



OPEN

Combining multi-objective genetic algorithm and neural network dynamically for the complex optimization problems in physics

Peilin Wang, Kuangkuang Ye, Xuerui Hao & Jike Wang

Neural network (NN) has been tentatively combined into multi-objective genetic algorithms (MOGAs) to solve the optimization problems in physics. However, the computationally complex physical evaluations and limited computing resources always cause the unsatisfied size of training set, which further results in the combined algorithms handling strict constraints ineffectively. Here, the dynamically used NN-based MOGA (DNMOGA) is proposed for the first time, which includes dynamically redistributing the number of evaluated individuals to different operators and some other improvements. Radio frequency cavity is designed by this algorithm as an example, in which four objectives and an equality constraint (a sort of strict constraint) are considered simultaneously. Comparing with the baseline algorithms, both the number and competitiveness of the final feasible individuals of DNMOGA are considerably improved. In general, DNMOGA is instructive for dealing with the complex situations of strict constraints and preference in multi-objective optimization problems in physics.

Multi-objective genetic algorithms (MOGAs) such as NSGA-II¹, MOEA/D², SPEA³ have shown good performance in many engineering optimization problems. Being inspired by the evolutionary theory of “survival of the fittest”, the competitive individuals can be obtained through operators of selection, mutation, and crossover by iteration. These individuals that cannot outperform each other on all objectives create a set, the so-called nondominated front. In terms of the physical optimization problems in which the evaluations are always computationally complex, their population sizes in MOGAs are generally small because of the limited computing resources. When the parameter space formed by the decision variables is very large and local optima exist in these questions, these algorithms tend to converge on local optima rather than global optima.

The field of accelerator can be taken as an example. There are many complex optimization problems that do not have the known optimal solution sets (the Pareto front) in accelerator fields, and good performance has been preliminarily obtained with MOGAs in some of these problems, such as the optimization of lattice⁴, Free Electron Laser^{5,6} and other accelerator facilities⁷. Designing the shape of radio frequency (RF) cavity is a sort of important problem as well, in which several objectives such as geometric shunt impedance (R/Q), Q factor and shunt impedance (R_a) are optimized simultaneously by tuning the geometric parameters of the cavity. At the same time, an equality constraint⁸ has to be considered in this design, which means the frequency of fundamental mode (f_{FM}) of cavity must be equal to a given target frequency, otherwise the cavity cannot be used even with great performance. These individuals that satisfy constraint are called feasible individuals, while others are infeasible. Although MOGAs have obtained some competitive individuals in RF cavity design^{9–11}, these works still rely on the professional knowledge of manual process to set the small and proper decision space. As a result, the manual optimization process as the most popular method in engineering^{12,13} cannot be replaced by MOGAs yet.

Neural network (NN) has been tentatively combined into MOGAs to increase the population size and speed up convergence^{14–16}. In these attempts, the common idea (named NBMOGA) is estimating all the individuals with NN instead of evaluating them directly, while the difference lies in the composition of training set and the number of times the NN is trained. In a specific algorithm¹⁵, after executing standard NSGA-II for several generations, NN is trained to estimate more individuals of which some are selected to be further evaluated. This combination performs good in the optimization of the dynamic aperture area and the Touschek lifetime. When the former objective remains unchanged, the latter one increases about 10% compared with the standard MOGA in similar time. It is clearly seen that NBMOGA do well in the convergence speed, but the shortcoming of entirely

The Institute for Advanced Studies, Wuhan University, Wuhan 430072, China. ✉ email: jike.wang@whu.edu.cn

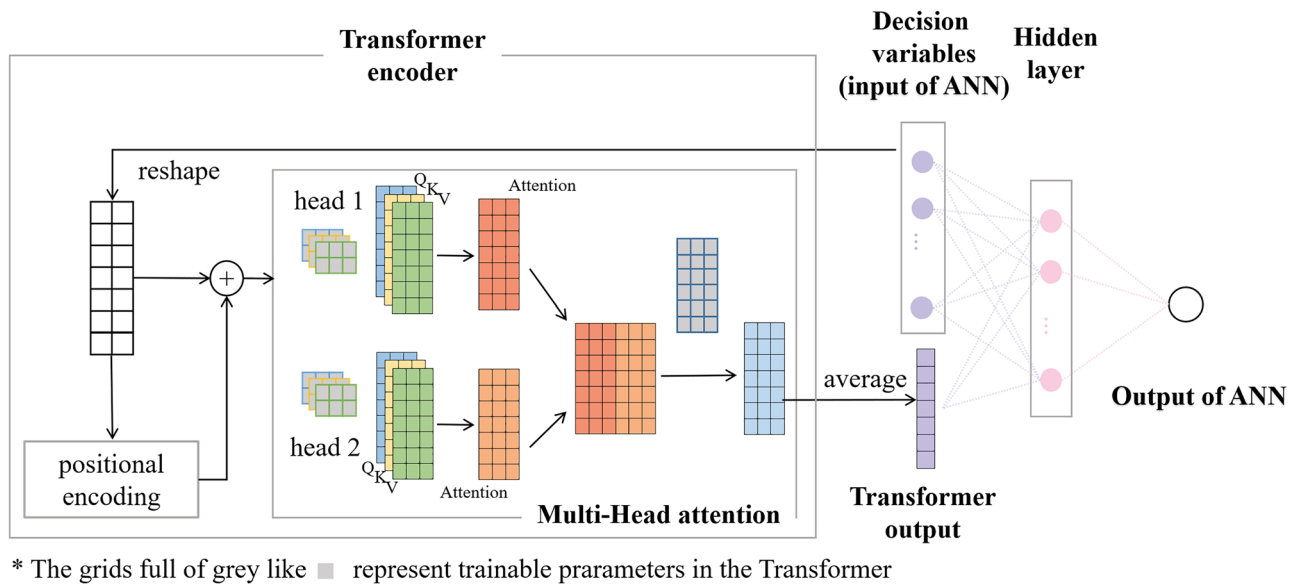


Figure 1. The model of NN combining ANN and Transformer.

