AI COMMENTS

# REVIEWER 1

## 1. COMMENT

I do not get the notation with an arrow above some parameters, do they mean a vector? Then maybe simply a bold font would be appropriate? Of course an arrow above a variable means a vector but rather as a Euclidean vector and it seems to not fit in this meaning. Besides g in eq. (3) seems to also be a vector, why it does not have an arrow? Please go through the equations and correct this issue.

## RESPONSE

We have add arrows in each vector of the manuscript to clarify this.

## 2. COMMENT

c\_i is not a "so-called center", it does not explain what it actually is, so please clarify it. Similar E in equation (3) is not a "so-called" training error, it is simply a loss function or training error.

## RESPONSE

a) The terminology center is used widely to express the vectors c\_i as mentioned also in wikipedia (see ref <https://en.wikipedia.org/wiki/Radial_basis_function>) as well as the recent review paper G.A. Montazer, D. Giveki, M. Karami, H. Rastegar, Radial basis function neural networks: A review. Comput. Rev. J 1, pp. 52-74, 2018.

b) The term training error is used as you mention for the loss function and it has been used widely in various research papers.

## 3. COMMENT

The section 2.1 starts with the sentence: "Grammatical evolution is a genetic algorithm where the chromosomes stand for the production rules of any given BNF (Backus–Naur form) grammar[47]."  
It is confusing for so many reasons. Please explain at first what are genetic algorithms.

## RESPONSE

The following phrase has been added at the beginning of subsection 2.1

“*Genetic Algorithms, suggested by John Holland [*[*Holland1*](#LyXCite-Holland1)*] are inspired by biology and the algorithm starts by creating an initial population of the so -called chromosomes that stand for potential solutions to the objective problem. These chromosomes are gradually altered using the genetic operators of selection, crossover and mutation[*[*Stender*](#LyXCite-Stender)*].* ”

## 4. COMMENT

Figure 1 has a caption above the Figure, while it should be belowe, but first of all I do not understand it, should it explain anything?

## RESPONSE

a) We have converted it to algorithm, since it is more suitable for this kind of information.

b) The caption has been changed to “*The BNF grammar used in the current work, to produce intervals for the RBF parameters. By using this grammar in the first phase of the proposed procedure, the optimal interval of values for the parameters of the neural network will be identified*”

## 5. COMMENT

What does % mean in Tab. 1? Is it also a mod function?

## RESPONSE

Yes, you are correct. We have replaced the % with the mod function.

## 6. COMMENT

Results do not seem "promising" as promised in the Abstract... Table 3 and 5 gives results in %, Table 4, also? What does the datasets contain? Maybe, because Authors use so many datasets it would be appropriate to describe those a bit more, give some examples? After reading the paper I am still confused with its contents. I understand a little bit what was done, but not enough. Authors write about RBFNN, but there is no NN in this paper described, no structure, no training. I also do not understand what was achived in the paper.

## RESPONSE

a) We have extended the abstract and now reads:

“*Radial basis function networks are widely used in a multitude of applications in various scientific areas in both classification and data fitting problems. These networks deal with the above problems by adjusting their parameters with various optimization techniques. However, an important issue to address is finding a satisfactory interval of values for the network parameters before adjusting these parameters. This paper proposes a two-stage technique, where in the first stage, using Grammatical Evolution, rules are generated to create the optimal value interval of the network parameters. In the second stage of the technique, the parameters of the network are fine-tuned with some robust global optimization method, such as a genetic algorithm. The proposed technique was tested on a number of problems from the recent literature and found to reduce the classification or data fitting error by over 40% on most datasets. Furthermore, the method appears highly stable as increasing the number of network parameters does not significantly affect its performance.*”

b) As we mention in Abstract the reduction in terms of test error is about 30-40% for the classification datasets and more than 50% for the regression datasets. For this reason we present the average results on a separete row for every experimental table.

c) An additional experiment was executed, where the number of nodes increases from 5 to 20 and the corresponding plot has been added to Experimental Results. The added text reads:

“In order to validate the results an additional experiment was executed for the classification datasets, where the number of nodes increases from 5 to 20 and the results are graphically outlined in Figure [fig:nodesExperiment](#fig_nodesExperiment). From this experiment, one can draw two conclusions: firstly, the proposed technique has a significant advantage over the others to a large extent in terms of average classification error, and secondly, the proposed method is shown to be robust and not significantly dependent on the increase of processing nodes , since 5–10 processing nodes are enough to achieve low classification errors.”

d) Tables 5 and 7 (we have added two additional tables in the manuscript) provide experimental results for the classification datasets, where we measure the average classification error. We also mention in these captions the followin: *“Every number in cells denotes average classification error as measured on the test set”*

e) The classification datasets as well as the regression datasets have been moved to tables for better readability (subsection 3.1)

# REVIEWER2

## 1. COMMENT

The contribution of the current study must be briefly discussed as bullet points in the introduction. And motivation must also be discussed in the manuscript.

