. (24) and (25):

$$\vec{D} = \left|\vec{C} \cdot \vec{X^*}(t) - \vec{X}(t)\right| \tag{24}$$

$$\vec{X}(t+1) = \vec{X^*}(t) - \vec{A}.\vec{D} \tag{25}$$

where t is the iteraction, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}$ is the position vector and $\vec{X^*}$ is the best solution.

The vector $\vec{A}$ and $\vec{C}$ are computed as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \tag{26}$$

$$\vec{C} = 2 \cdot \vec{r} \tag{27}$$

where $\vec{a}$ is linearly decreased from 2 to 0 over the course of iterations and $\vec{r}$ is a random vector in [0,1].

The bubble-net attacking method represents the exploitation phase and has two approaches: 1) shrinking encircling mechanism and 2) spiral updating position. There is a probability of 50% to choose either one or another.

The mathematical model is represented by Eq. (28):

$$\vec{X}(t+1) = \begin{cases} \vec{X^*}(t) - \vec{A} \cdot \vec{D} & \text{if } p < .5 \\ \vec{D'} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X^*}(t) & \text{if } p \geq .5 \end{cases} \tag{28}$$

where $\vec{D'}$ indicates the distance of the *ith* whale to the prey, b is a constant defining the shape of the logarithmic spiral, l is a random number in $[-1, 1]$ and p is a random number in [0, 1].

The search for prey represents the exploration phase, it starts with a set of random solutions, the whales update their position regarding either a randomly chosen search whale or the best positioned whale (prey).

The mathematical model is represented by Eqs. (29) and (30):

$$\vec{D} = \left|\vec{C} \cdot \overrightarrow{X_{rand}} - \vec{X}\right| \tag{29}$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \tag{30}$$

where $\vec{X}_{rand}$ is a random positioned whale. If $|\vec{A}| > 1$ a random whale is chosen otherwise the best positioned whale is chosen.

Recent works demonstrating the application of WOA in image segmentation can be seen in [61] and [62].

Since whales are known to hunt in spiral groups around prey, the algorithm is based on the position of each whale individually concerning hunting (optimal solution). At each iteration, parameters related to hunters, prey, and positions are updated toward the optimal solution: when a hunter whale catches a prey. Thus, the main factor in the complexity of this algorithm is the main loop, which controls the number of iterations O(T). At each iteration, the parameters of each hunt (number of whales, N) are updated to arrive at an optimal O(Fitness) solution. Thus, the total complexity of this algorithm is of the order of O(T N Fitness).

4.7. Grasshopper Optimization Algorithm (GOA)

The Grasshopper Optimization Algorithm (GOA), developed by Saremi et. al in [45], is an optimization algorithm which mimics the behavior of grasshopper swarm.

There are two tendencies of search process. The first one is the exploration which is an adulthood behavior with long-range and abrupt movement. The second one is the exploitation in the larval phase that is the slow movement and small steps of the grasshoppers.

The mathematical model proposed to solve a optimization problems is:

$$x_i^d = c\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} c\frac{ub_d - lb_d}{2} s\left(\left|x_j^d - x_i^d\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + \hat{T}_d, \tag{31}$$

where ub_d is the upper bond and lb_d is the lower bond in the dimension *dth* of social forces. $\hat{T}_d$ is the dimension value of the best solution found. The first adaptive parameter c, on the left, is used to reduces the movements of grasshoppers around the target. The second c, on the right, decreases the attraction zone, comfort zone and repulsion zone between grasshoppers.

Recent works demonstrating the application GOA in image segmentation can be found in [63] and [11].

Since GOA is a kind of swarm of insects based on the collective behavior of Grasshopper, each element of this algorithm is a proposed solution. Moreover, the cooperation between them and the search for food is taken as the primary optimization function. Thus, its complexity is also similar to other swarm proposes like the firefly, being of the order of $O(N^2T)$, where N is the number of individuals and T is the max number of iterations.

5. Methodology

The proposed methodology used in this paper to study the impact of non-extensive theory over bio-inspired algorithm is conducted empirically and can be divided into three steps: database preparation, optimization experiments and result analysis.

5.1. Database

According to the reviewed literature, most of the papers uses Brats database to study techniques of medical images segmentation. Multimodal Brain Tumor Image Segmentation Benchmark (Brats) database [64,65] was created in conjunction with the MICCAI 2012 and 2013 conferences to provide a common dataset to test medical image segmentation algorithms. Until then, several annual releases were made including new images.

The Brats 2017/2018 database contains multimodal brain scans from 19 institutions and are available as *NIfTI* files. There are 4 types of scans: native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes.

These dataset also contains a manually annotated files provided by neuro-radiologists specialists following the same annotation protocol. There are 3 labels in such annotations: GD-enhacing tumor, the peritumoral edema and the necrotic and non-enhancing tumor core.

In this work we used a collection of 100 patients with all four types of scans.

5.2. Experiments

Since the BRATS database contains full 3D scans, each of these scans is stored as a set of 155 slices of 240 by 240 pixels in *NIfTI* format. Therefore, for 100 patients there are 15,500 images. In order to make the empirical study feasible, we selected two slices per patient with the highest entropy, resulting in 200 slices. The motivation to use the entropy here is that we are interested in images with high amount of information and, therefore, harder for segmentation. Furthermore, some of these slices are totally blank (e.g. first and last slices) or contains too few information. This technique can filter out those images. This step is called database preparation.

The next step is the optimization algorithm benchmarks. All seven optimization algorithms described in Section 4 were tested with fixed parameters recommended by the original authors (see references therein). All the parameters used for each of the algorithms tested in this work are detailed in the address https://github.com/RMorelloS/bioInspiredalgorithms

For all tested optimization algorithms, the same objective functions based on non-additive entropy given by Eq. (3) was used, and the tunning parameter q.

In that equation, A and B are independent physical systems represented by a probability distribution and $S_q(x)$ is the Tsallis Entropy of physical system x given by Eq. (1).

Note that, as posed in [46], when $q \to 1$, the non-extensive theory converges to the traditional entropy function, so this is a generalization of traditional BGS entropy. Moreover, the non-additive entropy function can be expanded recursively when A or B is the union of two or more independent physical systems. Therefore, one can use the property represented by Eq. (3) to create the recursive Algorithm 1.

Algorithm 1 Pseudo additivity recursive algorithm.

```
function TsallisAdditivity($hist, lim, q, i = 0$)
    $s_1 \leftarrow$ TsallisEntropy($hist[lim[i] \ldots lim[i+1]], q$)
    if $i = thresholds - 3$ then
        $s_2 \leftarrow$ TsallisEntropy($hist[lim[i+1] \ldots], q$)
    else
        $s_2 \leftarrow$ TsallisAdditivity($hist, lim, q, i+1$)
    end if
    $s \leftarrow s_1 + s_2 + (1-q).s_1.s_2$
    return s
end function
```

Using Eq. (2) as the objective function enables the use of a large family of functions based on the entropic kernel q. Therefore, the idea in this paper is to analyze how changes in q affect the convergence and quality of segmentation.

All experiments were conducted using the open interval $q \in [0, 2]$ by steps of 0.1 and were repeated 10 times in order to measure the stability of convergence. The objective function is also used as a stop criteria using the standard deviation over the last 30 iterations. When the standard deviation reaches values smaller than 0.01 or the number of iterations reaches 200 rounds, the optimization stops.

Finally, the last step of the proposed methodology is the result analysis. Here, we use four metrics to compare the quality of segmentation: PSNR (Peak Signal-to-Noise Ratio), Jaccard and Dice.