Training sets and model		Number of trainable parameters	R ²				
			f _{FM}	R/Q _{FM}	R _{aFM}	Q _{FM}	f _{HOM}
3000 samples	Transformer with ANN	265	0.987	0.985	0.989	0.988	0.984
	2-layers ANN	76	0.923	0.954	0.957	0.969	0.942

Table 1. The values of R² in different NN models and indicators are compared.

relying on the estimated indicators to select parents since the training of NN begins^{14–16} still exists in them. Once the strict constraint is considered in the optimization, the feasible individuals are easily estimated as infeasible because of the small size of training set, and this is a major challenge for these algorithms.

To deal with this problem, a penalty operation that can be progressively stricter over generations is executed in fitness function to fulfill constraints gradually. Meanwhile, NN is included in an operator. This operator not only produces several individuals to be further evaluated, just like the operators of mutation and crossover do, but also screens in a great number of estimated individuals internally. The performance of these operators is different when the penalty changes, so the number of individuals that come from these operators to be further evaluated is dynamically redistributed. Namely, this algorithm is called the dynamically used NN-based MOGA (DNMOGA). In addition, accessibility algorithm is proposed as a new idea to deal with preference in NSGA-II, which doesn't depend on extra reference points that are set manually in other algorithms^{17,18} to lead the non-dominated front to approach them.

The shape of spherically shaped (SS) normal conducting cavity, a kind of RF cavity used in PEP II¹⁹, is optimized to prove the advantage of DNMOGA. Differences among various algorithms are compared, while two NN models are combined into DNMOGA respectively to illustrate the relationship of the performance of this optimizer and the accuracy of NN. Besides, some other details about DNMOGA are also discussed below. As a result, DNMOGA shows the potential to completely replace manual procession in this question, which also announces its ability to solve other optimization problems in physics that have the similar features with this design.

Results

NN models. One of the two NN models mentioned above is a simple artificial neural network (ANN) that has 5 neurons within one hidden layer^{20–23}. The other one, which is shown in Fig. 1, combines ANN and Transformer^{24,25} and has better accuracy. Note that neither of these models can accurately estimate the indicator of equality constraint in these experiments. The accuracies of these models are expressed with R², a criterion unrelated to MOGAs. This criterion is defined as

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \tag{1}$$

where y_i is the label, \hat{y}_i is the estimated value and \bar{y} is the average of y_i . The indicator estimated by NN is more precise when the value of R² is closer to 1. The values of R² of these models in 5 indicators are listed in Table 1. The 5 indicators are f_{FM} , R/Q of the FM (R/Q_{FM}), R_a of the FM (R_{aFM}), Q factor of the FM (Q_{FM}), frequency of

the higher order modes (f_{HOM}), in turn. From this table, the Transformer combined with ANN estimates more accurately in all indicators. The specific illustration of these models and the idea of designing these models are described in Supplementary materials.

Fitness function for dealing with constraint. Although feasible individuals are the favorites to decider, infeasible individuals can always provide the hidden information about global optima and the lower constraint violation (cf. Eq. 9). So, how to balance the numbers of feasible and infeasible individuals in parents is an important problem. Generally, there are four categories²⁶ to deal with it: adding a penalty to fitness function^{27,28}, judging constraint condition before adding the penalty^{29,30}, proposing novel selection strategy^{31–33} and regarding constraint as objective^{34,35}. The popular ways in engineering are the first two categories, while the specific operation of penalty in them is punishing infeasible individuals by increasing their fitness. Unfortunately, the uncontrollable numbers of feasible and infeasible parents in these algorithms can be dramatically different when the range of constraint changes or the number of generations is different, which results in a failure of balance.

To solve the mentioned problem, the feasible and infeasible parents are put into two independent groups as a detail of setting parents in this work, while a fitness function with a penalty operation that can make individuals fulfill constraints gradually is used. In the mathematical model shown below, $\vec{x} = (x_1, x_2, \dots, x_l) \in S$ is a l -dimensional decision vector, in which $S \subset \mathbb{R}^l$ is the decision space. $\vec{F}(\vec{x})$ is a fitness vector consisting of the fitness values of m objectives, and $F_i(\vec{x})$ ($i \in \{1, 2, \dots, m\}$) is the fitness function. The $g_j(\vec{x})$ and $h_j(\vec{x})$ are inequality and equality constraints respectively, n represents the total number of inequality and equality constraints, and x_q^{\min} and x_q^{\max} are the bound constraints of x_q ($q \in \{1, 2, \dots, l\}$).

$$\begin{aligned} \min \vec{F}(\vec{x}) &= (F_1(\vec{x}), F_2(\vec{x}), \dots, F_m(\vec{x}))^T \\ \text{s.t.} \begin{cases} g_j(\vec{x}) \leq 0, j = 1, \dots, p \\ h_j(\vec{x}) = 0, j = p + 1, \dots, n \\ x_q^{\min} \leq x_q \leq x_q^{\max}, q = 1, 2, \dots, l \end{cases} \end{aligned} \tag{2}$$

The general way to penalize the fitness function is shown here^{27,30}:

$$F_i(\vec{x}) = 1 - f_i(\vec{x}) + \sum_{j=1}^n r'_j c_j(\vec{x}) \tag{3}$$

$$c_j(\vec{x}) = \begin{cases} \max(0, g_j(\vec{x})), j = 1, \dots, p \\ \max(0, |h_j(\vec{x})| - \delta), j = p + 1, \dots, n \end{cases} \tag{4}$$

where $f_i(\vec{x})$ is the value of the i th objective, and δ is a very small positive tolerance value to relax the equality constraints. r_j as the penalty coefficient used to be adjusted in different problems to keep the balance^{28,36}. Based on these experiences, the r_j is redesigned to match the operator with NN, and the whole expression of penalty operation is described below.