## RESPONSE

The following text has been added in the Introduction section:

“*The main contributions of the suggested approach are:*

1. *The first phase procedure seeks to identify a range of values for the network parameters while also reducing the error of the network on the training data set.*
2. *The rules Grammatical Evolution uses in the first phase are simple and can be generalized to any data set for data classification or fitting.*
3. *The determination of the value interval is done in such a way that it is faster and more efficient to train the parameters of the neural network with some optimization method during the second phase of the method.*
4. *After identifying a promising value interval from the first phase, any global optimization method can be used on that value interval to effectively minimize the network training error.* ”

## 2. COMMENT

The introduction section must discuss the technical gaps associated with the current problem.

## RESPONSE

The main issues in the RBF training are mentioned in the introduction as below:

“*A common method of calculating the parameters in these neural networks uses a technique to calculate the centers of the functions φ ( x ) and then the weight vector w &xrarr; is calculated as a solution of a linear system of equations. Typically, the method used to calculate the centers is the well - known k-means method [*[*kmeans*](#LyXCite-kmeans)*]. In many cases this way of estimating the parameters of the neural network leads to over-fitting of the model so that it cannot generalize satisfactorily to unknown data. Furthermore, since there is no range of values for the parameters, there is the possibility that they will take extremely large or extremely small values, with the result that any generalizability of the model is lost.* ”

## 3. COMMENT

Figure 1 should be converted to plain text (poor sharpness, and the text is not a drawing).

## RESPONSE

a) We have converted it to algorithm, since it is more suitable for this kind of information.

b) The caption has been changed to “*The BNF grammar used in the current work, to produce intervals for the RBF parameters. By using this grammar in the first phase of the proposed procedure, the optimal interval of values for the parameters of the neural network will be identified*”

# REVIEWER3

## 1. COMMENT

It is suggested to organize the experimental datasets in a table for better readality.

## RESPONSE

The classification datasets as well as the regression datasets have been moved to tables for better readability (subsection 3.1)

## 2. COMMENT

When comparing with artificial neural networks, why limits the number of hidden nodes to 10? What if we use an optimized set of hyper-parameters for an ANN?

## RESPONSE

An additional experiment was executed, where the number of nodes increases from 5 to 20 and the corresponding plot has been added to Experimental Results. The added text reads:

“In order to validate the results an additional experiment was executed for the classification datasets, where the number of nodes increases from 5 to 20 and the results are graphically outlined in Figure [fig:nodesExperiment](#fig_nodesExperiment). From this experiment, one can draw two conclusions: firstly, the proposed technique has a significant advantage over the others to a large extent in terms of average classification error, and secondly, the proposed method is shown to be robust and not significantly dependent on the increase of processing nodes , since 5–10 processing nodes are enough to achieve low classification errors.”

## 3. COMMENT

What's the advantage of the proposed RBF network against other neural networks?

## RESPONSE

The following text has been added in the Introduction section:

“*The main advantages of RBF networks are:*

1. *They have simpler structure than other machine learning models such as multilayer perceptron neural networks (MLPs)[*[*nnreview*](#LyXCite-nnreview)*], since they have only one processing layer and therefore have faster training techniques and they have faster response times.*
2. *They can approximate any continuous function [*[*rbf\_approx*](#LyXCite-rbf_approx)*].* ”

## 4. COMMENT

Please analyze the computational complexity of different comparative models.

## RESPONSE

We have added a time comparison between the proposed method and the ADAM optimizer for the classification datasets, with the number of processing nodes increases from 5 to 20 and the added text reads:

“*However, the proposed technique consists of two stages and in each of them a genetic algorithm should be executed. This means that it is significantly slower in computing time compared to the rest of the techniques and, of course, it needs more computing resources. This is graphically shown in Figure* [*fig:timeExperiment*](#fig_timeExperiment)*, where the average execution time for the method ADAM and the proposed method is shown for the classification datasets when the number of processing nodes increases from 5 to 20. As expected, the proposed technique requires significantly more time than a traditional neural network training technique such as ADAM, since it consists of two sequential genetic algorithms*”

# REVIEWER4

## 1. COMMENT

In the first place, the abstract is a little thin and doesn't fully show the interesting results from the main paper. I would encourage the authors to extend the abstract more with the key results.