The PSNR metric is based on the mean squared error between two images I_1 and I_2 and is given by Eq. (32).

$$PSNR(I_1, I_2) = 10.\log_{10}\left(\frac{MAX^2}{MSE(I_1, I_2)}\right) \tag{32}$$

where MAX is the maximum possible difference between I_1 and I_2, and MSE is the Mean Squared Error.

The main motivation to use the *PSNR* metric is that we consider that segmentation is a process used to simplify an image without losing too much information. This means that segmentation is a pre-processing step in an image pipeline before more complex later tasks, such as recognition and interpretation. However, solely, it cannot be used for recognition task. Taking this into consideration, the output of the segmentation algorithm should be as similar as the input image, however, described through a restricted color domain.

On the other hand, the Jaccard similarity coefficient is a well-known metric used to compare two sets, given by Eq. (33), where A and B are sets of classified pixels. In this work, we used Jaccard to measure how close the output of a segmentation process is to the supervised brain images of BRATS dataset.

$$J(A, B) = \frac{A \cap B}{A \cup B} \tag{33}$$

The third metric used here is the Dice coefficient which is equivalent to Jaccard since the relationship between them is given by Eq. (34). We choose this metric for a broader literature comparative purposes only.

$$D(A, B) = \frac{2.J(A, B)}{1 + J(A, B)} \tag{34}$$

Finally, the effects of the non-extensive Tsallis entropy as evaluation function of the algorithms will also be analyzed in this paper as a metric of quality of the algorithms.

The overall experiments pipeline is illustrated in Fig. 1.

6. Results

In this paper, we claim that the use of non-extensive entropy as a segmentation objective function can contribute to bio-inspired algorithms to find the optimal thresholding(s) value(s). Therefore, it is important to establish a relationship between the q-value, the bio-inspired algorithm used and a metric of segmentation quality.

As usual, one of the first steps of Computer Aided Diagnosis (CAD) systems is the image segmentation process. Note that this process does not require semantic knowledge nor region recognition. The idea is to simplify the image representation without losing information.

Hence, in this paper we use the PSNR (Eq. (32)) to measure the difference between the segmented image and the original one. Note that the higher the value of the PSNR the better the segmentation process. This idea has been used by several authors to measure the quality of image segmentation when there is no use of ground truth [5]. In this case, the quality of measure, given by the PSNR, is given in relation to the original image itself. Thus, in Eq. (32), I_1 stands for the original image and I_2 for the corresponding segmented image.

Fig. 2 shows the average PSNR measure obtained from all meta-algorithms for different values of q in objective function, and Fig. 3 shows the areas under each curve of Fig. 2 ranked from highest to lowest in terms of PSNR. In this figure, one can observe that PSNR values between most of algorithms are similar each other. However, it is interesting the fact that PSNR decreases as q approaches to 0.6 and establishes at 2.0. Furthermore, all algorithms achieved better results as q approaches 0.1.This is a strong evidence that the studied problem, in this case segmentation, fits better for super-extensive system models than sub-extensive ones. Although there is a high correlation between algorithms, Krill Herd had the worst results and Firefly had the best one.

Considering the time consuming for each algorithm, we also studied here the number of generations (iterations) of each algorithm for each value of q. The results are shown in Figs. 4 and 5.

According to these figures, all algorithms took less iterations and absolute time when $q \simeq 2.0$. For lower values of q, one can observe that there is a great difference between themselves. Furthermore, although Cuckoo Search (CS) had more iterations for $q \simeq 0.1$, each iteration is more efficient in terms of absolute time. On the other hand, Firefly algorithm (FF) had the worst absolute time compared to others algorithms. Finally, *KH* was the bio-inspired algorithm that converged with fewer iterations.

Another interesting metric to evaluate is the normalized q-entropy value. This value is computed using Eq. (3) with a normalization factor. Fig. 6 shows the average normalized entropy computed for each value of q. Since the algorithms

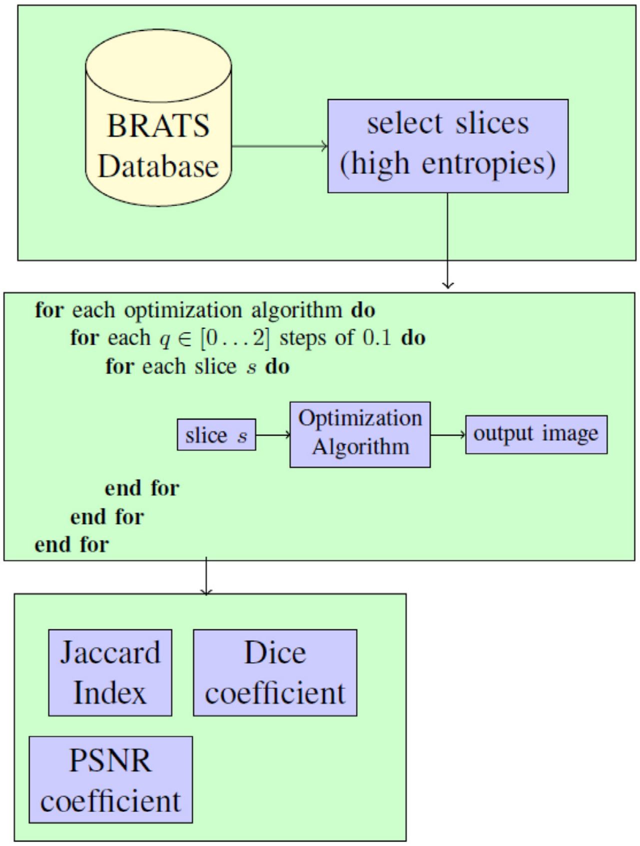

Fig. 1. Experimental pipeline.

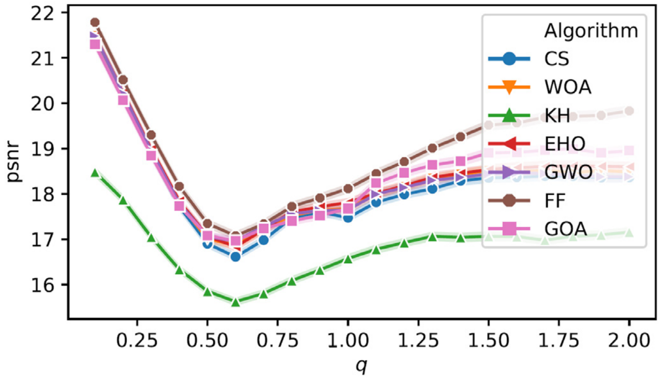

Fig. 2. Average PSNR between segmented image and the corresponding original (by each algorithm) for different values of q.

evaluated here use this objective function, these values are the average of the highest entropy achieved at each run of each algorithm.

The most notable observation between Figs. 2 and 6 is their high correlation, specially in the range $0 < q < 1$. Again, all algorithms behave similarly and entropy gets higher as q approaches to 2.

Furthermore, not only this high correlation between results from Figs. 2 and 6 suggests the positive effects of small values of q, but also the small values of those figures are near each other. This is a coherent results since entropy measures the amount of information in a physical system and PSNR measures how much information was preserved from the original image. Therefore, for example, when $q = 0.5$ PSNR shows that segmented image and original image are different themselves (Fig. 2) and normalized entropy shows that there is less information in the physical system as shown the corresponding Fig. 6.