The $F_i(\vec{x})$ used in Eq. (2) is set as

$$F_i(\vec{x}) = 1 - e_i(\vec{x}) + d(\vec{x}) \tag{5}$$

$$e_i(\vec{x}) = \begin{cases} f_i(\vec{x}) \cdot r_j, CV(\vec{x}) > 0 \\ f_i(\vec{x}), CV(\vec{x}) = 0 \end{cases} \tag{6}$$

$$d(\vec{x}) = \begin{cases} CV(\vec{x}) \cdot (1 - r_j), CV(\vec{x}) > 0 \\ 0, CV(\vec{x}) = 0 \end{cases} \tag{7}$$

where $CV(\vec{x})$ described in Eq. (9) is the overall value of constraint violation, and r_j is a penalty coefficient shown in Eq. (8).

$$r_j = \left(\frac{1}{1 + k_j \cdot \text{const} \frac{\text{pre}}{\text{total}}} \right)^4 \tag{8}$$

In Eq. (8), pre represents the number of generations (the evaluated generations, Fig. 3) at present, and $total$ represents the total number of evaluated generations. k_j is an adjustable value and written as k later, because there is only one constraint in the specific problem. $const$ is a constant to keep a suitable range of speed to converge, and it is set as $\frac{1}{e^k}$. In this penalty operation, $F_i(\vec{x})$ will be more and more large when $CV(\vec{x})$ grows. $r_j(pre)$ has a good mathematical property, because it is a continuous and derivable convex function when pre is regarded as an independent variable.

The $CV(\vec{x})$ is calculated as

$$CV(\vec{x}) = \sum_{j=1}^n r_j \cdot c_j(\vec{x}) \tag{9}$$

where the operation of normalization can be calculated according to Eq. 10.

Note that the values of $F_i(\vec{x})$, $CV(\vec{x})$ and $f_i(\vec{x})$ must be normalized. In each generation, the values that come from the same function create a set, and the method of normalization in each set is expressed below:

$$\text{norm}(Y) = \frac{Y - \min(Y)}{\max(Y) - \min(Y)}, \quad (10)$$

where the Y is an element in set, and $\max(Y)$, $\min(Y)$ is the maximum and minimum in this set respectively.

Dynamically redistributing the numbers of individuals to operators. The way to redistribute is defined as

$$N_{v+1} = T \times \frac{PN_v}{PT_v}. \quad (11)$$

In Eq. 11, T is fixed as the total number of evaluated individuals in each generation, while PT_v represents how many individuals are selected as parents totally in the v th generation. For each operator, N_{v+1} represents the redistributed number of individuals in the $(v+1)$ th generation, and PN_v represents how many individuals from this operator are selected as parents in the v th generation.

The accessibility algorithm. Sometimes the decider might rank the degrees of importance to all objectives, which means preference exists. The common idea to deal with preference is introducing manual reference points^{17,18,37} and calculating distances between individuals and these points, then leading the nondominated front to approach them. But these methods only concentrate on the narrow regions around the reference points where too much hidden information about global optima is lost, and consequently converge to local optima in complex problems. In DNMOGA, accessibility which is frequently used in traffic problems³⁸ is introduced instead of original crowding-distance, and it can be expressed as

$${}_1A = \sum_{o=1}^M {}_1A_o = \sum_{o=1}^M \frac{S_o}{T_{1,o}^Y} = \sum_{o=1}^M \frac{\vec{H} \cdot \vec{F}_o + 1}{D_{1,o}}. \quad (12)$$

The first three terms of Eq. 12 is expressed in³⁸ as well, in which the M represents the number of areas, ${}_1A$ is the accessibility of the first area, and ${}_1A_o$ is the accessibility from the first area to the o th. At the same time, S_o represents the vitality in the o th area, and it can be described as population, gross domestic product, etc. Besides, $T_{1,o}$ represents the travel time between these areas, while Y is an exponent to describe the effect of travel time. In DNMOGA, $T_{1,o}^Y$ is replaced by $D_{1,o}$ to represent the Euclidean distance of objectives between individuals. Then, accessibility can be used to deal with preference by relating S_o to every objective as a number larger than 1. In the last term of Eq. 12, \vec{H} represents the preference vector with a series of values, and each of them is corresponded to a specific objective. The larger value in \vec{H} represents that the corresponding objective is more important. In addition, the \vec{F}_o represents the fitness vector for these objectives.

There are two sets in accessibility algorithm, one is the pending set in which all the individuals of nondominated front are initially included as the pending parents of the next generation, and the other one is the selected set to collect the selected individuals. After normalizing the objective value and putting the boundary solutions to the selected set from the pending one, the accessibility of each pending individual to all the selected individuals are calculated, then the most inaccessible one is taken to the selected set. This procession is repeated until the size of selected set is satisfied. The picture below (Fig. 2) shows the advantage of accessibility. The fake code is released in Supplementary materials.

The complete process of DNMOGA. The flow chart of DNMOGA is shown in Fig. 3. Before describing the complete process of DNMOGA, it is necessary to detail the operator including NN first. In this operator (the orange blocks in Fig. 3), the crossover, mutation and Latin hypercube sampling (LHS)³⁹ are used to produce astronomical individuals together, then NN is used to estimate the indicators of these individuals. After executing the fast nondominated sort algorithm, the nondominated front is generated as the parents of next generation. This series of operations constitutes the estimated generation, which is carried out five times continuously. Then the accessibility algorithm is executed among the nondominated front of the fifth estimated generation to pick up a certain number of individuals that will be further evaluated.

