GE LOCAL COMMENTS

# Reviewer 1

## 1. Comment

The technique of simulated annealing is inadequately described, despite being quite critical to the paper's whole concept. A clear description of simulated annealing would be extremely helpful.

## Response

**(PERIGRAFI SA ISOS KAI SE DIAGRAMA ROIS)**

## 2. Comment

There are many tables of numbers in the results section, and they can be difficult to interpret. Perhaps some comments on those numbers, pointing out what is notable and important, combined with shading or emphasis of specific areas of the tables, would help the reader to interpret them.

## Response

**(COMMENTS ON THE RESULTS)**

## 3. Comment

Some of the graphs are also a bit hard to read. In particular, when the points are all clustered in one area as in Figure 4, it can be hard to read. The authors might consider whether a zoomed-in view of the denser part of the plot, or a log scale on the Y axis, could be a useful additional view for the reader.

## Response

**(ALEXANDROS)**

## 4. Comment

The authors might also consider whether some kind of graphical representations might help the readers to understand the problems and the solutions they are discussing. In many papers on optimization, example functions of two variables are shown as 3D surfaces, enabling readers to visualize issues such as local minima, the progress of algorithms as they approach optimality, and similar issues. As this is often done as I said, they may be able to find some existing figures for which they could obtain permission to include in the paper, as opposed to doing the work to make up their own.

## Response

**(DEN EGINE OPTIMIZATION EDO)**

## 5. Comment

Please ensure that all acronyms are defined at the time of first usage. As an example, the acronym BNF is used in the introduction without having been previously defined. There are also some spelling/typing errors, such as on the X axis of Figure 3, where a letter is omitted from the word "Classification". In addition, there are a few places where the grammar is slightly awkward.

## Response

Done.

The paper has been checked using the freely available tools of ispell and grammar check.

# Reviewer 2

## 1. Comment

In fact, this study belongs to a very classic research domain: neural architecture search. In this field, numerous studies have utilized metaheuristic algorithms to optimize parameters of neural networks. However, the literature review in this study is not comprehensive.

## Response

We have added the following paragraph at the end of Introduction:

“*In the same direction many researchers have published papers in neural network initialization or construction, such as the usage of decision trees to initialize the weights of a neural network[*[*nninit1*](#LyXCite-nninit1)*], initialization of the weights using Cauchy's inequality [*[*weight\_init2*](#LyXCite-weight_init2)*], application of discriminant learning [*[*weight\_init3*](#LyXCite-weight_init3)*]. Also, the issue of constructing the structure of artificial neural networks have been discussed in various papers, such as incorporation of genetic algorithms [*[*nngencon*](#LyXCite-nngencon)*], construction and prunning of the weights [*[*nnprunncon*](#LyXCite-nnprunncon)*], application of of cellular automata [*[*nncell*](#LyXCite-nncell)*] etc.* ”

## 2. Comment

The author has employed a hybrid of genetic algorithms and simulated annealing mechanisms, both of which are quite classical. The choice of these two algorithms needs further explanation. Additionally, the combination of algorithms is a well-established technique to overcome local optima issues. Therefore, the author needs to devote more space to elucidating the innovative aspects of this approach.

## Response

The following text has been added in the beginning of subsection 2.2, in order to clarify the selection of the Simulated Annealing method:

*“Simulated Annealing was chosen as a local search method, since it offers the possibility of representing the considered solutions in an integer form, which is critical for the Grammatical Evolution representation of chromosomes. In addition, this method has been distinguished for its easy adaptation to a multitude of problems but also for its ability to find the total minimum of functions through the point acceptance mechanism at high temperatures. In the present work, the Simulated Annealing will initiate from the current representation of a chromosome and gradually try to find other nearby representations that might lead to lower values of the fitness value.”*

## 3. Comment

Although there is an improvement in performance with the author's method, metaheuristic algorithms are known to be highly computationally intensive. Therefore, it is necessary for the author to provide an analysis of time complexity or a comparative account of actual computation times to ensure that the performance gains justify the computational costs.