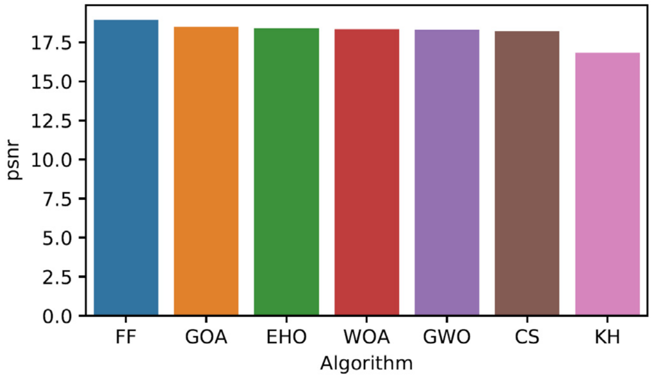

Fig. 3. Comparison of PSNR sorted by area under the curves shown in Fig. 2.

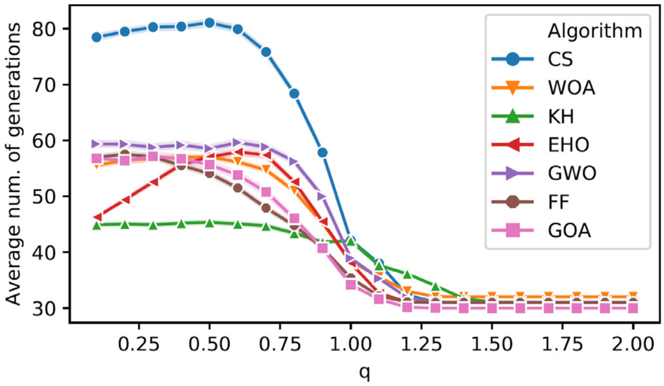

Fig. 4. Average number of generations for each algorithm as a function of *q*.

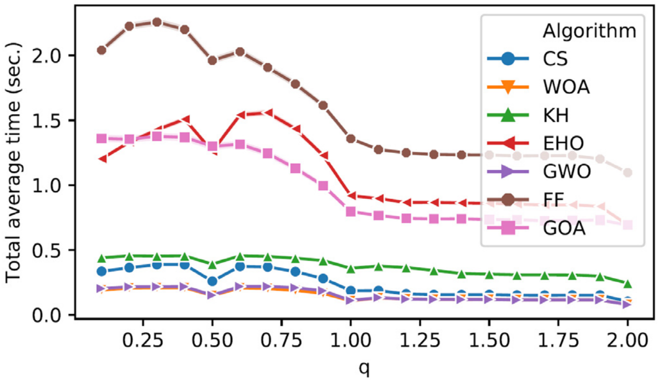

Fig. 5. Average time for each algorithm as a function of *q*.

These results motivated us to investigate the behavior of objective function. Fig. 7 shows the results of an optimization made in one image of the database that was segmented by two thresholds with KH algorithm (recall that KH was the worst algorithm in terms of PSNR according to Fig. 2

In this figure, the first column corresponds to the optimization plane created by the objective function (non-extensive entropy) using two thresholds. The *y*-axis represents the values of the first threshold and the *x*-axis corresponds to the second threshold. The second column is the image histogram with the thresholds cuts. Finally, the third column plots the sorted values of the optimization plane. Each row of this image was computed taken different values of $q \in \{0.1, 0.5, 1.0, 1.5, 2.0\}$.

First, we will consider the special case $q = 1$, the traditional additive entropy objective function (Eq. (2)). The figure shows that the found solution (blue dot) is near to the optimal one (red dot). However, the *psnr* value is higher when $q = 0.1$ compared to $q = 1.0$, so there is a better thresholding solution to the image. The issue here is that the traditional

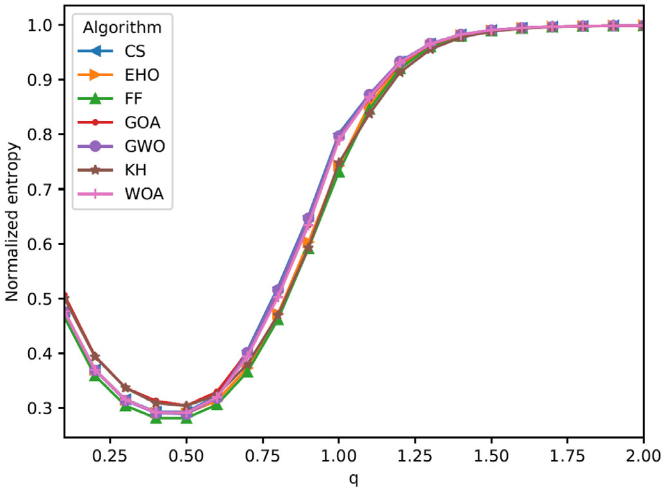

Fig. 6. Average normalized entropy for each value of q.

entropy could not model the search space correctly; that is, the algorithm found a near to optimal value, but this optimal value is not correlated to the best values of *psnr*.

Frequently, when working with medical images, the number of occurrences of black pixels (corresponding to the background) is very high. Therefore, the histogram presents a peak around the zero 0 bin. Since the approach taken in this work to segment the images is based on information theory, each bin of the histogram h corresponds to a state in a system and the objective is to find a partition (or set of partitions) that balances the information of all systems. However, since the magnitude of the first bin of h is much higher compared with others states, a plausible solution for the balancing is to set $h(0) = 0$. This explains the behavior of the histogram of Fig. 7 for $q = 1.0$ (column two, row three).

When looking to the non-extensive results, the effects are quite different. Since Tsallis Entropy uses $q \in \mathbb{R}$ to power the system probabilities, this parameter can give more (or less) importance for higher probabilities. When $q < 1$, higher probabilities values of histogram will decrease their magnitude. The effect is that such values receive less importance when searching for thresholding values. This explains the optimization planes (first column) in Fig. 7.

Another important observation is that the sorted values of optimization plane (third column) behaves as a single shape (with less inflection points and more positive second derivatives). This kind of shape helps the optimization algorithm to find the maximum point, since most values differ themselves.

On the other hand, when $q \sim 2$, the effect of background pixels over the system are even higher. In addition to this, as q approaches to 2, the objective function morphs into a plateau. This result also reinforces the fact that high values of q make the optimization algorithms finish earlier, since the stop criteria is based on changes in objective function.

The discussion of these results where carried out over KH results, which was the algorithm with the worst results as can be observed in Fig. 2. Nevertheless, all observations made here for KH approach are still valid for the remaining algorithms since the objective functions are the same.

As explained before, the difference of magnitude between occurrences of background pixels and other values of the histogram can affect significantly the results. Thus, we carried out the same experiment but seting $h(0) = 0$ (first bin of histogram). Fig. 8 illustrates two image histograms. The left histogram is the probability distribution of all pixel values and, the right is the probability distribution of all pixel values higher that 0 (strictly non zero entries) of the same image.

The results of the optimization are shown in Fig. 9, where we can observe that, for all values of q, the optimal points are the same. This evidence reinforces the fact that the background has strong interference in optimization process. Moreover, since the optimal point is the same, we can conclude that the optimization is better when $q < 1$. On the other hand, when $q > 1$, we find a plateau similar to that one in Fig. 7 (Figs. 10 and 11).

These behaviors can be clearly observed by comparing the three different segmentations for different slices shown in Figs. 10 and 11) First, Fig. 10 shows the segmentation with the algorithm KH for three values of $q = \{0.1, 0.5, 1.5\}$, while Fig. 11 shows the same process but setting $h(0) = 0$.

Also, we can observe two behaviors: (i) when $q < 1$, background intensities have less interference over the segmentation; (ii) bio-inspired optimization algorithms can achieve better precision in terms of segmentation quality since the objective function is steeper.

The parameters and settings of the algorithms compared here were suggested by the literature that was proposed or applied. In this paper, the GWO algorithm was executed with the parameters defined by the author in the article [8,43], which compared with the PSO, and three other evolutionary algorithms.