The first step of the complete process in DNMOGA is evaluating a certain number of individuals that are produced by LHS as the training set of NN and the first generation of NSGA-II, after which the parents are selected and the main loop begins. In this cycle, crossover and mutation are used again as two of the operators (the green and yellow blocks in Fig. 3), and they produce the individuals to be directly put in the actual evaluator together with the operator with NN. Then, all the evaluated individuals in this generation are mixed with these earlier evaluated, after that the fitness function is executed. Once the fast nondominated sort algorithm and accessibility algorithm are executed in feasible and infeasible individuals respectively, the two groups of parents for next generation can be obtained. As the last step, the numbers of individuals generated by these operators in the next generation are dynamically redistributed. This generation produces new individuals through actual evaluator, and it is naturally called the evaluated generation. The fake code of the complete process is released in Supplementary materials.

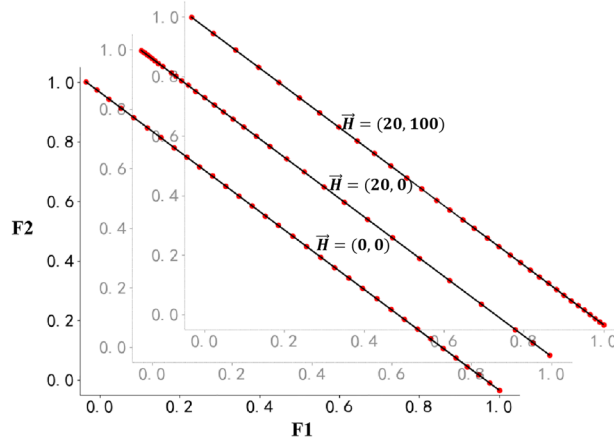


Figure 2. The distribution of selected individuals which is obtained by using accessibility algorithm. 1000 nondominated individuals are set and 33 individuals are selected. There are two objectives: F1 and F2, and both should be minimized. When the \vec{H} is set as (0, 0), it means no preference in F1 and F2, and the distribution of selected individuals is very uniform. When the \vec{H} is (20, 0), the individual with lower F1 score prefers to be taken. When the \vec{H} is (20, 100), although preference value of F1 is 20, but F2 has higher preference value, and more individuals good at F2 are taken.

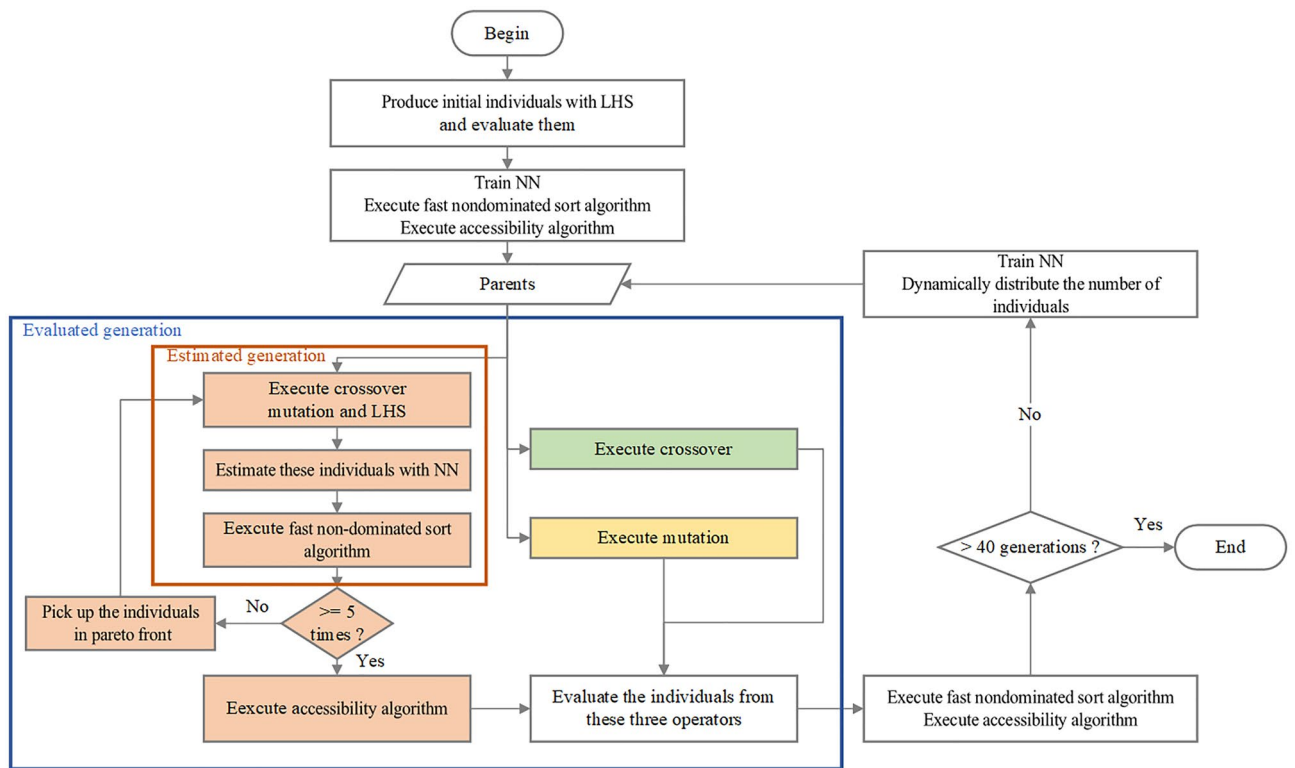


Figure 3. The flow chart of DNMOGA. The green, yellow, and orange blocks represent the three operators, and they are named as crossover, mutation, and the operator with NN, respectively.