## Response

We have added a time comparison between the original method and the three cases of the proposed method where the parameter g varies from 5 to 20. The added text in subsection 3.3 now reads:

“*In addition, in the graph of the figure* [*fig:Timecomparison*](#fig_Timecomparison)*, a comparison is made for the average execution time of each experiment for the classification datasets. The comparison was made between the initial method (denoted as NNC in the graph) as well as the various cases of the proposed technique by changing the parameter g to 5, 10 and 20. As expected, adding the local minimization technique to the Grammatical Evolution method significantly increases the execution time of the method, however this increase remains almost constant for different values of the critical parameter g . Moreover, by using parallel techniques that appear in the relevant literature this increase in execution time could be reduced.*”

## 4. Comment

The machine learning algorithms used by the author are too classic, which might reduce the credibility of the proposed technique's applicability to new neural networks. The author needs to offer more discussion on this aspect.

## Response

The following text has been added in subsection 3.3 to justify the selection of these methods:

“*The above techniques were used in the experiments as they are widespread in machine learning, such as the Adam method [*[*adam1*](#LyXCite-adam1)*,* [*adam2*](#LyXCite-adam2)*] and because they have a similar complexity to the present technique such as Genetic Algorithms.*”

# Reviewer 3

## 1. Comment

Since the authors propose a new optimization technique (i.e., an extension of the conventional Grammatical Evolution approach), the technique should also be tested on (a) benchmark optimization problems/test functions for optimization or (b) real-world optimization problems.

## Response

## 2. Comment

Please include comparative analysis on the execution time taken by the proposed approach versus the conventional approach.

## Response

We have added a time comparison between the original method and the three cases of the proposed method where the parameter g varies from 5 to 20. The added text in subsection 3.3 now reads:

“*In addition, in the graph of the figure* [*fig:Timecomparison*](#fig_Timecomparison)*, a comparison is made for the average execution time of each experiment for the classification datasets. The comparison was made between the initial method (denoted as NNC in the graph) as well as the various cases of the proposed technique by changing the parameter g to 5, 10 and 20. As expected, adding the local minimization technique to the Grammatical Evolution method significantly increases the execution time of the method, however this increase remains almost constant for different values of the critical parameter g . Moreover, by using parallel techniques that appear in the relevant literature this increase in execution time could be reduced.*”

## 3. Comment

Please include discussions on the potential for the proposed technique to be employed for multi-objective, multi-constraint, nonlinear, nonconvex  problems as well as problems with noisy/uncertain decision parameters.

## Response

**(STO CONCLUSIONS)**

## 4. Comment

Please include justifications on the reason for using simulated annealing and not other evolutionary techniques - e.g. genetic algorithm, differential evolution, swarm-based optimizer,..., etc.

## Response

The following text has been added in the beginning of subsection 2.2, in order to clarify the selection of the Simulated Annealing method:

*“Simulated Annealing was chosen as a local search method, since it offers the possibility of representing the considered solutions in an integer form, which is critical for the Grammatical Evolution representation of chromosomes. In addition, this method has been distinguished for its easy adaptation to a multitude of problems but also for its ability to find the total minimum of functions through the point acceptance mechanism at high temperatures. In the present work, the Simulated Annealing will initiate from the current representation of a chromosome and gradually try to find other nearby representations that might lead to lower values of the fitness value.”*

## 5. Comment

Please include limitations of the proposed approach (i.e., trade-offs compared with conventional techniques).

## Response

**(STO CONCLUSIONS)**

## 6. Comment

Please check and correct overall quality of English in the manuscript - e.g. correct 'In current work...' in the abstract.

## Response

**(NA GRAFEI POS EGINE CHECK ME ISPELL KAI GRAMMARCHECK)**