Regarding the GOA algorithm, it the work of [45] was tested in unimodal, multimodal, and composite functions with the same parameters used in this work. Also, the comparison was made with other algorithms.

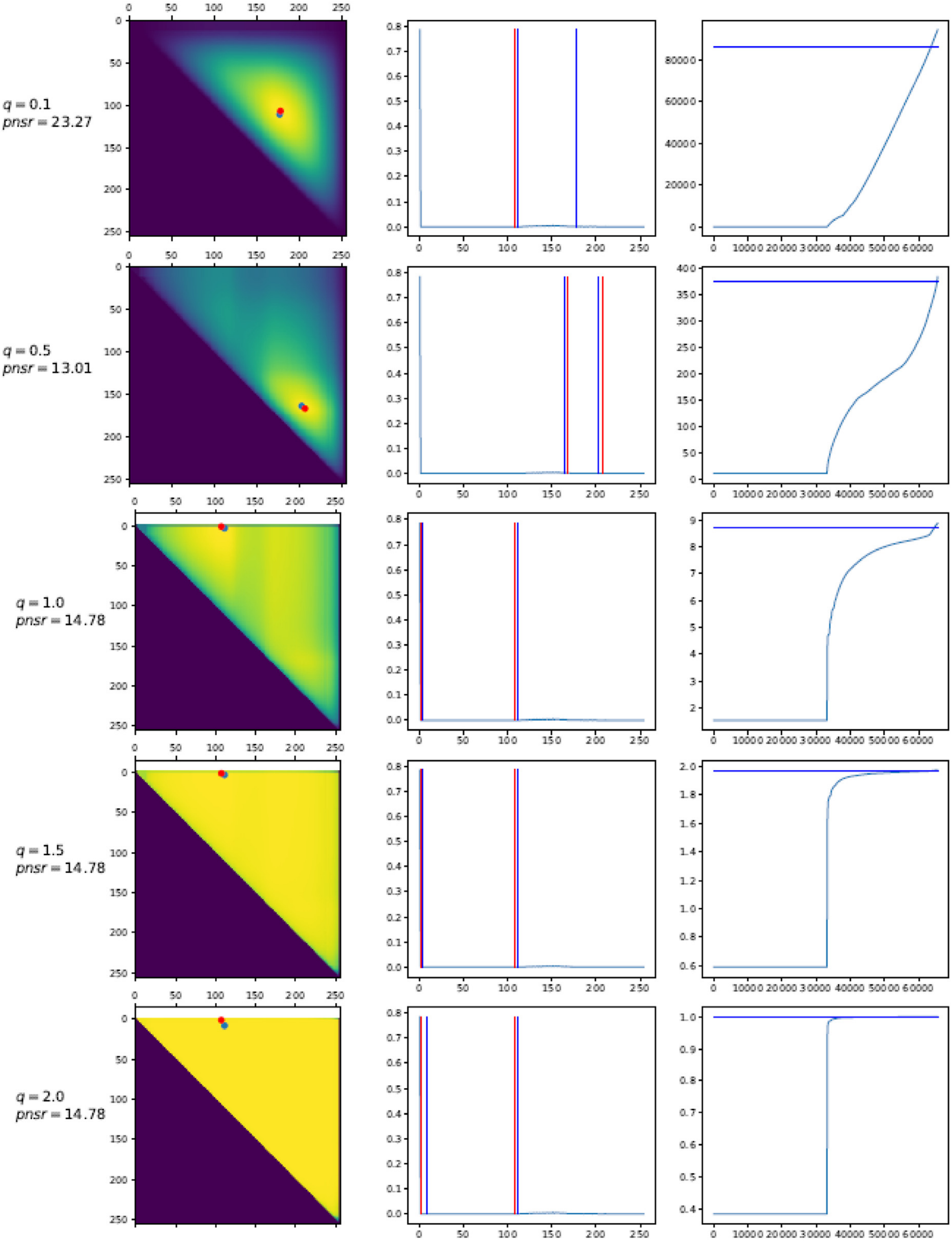

Fig. 7. Performance of the KH algorithm for two thresholds. Left: combination of the two ideal thresholds (red point) and the two thresholds found (blue point) under all possible psnr values (yellow are greater psnr than black); Center: comparison of the two ideal thresholds (red vertical lines) with two thresholds found by KH (blue vertical lines); Right: sorted values of psnr from left column and the cut threshold found (blue horizontal line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

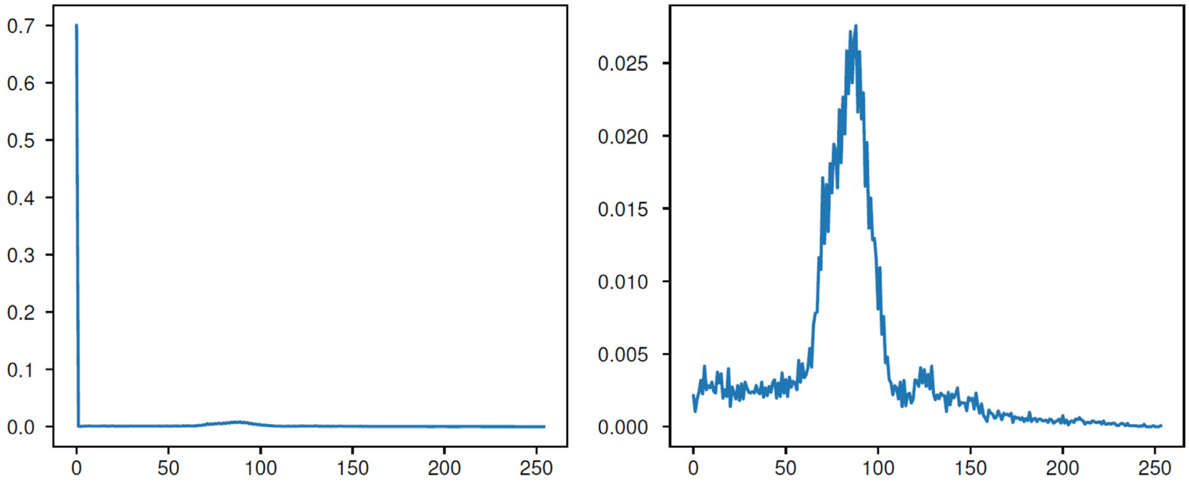

Fig. 8. Two histograms of the same image. The left histogram considers all states (including the background intensity $h(0)$). The right histogram excludes the background intensity $h(0)$.

Similarly, the results achieved by the WOA algorithm concerning the other algorithms in the literature [60] proved to be prominent or similar to different algorithms. So, those same parameters were used here in our comparison.

The work of [24], who proposed the Firefly algorithm using non-extensive entropy, used parameters similar to those suggested by the author of the algorithm [5]. Thus, in this work, the same parameter specifications for this algorithm were followed. In turn, for CS, this article adopted parameters that were used successfully in consolidated benchmark functions [6].

Besides, the KH algorithm was tested in four different variations and compared with eight other algorithms in [25]. Thus, twelve algorithms were considered. Then, the KH with crossover operator obtained the best result among the twelve methods, so this specific variation of the algorithm was considered in this work. Finally, the EHO algorithm [9] was compared with three other algorithms in fifteen benchmark functions.

Thus, in this work, the ideal configuration of the algorithms was maintained, optimizing the parameterization according to the specifications of each author and their respective proposed models, keeping the comparison of the algorithms fair.