In this study, some specific details can also be helpful to obtain the competitive individuals. To begin with, as an assisting method to satisfy constraint, standard mutation operation is replaced by adjusting the f_{FM} to produce more feasible individuals, and the specific operation of adjusting frequency is shown below:

$$Req' = \frac{f_{FM} - 499.65}{2.55} + Req. \tag{13}$$

It is assumed that f_{FM} is only related to Req , and because of the linear relationship in physics between Req and f_{FM} , an appropriate coefficient is empirically set as 2.55 [MHz/mm]. The Req' in Eq. 13 is the new decision

variables obtained by this operation. The comparison between this method and the standard mutation will be described below. For the optimization problems in physics, tuning frequency might not suit for all scenarios, but the relationship between the objectives and decision variables can be expressed in a formula, and the positive or negative correlation can be found locally as the basis to design the assisting method even if it sometimes is hard for the objectives to be calculated correctly according to this relationship. The second detail is about the two groups of parents. To increase the variety of parents, the number of parents in each group is the greater value between 90% of the size of their nondominated front and a constant. Note that all the values mentioned below about the parents are this constant, and the operation of two groups of parents is not used in selecting estimated individuals because of the dissatisfied accuracy.

Optimization of the SS cavity. The shape and the geometric parameters of the SS cavity are shown in Fig. 5c, in which the tube parameters Rt and Lt that we do not care about are fixed as 200 and 16 mm respectively.

In order to obtain the shapes with performance as good as possible, larger space should be explored, which means the limits among the geometric parameters must be considered. For example, the RO_l , RO_r should be smaller than Req , while the sum of Rt , $R3_l$ and nose should be smaller than the Req as well. The way to deal with these limits is executing nonlinear transformations to relate the geometric parameters with independent variables, and the variables should have the same degrees of freedom to the parameters. As a result, 13 independent variables are set as decision variables in DNMOGA. The specific procession is discussed in Supplementary materials, and the range of Leq is in the range of 11 and 630 mm, while Req is between 116 and 216 mm.

Besides the indicators mentioned in the introduction, there are also many other indicators in the optimization problems of RF cavity, such as the Q factor, the normalized peak electric field on the cavity surface, and so on. Principally, Q factor and R_a of the FM should be maximized, but according to the previous papers^{9,10,40} and mathematical relationship among R/Q , R_a and Q factor (as shown in Eq. 14), R_a and R/Q are usually considered as the objectives. In addition, the normalized peak electric field on the surface of the optimized normal conducting cavity can always satisfy the decider's requirement, so it is meaningless to be seen as an objective. Moreover, the indicators in HOM are important for the cavity of the 4th generation synchrotron radiation sources, in which the R/Q should be minimized, and the frequency should be maximized. As a result, four indicators shown in Eq. (15) are selected as the objectives of the optimization. Note that the indicators with * should be minimized, so they are transferred to satisfy the mathematical model mentioned above. This is shown in Eq. 15 as well. Besides, HOM is the first higher order mode we meet during the calculation. The δ in equality constraint is 0.05.

$$R/Q = \frac{R_a}{Q} \quad (14)$$

$$\begin{aligned} \min \vec{F}(\vec{x}) = & \begin{pmatrix} F_1(R/Q_{FM}(\vec{x})), \\ F_2(R_a_{FM}(\vec{x})), \\ F_3(f_{HOM}(\vec{x}) - f_{FM}(\vec{x})), \\ F_4(*R/Q_{HOM}(\vec{x})) \end{pmatrix}, \\ \text{s.t. } & f_{FM} - 499.65 = 0, \\ & *A = \max(A) - A. \end{aligned} \quad (15)$$

When the range of FM frequency in 1000 initial individuals is between 294.44 MHz and 954.05 MHz, which is 6600 times larger than the constraint, approximately 300 feasible individuals are found after evaluating 2000 individuals within 40 evaluated generations. This algorithm takes about 34 h totally when two CST processions work simultaneously.

In order to finish following discussion, a suitable k (which is mentioned in Eq. 8) is set as 0.1. Fit sizes of two groups of parents are set as well, which are 100 and 50 to feasible and infeasible groups respectively. The trends of changing k and the sizes of two groups of parents are discussed in Supplementary materials. These parameters in DNMOGA mainly influence the speed of convergence, while the performance of nondominated individuals would be affected tremendously only with harsh settings.

The results of experiments are shown in Fig. 4. The first comparison is about the two groups of parents and the general way of setting parents. In Fig. 4a,b, both the constants (mentioned above) of setting the total number of parents are the same as 150. The performance of nondominated individuals produced by the two groups of parents is approximately like that with the general method, but the former one creates a larger size of nondominated front.

Figure 4c illustrates the dynamic redistribution of evaluated individuals. As the algorithm goes, the penalty is stricter so that the individuals from NN performs worse in fitness values. As a result, less individuals from NN are picked up, and this is the reason why DNMOGA performs better than others when the accurate of NN is dissatisfied.

The results of DNMOGA with two different preferences are shown as well, and their \vec{H} vectors (described in Eq. 12) are set as (5, 1, 1, 0) (Fig. 4d) and (10, 1, 1, 0) (Fig. 4e) respectively. The four values in \vec{H} correspond with the four objectives ranging from F1 to F4 in turn. When the preference of F1 is improving gradually, more individuals that perform good in this objective are obtained. In Figs. 4f,i, individuals from Fig. 4a,d,e are mixed, and all these individuals are ranked by F1 and F4 respectively. From these figures, individuals generated with (10, 1, 1, 0) perform well in F1 and poorly in F4.