## RESPONSE

The new abstract now reads:

“*Radial basis function networks are widely used in a multitude of applications in various scientific areas in both classification and data fitting problems. These networks deal with the above problems by adjusting their parameters with various optimization techniques. However, an important issue to address is finding a satisfactory interval of values for the network parameters before adjusting these parameters. This paper proposes a two-stage technique, where in the first stage, using Grammatical Evolution, rules are generated to create the optimal value interval of the network parameters. In the second stage of the technique, the parameters of the network are fine-tuned with some robust global optimization method, such as a genetic algorithm. The proposed technique was tested on a number of problems from the recent literature and found to reduce the classification or data fitting error by over 40% on most datasets. Furthermore, the method appears highly stable as increasing the number of network parameters does not significantly affect its performance.*”

## 2. COMMENT

The importance of the design carried out in this manuscript can be explained better than other important studies published in this field. I recommend that the authors review other recently developed works in the section on introduction.

## RESPONSE

The following text has been added at the Introduction:

“*This problem was also tackled by various researchers during the past years such as the work of Agarwal and Bhanot [*[*rbftune*](#LyXCite-rbftune)*] proposed to adapt the RBF parameters, the usage of the ABC algorithm[*[*rbfABC*](#LyXCite-rbfABC)*], the incorporation of the Firefly algorithm[*[*fireflyCancer*](#LyXCite-fireflyCancer)*].*”

## 3. COMMENT

In the introduction section, some recent related work on the topics of advanced machine learning models should be strengthened. recent investigations are recommended for the author’s references, i.e., "10.1016/j.eswa.2021.115982 "; "10.1016/j.ast.2023.108325".

## RESPONSE

The following text has been added in the Introduction section:

“*Furthermore, Gyamfi et al [*[*rbfSeq*](#LyXCite-rbfSeq)*] proposed recently a differential RBF network that incorporates partial differential equations aiming to make the network more robust in the presense of noise data. Also, Li et al [*[*mlHier*](#LyXCite-mlHier)*] proposed a multivariate ensembles-based hierarchical linkage strategy (ME-HL) for system reliability evaluation of aeroengine cooling blades.*”

## 4. COMMENT

In Section 3.2, it is mentioned that the proposed method performs well; however, the performance of the compared methods (i.e. RPROP, ADAM, NEAT) varies greatly. It is recommended that the author provide a solid comparative study by providing a detailed description of these compared methods.

## RESPONSE

a) An additional experiment was executed, where the number of nodes increases from 5 to 20 and the corresponding plot has been added to Experimental Results. The added text reads:

“In order to validate the results an additional experiment was executed for the classification datasets, where the number of nodes increases from 5 to 20 and the results are graphically outlined in Figure [fig:nodesExperiment](#fig_nodesExperiment). From this experiment, one can draw two conclusions: firstly, the proposed technique has a significant advantage over the others to a large extent in terms of average classification error, and secondly, the proposed method is shown to be robust and not significantly dependent on the increase of processing nodes , since 5–10 processing nodes are enough to achieve low classification errors.”

b) We have added a time comparison between the proposed method and the ADAM optimizer for the classification datasets, with the number of processing nodes increases from 5 to 20 and the added text reads:

“*However, the proposed technique consists of two stages and in each of them a genetic algorithm should be executed. This means that it is significantly slower in computing time compared to the rest of the techniques and, of course, it needs more computing resources. This is graphically shown in Figure* [*fig:timeExperiment*](#fig_timeExperiment)*, where the average execution time for the method ADAM and the proposed method is shown for the classification datasets when the number of processing nodes increases from 5 to 20. As expected, the proposed technique requires significantly more time than a traditional neural network training technique such as ADAM, since it consists of two sequential genetic algorithms*”

## 5. COMMENT

It is recommended that the author add pictures comparing the optimization and iteration process of the proposed algorithm with those in the literature to intuitively discover the superiority of the proposed method.

## RESPONSE

There is a flowchart of the suggested algorithm in Figure 3, for better understanding of the optimization process.