7. Discussion

The graphs shown in Fig. 4 show a comparison of convergences in terms of the average number of generations for different values of the non-extensibility parameter q. It can be observed that all the algorithms studied here converge to the same low number of iterations when the q parameter is greater than or equal to 1.0. Besides with the behavioral observations shown in Figs. 7 and 9, for the same range of q value, it is noted that these convergences may indicate local minimums reached due to the strong influence of the background for this class of images. Thus, these results suggest the q values less than 1.0, the so-called super-extensive system, for this type of image. Also, all algorithms present low convergence rates, with KH having the fastest and most stable convergence, and CS having the most significant convergent iterations.

In Fig. 5, we can observe the convergences regarding the average meantime. This figure shows some stability for all algorithms studied here, being them KH, CS, GWO, WOA, regardless of the value of q. On the other hand, the FF, GOA, and EHO algorithms present a harder delay in the range of q values less than 1.0 when they reach a global minimum.

These observations show that both the convergence time and the number of iterations have acceptable values for all algorithms. In terms of general time, both are similar, but it was GWO and WOA that had the lowest total times. Also, it can be noted that the value of $q = 0.5$ (see in Fig. 2) shows the lowest error rates. Since, for $q = 0.5$, the number of iterations and total average-time are the smallest for most algorithms, this value of q seems to be the most recommended for these algorithms when used in this class of images.The reasons for why these convergences occur are discussed in the following.

The nature-inspired algorithms studied in this article are all representatives of the swarm category, having been created to solve optimization problems in general. All of them have an evaluation function that, maximized, influences the convergences for the best results. The strategy presented here is to use these functions as non-extensive additivity, based on Tsallis's non-extensive entropy. According to the theory demonstrated in [66], it is proved that non-extensive entropy also encompasses Shannon's traditional extensive entropy, when $q = 1.0$. This means that the results obtained with the conventional Shannon method are also evaluated when the evaluation function is extensive when $q = 1.0$. Therefore, it is expected that better results can be obtained in some range of values for $q \neq 1.0$.

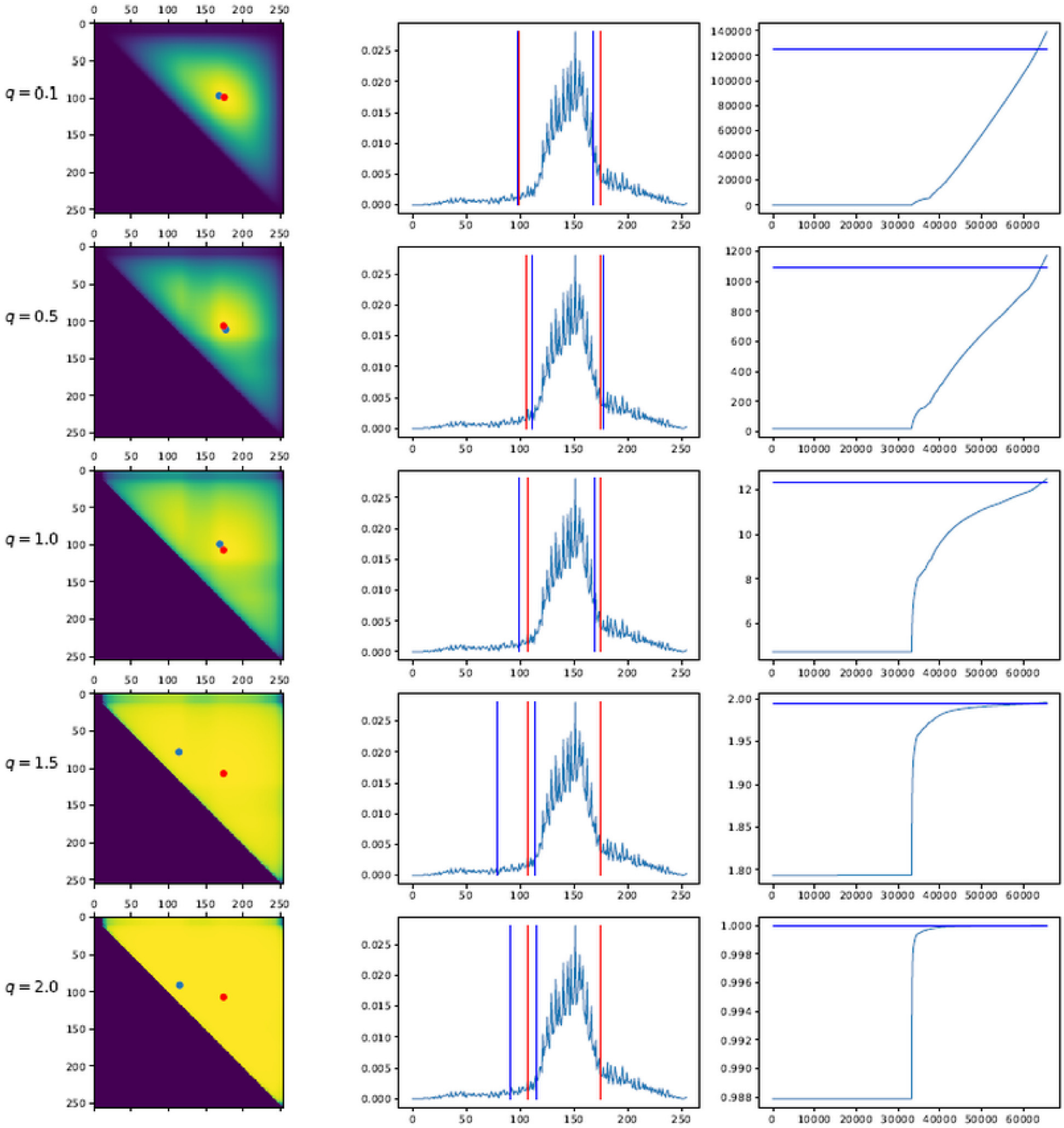

Fig. 9. Results for KH algorithm for one image of the database removing the occurrences of background intensities. The first column of graphics stands for the optimization plane considering two thresholds (the red dot corresponds to the maximum point and the blue dot corresponds to the output of the optimization algorithm). The second column stands for the image histogram with the thresholding cuts (red for the optimal cuts and blue for the optimization result). The third column stands for the sorted values of optimization plane. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This hypothesis seems to be confirmed in Figs. 4 and 7, where it is noted that the best convergences are obtained for values of $q < 1.0$, the so-called super-extensive systems. However, there is still no formal way in the literature to calculate the value of q that maximizes the results of convergence, neither approaching zero nor concerning time.

Thus, the automatic calculation of the ideal value of q is a disadvantage of the proposed methodology and an open topic. However, the range of values for $q < 1.0$ seems to approximate the ideal results. The same occurs for the parameterization

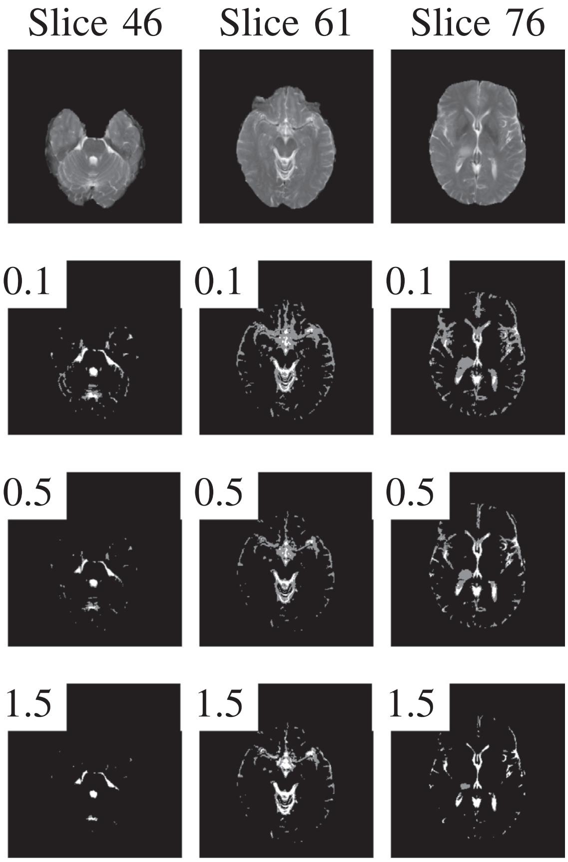

Fig. 10. A visual comparative example of the segmentation using the KHA for three slices from the BRATS database and the respective outputs for $q \in \{0.1, 0.5, 1.5\}$. Each column corresponds to slice and each line corresponds to a value for q. Two thresholds and 100 generations were considered in this case. Also, bin $h(0)$ was considered in the segmentation.

of the algorithms, which followed the suggestions in the literature. Since in this article, we show that the value range of $q < 1.0$ recovers optimal values, it can be said that the disadvantage, in this case, is mitigated.