By comparing the results of DNMOGA with two different NN models, it is concluded that the accuracy of NN only influences the convergence speed. When the total number of evaluated generations is 40, the performance

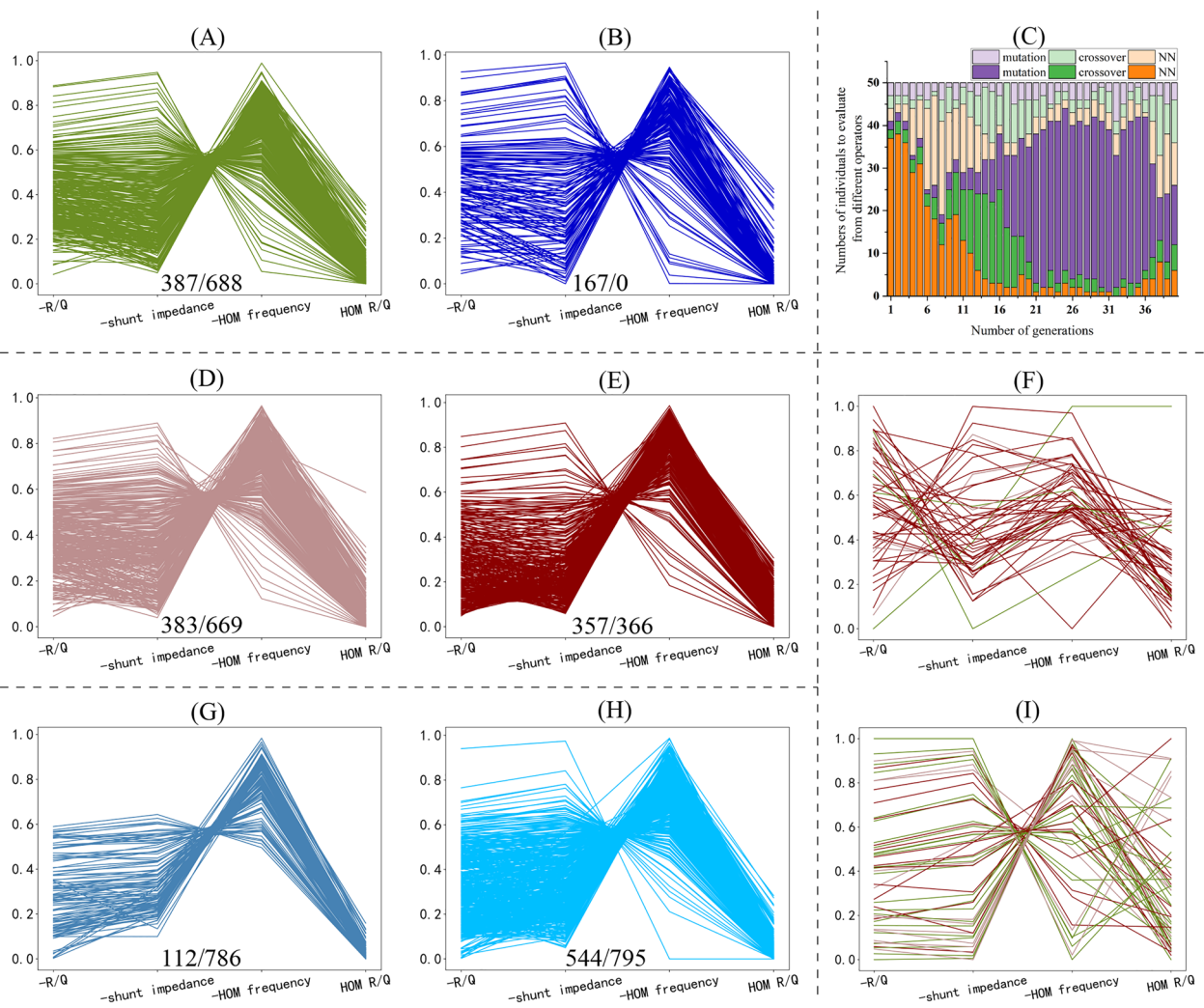


Figure 4. The results of DNMOGA. **(a)** Result of DNMOGA ($k=0.1$, sizes of the two groups of parents = (100,50), initial size = 1000, generation = 40, $\bar{H} = (0,0,0,0)$, using high accurate NN). **(b)** Result of DNMOGA ($k=0.1$, size of parents = 150, initial size = 1000, generation = 40, $\bar{H} = (0,0,0,0)$, using high accurate NN). **(c)** Redistribution quantities of evaluated individuals in Fig. 4a. The darker bars represent the numbers of individuals which are selected as parents of the next generation, and the lighter bars represent the numbers of individuals which are not selected. **(d)** Result of DNMOGA ($k=0.1$, sizes of the two groups of parents = (100, 50), initial size = 1000, generation = 40, $\bar{H} = (5,1,1,0)$, using high accurate NN). **(e)** Result of DNMOGA ($k=0.1$, sizes of the two groups of parents = (100,50), initial size = 1000, generation = 40, $\bar{H} = (10,1,1,0)$, using high accurate NN). **(f)** 50 ranked and selected individuals that come from Fig. 4a,d,e according to the R/Q of the FM. **(g)** Result of DNMOGA ($k=0.1$, sizes of the two groups of parents = (100, 50), initial size = 1000, generation = 40, $\bar{H} = (0,0,0,0)$, using low accurate NN). **(h)** Result of DNMOGA ($k=0.1$, sizes of the two groups of parents = (100,50), initial size = 1000, generation = 60, $\bar{H} = (0,0,0,0)$, using low accurate NN). **(i)** 50 ranked and selected individuals that come from Fig. 4a,d,e according to the R/Q of the HOM. Note that the Fig. 4a,b,d,e,g,h are the parallel coordinate plots of the feasible nondominated fronts in the last generation, in which every line represents an individual. The first objective in these subfigures is the R/Q of the FM, and the second is the Ra of the FM. The two values at the bottom of subfigures represent the numbers of feasible and infeasible nondominated individuals of the last generation respectively.

of DNMOGA with ANN is not good (Fig. 4g). But if continuing to evaluate 20 generations (Fig. 4h), it is nearly the same as the result of 40 total generations with a better NN model (Fig. 4a).