Observing the general behavior of the algorithms studied here, we can divide them into two convergence strategies: one that converges in large groups of agents (GOA, EHO, and KH) and another that converges in smaller groups or individually (WOA, GWO, FF, and CS). These two strategies seem to have an impact both on the number of iterations and on the convergence time of the algorithms, depending on the topology of the threshold distribution (search space). The topologies of the search spaces seem to influence more on the final value of the thresholds found and, therefore, on the convergence of the number of generations, as demonstrated in [24], and also observed in Figs. 7 and 9. The more homogeneous areas the image being segmented has, the more plateaus in the distribution will occur. And the more significant the plateau of a threshold distribution, the smaller the number of iterations. The existence of plateaus favors the search for optimal thresholds under large group-based algorithms.

However, a lower number of convergence does not precisely mean a direct correlation with the total time of convergence, since at each iteration the calculation of the movements of the agents in each algorithm can be excessive, depending on individual factors or the position of other agents, increasing computational time and, consequently, the total convergence time. Only if an algorithm has a low number of generations combined with a shorter time to converge, it leads to an ideal combination of convergence and time. Figs. 7–9 show the influence of the distribution's plateaus on the number of generations, as well as on the time of total convergence. In the case of the images used in this work, the plateaus are due to the large dark homogeneous areas in the backgrounds for this type of image.

Thus, a hypothesis that can be raised here is that algorithms based on group convergence, such as GOA, EHO, and KH, are more adapted to images with more massive plateaus, and algorithms based on more individual convergences, such as WOA, GWO, FF, and CS, tend to adapt better to distributions with smaller plateaus.

Based on these ideas, each of the algorithms is evaluated below.

The FF is an algorithm where each agent (firefly) converges individually, being, therefore, more suitable for images with smaller plateaus. In images with long plateaus (more homogeneous areas), this algorithm can have a higher convergence time, as shown in Fig. 5. However, the number of epochs is compatible with most algorithms studied here.

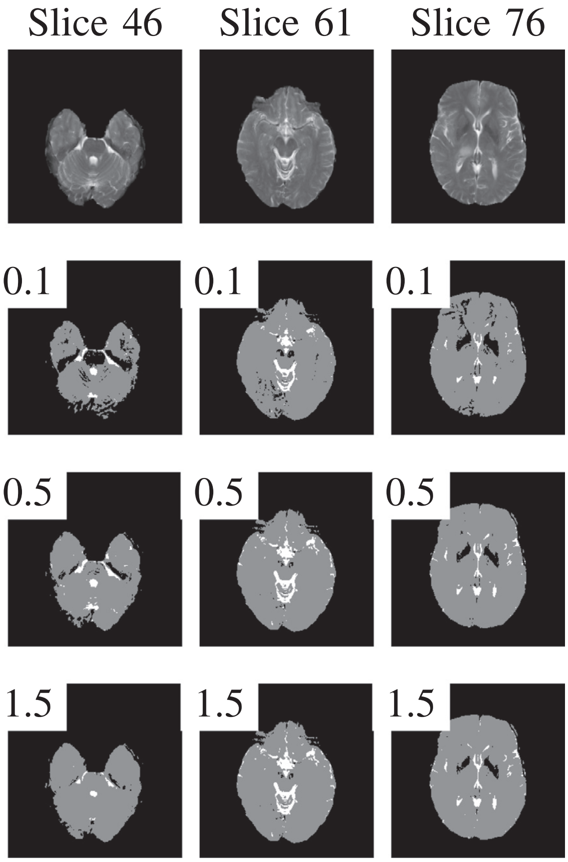

Fig. 11. A visual comparative example of the segmentation (setting histogram bin $h(0) = 0$) using the KHA for three slices from the BRATS database and the respective outputs for $q \in \{0.1, 0.5, 1.5\}$. Each column corresponds to slice and each line corresponds to a value for q. Two thresholds and 100 generations were considered in this case.

Like WOA, GOA, and FF, CS is an algorithm based on the individual convergence of its agents (the cuckoos), suggesting the same limitations in the case of images with topologies of long plateaus distributions. This behavior is reinforced by the levy-Flighty swarm pattern. Among all methods studied here, the CS has the most significant number of iterations, as shown in Fig. 4. However, this number of iterations is reached in a time compatible with the fastest ones: GWO, WOA, and KH. This low convergence time may be linked to its low linear computational complexity, $O(P\log(P)) + O(PD)$, as calculated above. Unlike most algorithms, Cukoo Search does not compare your solution with the other agents, but only estimates the probability of the solution of each algorithm to be good or not.

Unlike FF, CS, WOA, and GWO, GOA, similar to EHO and KH, is a nature-inspired algorithm where solutions are proposed in groups, subdividing the swarm into competitive parts, as if they were large agents. This strategy can better adapt these algorithms to images with significant plateau, producing lower iteration rates (see Fig. 4), but the computational cost can increase the convergence time (see Fig. 5) due to the need for constant calculation of intra-individual solutions. In the particular case of GOA, the complexity, as previously shown, is of the order of $O(N^2T)$. In Fig. 5, GOA was only the 5*th* faster in terms of convergence time.

The GWO is also a nature-inspired algorithm based on agents that propose individual solutions, for this reason, it is more adapted to intensity distributions with fewer plateau topologies. Thus, as shown in Fig. 4, their convergence in the number of iterations is similar to that of FF and WOA. However, due to its low computational complexity ($O(TN)$, as described above), it presented the best convergence time among all the algorithms tested here, having only WOA as its competitor.

The EHO is an algorithm based on the convergence of agents in groups (called clans). The best solutions depend on similar solutions within each clan, which makes this algorithm more suitable for working with images with many plateaus, reducing the number of converging generations (see Fig. 4). Although the number of agents is divided into groups (clans), its complexity is a quadratic function of the total number of agents, as shown before, which makes its computational time similar to that of the algorithms studied here with longer times, such as FF (see Fig. 5).

The KH is the algorithm that converges with the least number of iterations, and in terms of time, it is compatible with the fastest. As it is also an algorithm based on groups, it is more adapted to images such as those used in this article, as they are images with vast plateaus. Among the seven studied algorithms, it is the only one that has a complexity of N^2 but converges in a short time, in comparison with the others. The hypothesis of the reason for this fast convergence time is that KH, for the calculation of inter-individual dependency, uses only the dependency neighborhood. Therefore, the groups of agents and solutions (krills) are smaller than the others based on groups (WOA and GOA).

The WOA is the fourth algorithm studied here that is based on individual behavior, parameters, the distance between each agent, and the distance of each agent to another agent chosen at random, which leaves the complexity for finding the best solution around the order of $O(N)$. Basically, there are two stages in this algorithm: the first is that of convergence in a spiral direction towards the hunting, and the second is that of spiral formation (the relative location between agents). Due only to these two alternating stages, and the order of complexity by iteration $O(N)$, WOA has a total linear complexity of the order of $O(NT)$. Thus, due to the randomness of the alternation between the two stages, the number of generations is low but compatible with most of the algorithms studied here (see Fig. 4). However, since its convergence time is also short (see Fig. 5), it becomes one of the two fastest algorithms among all studied here, along with GWO.

8. Conclusion

Recently, the non-extensive Tsallis entropy has been used as a promise alternative for image thresholding. When the entropy is maximum many physical systems, such as images, get better equilibrium. However, finding the optimal thresholding values to obtain the highest entropy is not a trivial task, specially when working with multi-thresholding image segmentation. The brute-force algorithm is not a feasible solution for such problem, since it scales in exponential complexity order.

Nevertheless, recent developments of bio-inspired optimization algorithms brought new alternatives that fit into this problem. Since then, few researchers have explored such algorithms to solve the multi-thresholding image segmentation.

Although these works found promising results in terms of image segmentation, the benchmark of such algorithms and how the non-extensive Tsallis entropy contributes to the convergence and quality of results in the context of image segmentation is a topic of study that should be explored in more details.

In this paper, we explored seven bio-inspired optimization algorithms considering the non-extensive Tsallis additive equation as objective function. Those algorithms are compared in terms of quality and convergence time in the context of medical images. Moreover, we study some benefits of using Tsallis theory for such types of algorithms.

Results show that all algorithms behave similarly when considered the quality of segmentations. However, Krill Herd had the worst results. The Grey Wolf and Whale Optimization algorithms had the fastest performance in terms of time. These algorithms are among the most recent proposed, however they already have great interest among researchers.

Moreover, one of our main findings is that the q parameter of Tsallis theory changes the optimization hiper-plane increasing the differences between possible solutions and highlighting the optimal one. This effect is observed as q approaches to 0 and this behavior also contributes to balance background intensities with the foreground ones.

The results found in this paper brings new fronts of research. The first one is to study how to compute previously the best value of q based on the image histogram distribution. However, our experiments show the ranges of values that lead to the best performances. The second one is to test other objective functions including other versions of entropy. The last one is to test all these algorithms in higher spaces, such as, color images.

Although all results from this paper comes from empirical benchmarks, they explore new paths of researches on image segmentation and information theory. We hope that advances in segmentation step is fundamental in an object recognition pipeline. Moreover, the use of Tsallis entropy in bio-inspired algorithms seems to be a promising segmentation technique.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

G.A. Wachs-Lopes: Conceptualization, Software, Validation. **R.M. Santos:** Software, Validation, Writing - review & editing. **N.T. Saito:** Software, Validation. **P.S. Rodrigues:** Conceptualization, Writing - original draft, Validation, Writing - review & editing, Supervision.

References

[1] Poli R, Kennedy J, Blackwell T. Particle swarm optimization. Swarm Intell. 1995;1:33–57.
[2] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing.. Science 1983;220 4598:671–80.
[3] Dorigo M. Optimization, learning and natural algorithms; 1992.
[4] Glover FW. Tabu search - part I. INFORMS J Comput 1989;1:190–206.
[5] Horng M-H. Multilevel minimum cross entropy threshold selection based on the honey bee mating optimization. Expert Syst Appl 2009;37:4580–92.
[6] Yang X, Deb S. Cuckoo search via lvy flights. In: 2009 World Congress on Nature Biologically Inspired Computing (NaBIC). p. 210–214. doi:10.1109/NABIC.2009.5393690.
[7] Beevi K S, S Nair M, G R B. Automatic segmentation of cell nuclei using krill herd optimization based multi-thresholding and localized active contour model. Biocybernet Biomed Eng 2016;36. doi:10.1016/j.bbe.2016.06.005.
[8] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw 2014;69:46–61. doi:10.1016/j.advengsoft.2013.12.007.

[9] Wang G-G, Deb S, Coelho LdS. Elephant herding optimization. In: Proceedings of the 2015 3rd International Symposium on Computational and Business Intelligence (ISCBI). IEEE Computer Society; 2015. p. 1–5. doi:10.1109/ISCBI.2015.8. ISBN 978-1-4673-8501-5

[10] Kumar CS, Rao RS. A Novel Global MPP Tracking of Photovoltaic System based on Whale Optimization Algorithm. Int J Renew Energy Dev 2016;5(3):225–32. doi:10.14710/ijred.5.3.225-232. 00011

[11] Liang H, Jia H, Xing Z, Ma J, Peng X. Modified grasshopper algorithm-based multilevel thresholding for color image segmentation. IEEE Access 2019;7:11258–95. doi:10.1109/ACCESS.2019.2891673.

[12] Alkemade A, Keuken MC, Forstmann BU. A perspective on terra incognita: uncovering the neuroanatomy of the human subcortex. Front. Neuroanat.; 2013.

[13] Pun T. Entropic thresholding, a new approach; 1981.

[14] Kapur JN, Sahoo PK, Wong AKC. A new method for gray-level picture thresholding using the entropy of the histogram. Comput Vis Graph Image Processing 1985;29:273–85.

[15] Hassanzadeh T, Vojodi H, Eftekhari-Moghadam A-M. An image segmentation approach based on maximum variance intra-cluster method and firefly algorithm. 2011 Seventh international conference on natural computation 2011;3:1817–21.

[16] Luiz C. Sampaio MPd. A, de Menezes FS. Nonextensivity and tsallis statistics in magnetic systems 55(9):5611–5614. doi:10.1103/PhysRevB.55.5611.

[17] Oliveira IS. Some metallic properties in the framework of tsallis generalized statistics. Eur Phys J B - Condensed Matter Complex Syst 2000;14:43–6.

[18] Martin M, Plastino A, Plastino A. Tsallis-like information measures and the analysis of complex signals. Physica A 2000;275(1):262–71. doi:10.1016/S0378-4371(99)00424-0.

[19] Tong S, Bezerianos A, Paul J, Zhu Y, Thakor N. Nonextensive entropy measure of eeg following brain injury from cardiac arrest. Physica A 2002;305(3):619–28. doi:10.1016/S0378-4371(01)00621-5.

[20] Landsberg PT, Vedral V. Distributions and channel capacities in generalized statistical mechanics. Phys Lett A 1998;247(3):211–17. doi:10.1016/S0375-9601(98)00500-3.

[21] Capurro A, Diambra L, Lorenzo D, O M, Martin M, C M, et al. Tsallis entropy and cortical dynamics: the analysis of eeg signals. Phys A Stat Theor Phys 1998;257. doi:10.1016/S0378-4371(98)00137-X.

[22] Eberhart RC, Kennedy J. A new optimizer using particle swarm theory; 1995.

[23] Yang X-S. Nature-inspired metaheuristic algorithms; 2008.

[24] Rodrigues PS, Wachs-Lopes GA, Erdmann HR, Ribeiro MP, Giraldi GA. Improving a firefly meta-heuristic for multilevel image segmentation using tsallis entropya; 20:1–20. doi:10.1007/s10044-015-0450-x.

[25] Gandomi A, Alavi A. Krill herd algorithm: a new bio-inspired optimization algorithm. 2012.

[26] Agrawal PK, Pandit M. Improved krill herd algorithm with neighborhood distance concept for numerical optimization. In: 2016 IEEE International Conference on Engineering and Technology (ICETECH); 2016. p. 1105–10. doi:10.1109/ICETECH.2016.7569421.