The results of different algorithms (NSGA-II, DNMOGA and the NBMOGA in¹⁵) are shown in Fig. 5a,b, while the method proposed by²⁹ is combined in NSGA-II and NBMOGA to deal with constraints. All these experiments evaluated 50 individuals per generation, and the total number of individuals is 3000. The initial population and the numbers of the total generations in NSGA-II are the same as DNMOGA, while these values are 50 and 500 for NBMOGA to keep the same number of estimated generations as DNMOGA (the training of NN in NBMOGA begins at the 10th generation). From the result of Fig. 5a,b, adjusting frequency is a useful assisting method to obtain more feasible individuals, while the gaps in the distributions and sizes of nondominated front between the two algorithms and DNMOGA is clear.

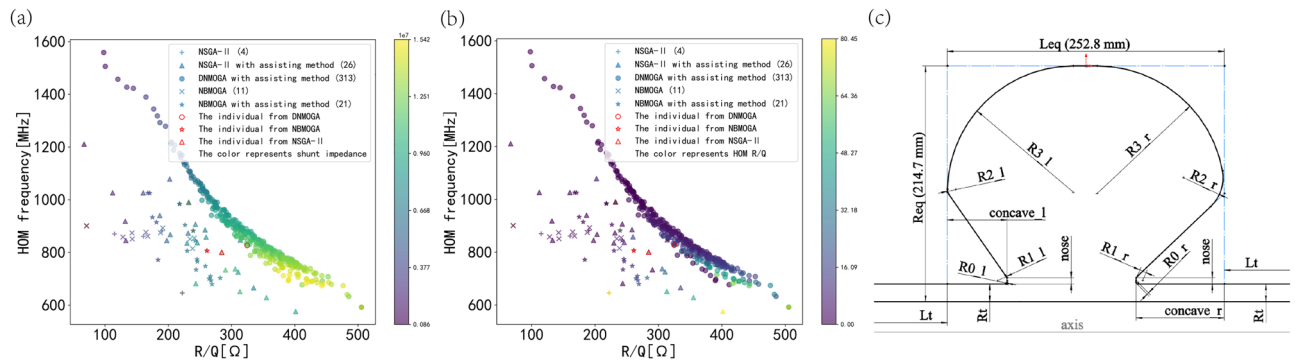


Figure 5. Comparing the results of different algorithms. (a) The scatter graphs of feasible and nondominated fronts from different algorithms. The colored z-axis represents the $R_{a,FM}$. The abscissa and ordinate represent R/Q of the FM and HOM frequency respectively, and the number in brackets represents how many feasible and nondominated individuals exist in each algorithm. The red shapes represent the individuals picked up from these algorithms. (b) The scatter graphs of feasible and nondominated fronts from different algorithms, in which the colored z-axis represents the R/Q_{HOM} . (c) The geometric parameters of the SS cavity and the shape of an individual picked up from Fig. 4a.

Then the individuals that have the similar frequency of the HOM are picked up from different algorithms, and their locations in the nondominated fronts are signaled on Fig. 5a,b. The indicators of these individuals are shown in Table 2, in which the advantage of DNMOGA can be discovered. The R/Q_{FM} is improved by about 24% and 14% compared with using the NBMOGA and NSGA-II, while the $R_{a,FM}$ is increased by approximately 55 and 22% respectively. Besides, only the R/Q_{HOM} that comes from the individual of DNMOGA tends to zero, which is beneficial for further analysis of HOM. The geometric parameters of the individual from DNMOGA are shown in Fig. 5c.

Some benchmark optimization problems, such as CEC2009, DTLZ, and CMOP, are also used to validate the performance of DNMOGA, and the results have been added in the Supplementary materials.

Discussion

Various machine learning models have potential to be combined into MOGAs to solve multi-objective optimization problems, but the way to combine is the key factor to influence the performance. In this paper, DNMOGA in which NN is dynamically used in a novel way of combination, is proposed and demonstrated. It's good at dealing with the optimizing problem in physics, especially the complex questions with constraints and preference. It is easy to find the advantage of DNMOGA compared to NBMOGA and NSGA-II through the design of RF cavity. At the same time, using assisting methods is functional to make individuals more feasible in problems with strict constraints, while these methods are easier to operate in physically meaningful problems.

Not only can all kinds of RF cavity optimizations be well handled, such as multi-cell cavity¹¹ and heavy ion cavities¹¹, but also other accelerator optimization are principally suitable, like free electron laser^{5,6}, nonlinear beam dynamics⁴ and so on. If we look at optimization designs in other physics fields, such as radio apparatus⁴² and structural components of materials⁴³, more questions would be solved. In aerospace, MOGAs have been used to optimize 3D Wing-Shape, but response surface methodology is executed to meet the limited computing resources⁴⁴; if suitable estimator is combined into MOGAs, the limit of calculation resource might be solved. All in all, it is obvious to see that combining machine learning and MOGAs have great potential to be dug out, the DNMOGA is one of the excellent methodology output.