[27] Wang G, Gandomi AH, Alavi AH, Deb S. A multi-stage krill herd algorithm for global numerical optimization. Int J Artif Intell Tools 2016;25:1–17.

[28] Wang G, Guo L, Gandomi AH, sheng Hao G, Wang H. Chaotic krill herd algorithm. Inf Sci 2014;274:17–34.

[29] Abdel-Basset M, Hessin A-N, Abdel-Fatah L. A comprehensive study of cuckoo-inspired algorithms. Neur Comput Appl 2016;29:345–61.

[30] Brajevic I, Tuba M, Bacanin N. Multilevel image thresholding selection based on the cuckoo search algorithm. Advances in sensors, signals, visualization, imaging and simulation; 2012.

[31] Ilunga Mbuyamba E, Cruz-Duarte J, Avina-Cervantes J, Rodrigo Correa-Cely C, Lindner D, Chalopin C. Active contours driven by cuckoo search strategy for brain tumour images segmentation 56. doi:10.1016/j.eswa.2016.02.048.

[32] Kamat S, Karegowda AG. A brief survey on cuckoo search applications. In: Proceedings of international conference on advances in computer & communication engineering; 2014.

[33] Malik M, Ahsan F, Mohsin S. Adaptive image denoising using cuckoo algorithm. Soft comput 2016;20:925–38.

[34] Mousavirad SJ, Ebrahimpour-Komleh H. Entropy based optimal multilevel thresholding using cuckoo optimization algorithm. 11th Int Conf on Innov in Inf Tec (IIT) 2015:302–7.

[35] Ouyang A, Pan G, Yue G, Du J. Chaotic cuckoo search algorithm for high-dimensional functions. JCP 2014;9:1282–90.

[36] Yang X-S, Deb S. Cuckoo search: recent advances and applications. Neural Comput Appl 2013;24:169–74.

[37] Ye Z, Wang M, Hu Z, Liu W. An adaptive image enhancement technique by combining cuckoo search and particle swarm optimization algorithm. Computational Intelligence and Neuroscience; 2015.

[38] Zhao W, Ye Z, Wang M, Ma L, Liu W. An image threholding approach based on cuckoo search algorithm and 2d maximum entropy. 2015 IEEE 8th international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS) 2015;1:303–8.

[39] ge Wang G, Deb S, dos S Coelho L. Elephant herding optimization. In: 2015 3rd international symposium on computational and business intelligence (ISCBI); 2015. p. 1–5.

[40] Deb S, Fong SJ, Tian Z. Elephant search algorithm for optimization problems. In: Tenthinternational conference on digital information management (ICDIM); 2015. p. 249–55.

[41] Tian Z, Wong RK, Mohammed S, Deb S, Fong S, Fiaidhi J. Finding approximate solutions of nphard optimization and tsp problems using elephant search algorithm. J Supercomput 2016;72:3960–92.

[42] Tuba E, Alihodzic A, Tuba M. Multilevel image thresholding using elephant herding optimization algorithm. In: 14th International Conference on Engineering of Modern Electric Systems (EMES); 2017. p. 240–3. doi:10.1109/EMES.2017.7980424. ISSN:

[43] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer 69:46–61. doi:10.1016/j.advengsoft.2013.12.007.

[44] Mirjalili SM, Lewis AW. The whale optimization algorithm. Adv Eng Softw 2016;95:51–67.

[45] Saremi S, Mirjalili S, Lewis A. Grasshopper optimisation algorithm: theory and application. Adv Eng Softw 2017;105:30–47. doi:10.1016/j.advengsoft.2017.01.004.

[46] Tsallis C. Possible generalization of boltzmann-gibbs statistics. J Stat Phys 1988;52:479–87.

[47] Portes de Albuquerque M, Esquef IA, Gesualdi Mello AR, Portes de Albuquerque M. Image thresholding using tsallis entropy 25(9):1059–1065. 10.1016/j.patrec.2004.03.003

[48] Rodrigues PS, Chang R-F, Suri JS. Non-extensive entropy for CAD systems of breast cancer imagesb;:121–128 doi:10.1109/SIBGRAPI.2006.31.

[49] Rodrigues PSS, Giraldi GA. Improving the non-extensive medical image segmentation based on tsallis entropy. Pattern Anal Appl 2011;14:369–79.

[50] Curado EM, Tempesta P, Tsallis C. A new entropy based on a group-theoretical structure. Ann Phys (N Y) 2016;366:22–31. doi:10.1016/j.aop.2015.12.008.

[51] Yang X-S. Nature-inspired metaheuristic algorithms. Luniver Press; ISBN 1-905986-10-6 978-1-905986-10-1.

[52] He L, Huang S. Modified firefly algorithm based multilevel thresholding for color image segmentation. Neurocomputing 2017;240. doi:10.1016/j.neucom.2017.02.040.

[53] Suresh K. Image multithresholding based on kapur and tsallis entropy and firefly algorithm. Indian J Sci Technol 2016;9. doi:10.17485/ijst/2016/v9i12/89949.

[54] Bhandari A, Kumar A, Chaudhary S, Singh G. A novel color image multilevel thresholding based segmentation using nature inspired optimization algorithms. Expert Syst Appl 2016;63:112–33. doi:10.1016/j.eswa.2016.06.044.

[55] Thomas RM, John J. A review on cell detection and segmentation in microscopic images. In: International conference on circuit,power and computing technologies (ICCPCT); 2017. p. 1–5. doi:10.1109/ICCPCT.2017.8074189. ISSN:

[56] Kayom Md Khairuzzaman A, Chaudhury S. Multilevel thresholding using grey wolf optimizer for image segmentation. Expert Syst Appl 2017;86.

[57] Zidan A, Fouad A, Houseni M, Allam N, Hassanien AE, Hefny H. A hybrid grey wolf based segmentation with statistical image for ct liver images. In: Proceedings of the international conference on advanced intelligent systems and informatics; 2017.

[58] Tuba E, Alihodzic A, Tuba M. Multilevel image thresholding using elephant herding optimization algorithm. In: 2017 14th international conference on Engineering of Modern Electric Systems (EMES); 2017. p. 240–3. doi:10.1109/EMES.2017.7980424.

[59] Wachs-Lopes GA, Beltrame FS, Santos RM, Rodrigues PS. Comparison of bio-inspired algorithms from the point of view of medical image segmentation. In: 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI); 2018. p. 1–7. doi:10.1109/IWOBI.2018.8464218.

[60] Mirjalili S, Lewis A. The whale optimization algorithm. Adv Eng Softw 2016;95:51–67.

[61] Aziz MAE, Ewees AA, Hassanien AE. Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation. Expert Syst Appl 2017;83(C):242–56. doi:10.1016/j.eswa.2017.04.023.

[62] Mostafa A, Hassanien AE, Houseni M, Hefny H. Liver segmentation in MRI images based on whale optimization algorithm. Multimed Tools Appl 2017;76(23):24931–54. doi:10.1007/s11042-017-4638-5.

[63] Ozturk U, Ozkaya U, Akdemir B, Seyfi L. Soft tissue sacromas segmentation using optimized otsu thresholding algorithms. Int J Eng Technol Manag Appl Sci 2017;5:49–54.

[64] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, et. al. The multimodal brain tumor image segmentation benchmark (BRATS) 34(10):1993–2024. doi:10.1109/TMI.2014.2377694.

[65] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data 2017;4:170117. doi:10.1038/sdata.2017.117.

[66] Tsallis C. Nonextensive statistics: theoretical, experimental and computational evidences and connections I introduction. Brazili. J. Phys. 1999. 1-35. ISSN 0103-9733