Methods

Actual evaluator. The actual evaluator in these experiments is CST Studio Suite⁴⁵, a software that can parametrically produce 3D models and obtain the electromagnetic field of the cavity by finite element analysis. The powerful post-processing functions in it can calculate various indicators. In this experiment, the mesh of the finite element analysis is set as 20 cells per wavelength.

Calculation facility. The calculation facility used is a workstation with Xeon W-2265 CPU and 64 GB memory, and it takes about 45 s for this workstation to evaluate a cavity.

Algorithm	R/Q _{FM} [Ω]	R _{FM} [MΩ]	f _{HOM} [MHz]	R/Q _{HOM} [Ω]
DNMOGA	324.72	12.15	826.75	8.28E-7
NBMOGA	261.09	7.82	806.46	9.46
NSGA-II	284.54	9.93	800.56	18.58

Table 2. The indicators of individuals picked up from the feasible and nondominated fronts of last generation.

Data availability

The raw data generated during the current study is available in the github repository, <https://github.com/PeilinWangWHU/Combining-multi-objective-genetic-algorithm-and-neural-network-dynamically>.

Received: 30 July 2022; Accepted: 3 January 2023

Published online: 17 January 2023

References

- Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197. <https://doi.org/10.1109/4235.996017> (2002).
- Zhang, Q. & Li, H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evol. Comput.* **11**, 712–731. <https://doi.org/10.1109/TEVC.2007.892759> (2007).
- Zitzler, E. & Thiele, L. Multiobjective evolutionary algorithms: A comparative case study and the Strength Pareto approach. *IEEE Trans. Evol. Comput.* **3**, 257–271. <https://doi.org/10.1109/4235.797969> (1999).
- Gao, W., Wang, L., Li, W. & Beams, S. Simultaneous optimization of beam emittance and dynamic aperture for electron storage ring using genetic algorithm. *Phys. Rev. Spec. Top.-Accel.* **14**, 094001. <https://doi.org/10.1103/PhysRevSTAB.14.094001> (2011).
- Yan, J. & Deng, H. Generation of large-bandwidth x-ray free electron laser with evolutionary many-objective optimization algorithm. *Phys. Rev. Accel. Beams* **22**, 020703. <https://doi.org/10.1103/PhysRevAccelBeams.22.020703> (2019).
- Wu, J. *et al.* Multi-dimensional optimization of a terawatt seeded tapered free electron laser with a multi-objective genetic algorithm. *Nucl. Instrum. Methods Phys. Res. Sect. A-Accel. Spectrom. Detect. Assoc. Equip.* **846**, 56–63. <https://doi.org/10.1016/j.nima.2016.11.035> (2017).
- Hofler, A. *et al.* Innovative applications of genetic algorithms to problems in accelerator physics. *Phys. Rev. Spec. Top.-Accel.* **16**, 010101. <https://doi.org/10.1103/PhysRevSTAB.16.010101> (2013).
- Courant, R. & Hilbert, D. Methods of mathematical physics. *Bull. Am. Math. Soc.* **60**, 578–579 (1954).
- Kranjčević, M., Adelman, A., Arbenz, P., Citterio, A. & Stingelin, L. Multi-objective shape optimization of radio frequency cavities using an evolutionary algorithm. *Nucl. Instrum. Methods Phys. Res., Sect. A* **920**, 106–114. <https://doi.org/10.1016/j.nima.2018.12.066> (2019).
- Kranjčević, M., Zadeh, S. G., Adelman, A., Arbenz, P. & Van Rienen, U. Constrained multiobjective shape optimization of superconducting rf cavities considering robustness against geometric perturbations. *Phys. Rev. Accel. Beams* **22**, 122001. <https://doi.org/10.1103/PhysRevAccelBeams.22.122001> (2019).
- Luo, T. *et al.* RF design of APEX2 two-cell continuous-wave normal conducting photoelectron gun cavity based on multi-objective genetic algorithm. *Nucl. Instrum. Methods Phys. Res. Sect. A* **940**, 12–18. <https://doi.org/10.1016/j.nima.2019.05.079> (2019).
- Li, Z.-Q. & Zhang, C. Study of heavily damped SC RF cavity. *Chin. Phys. C* **27**, 919–925 (2003).
- Zheng, H.-J., Gao, J. & Liu, Z.-C. Cavity and HOM coupler design for CEPC. *Chin. Phys. C* **40**, 057001. <https://doi.org/10.1088/1674-1137/40/5/057001> (2016).
- Kranjčević, M., Riemann, B., Adelman, A. & Streun, A. Multiobjective optimization of the dynamic aperture using surrogate models based on artificial neural networks. *Phys. Rev. Accel. Beams* <https://doi.org/10.1103/PhysRevAccelBeams.24.014601> (2021).
- Wan, J., Chu, P. & Jiao, Y. Neural network-based multiobjective optimization algorithm for nonlinear beam dynamics. *Phys. Rev. Accel. Beams* **23**, 081601. <https://doi.org/10.1103/PhysRevAccelBeams.23.081601> (2020).
- Edelen, A. *et al.* Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems. *Phys. Rev. Accel. Beams* <https://doi.org/10.1103/PhysRevAccelBeams.23.044601> (2020).
- Ben Said, L., Bechikh, S. & Ghedira, K. The r-Dominance: A new dominance relation for interactive evolutionary multicriteria decision making. *IEEE Trans. Evol. Comput.* **14**, 801–818. <https://doi.org/10.1109/tevc.2010.2041060> (2010).
- Molina, J., Santana, L. V., Hernandez-Diaz, A. G., Coello, C. A. C. & Caballero, R. g-dominance: Reference point based dominance for multiobjective metaheuristics. *Eur. J. Oper. Res.* **197**, 685–692. <https://doi.org/10.1016/j.ejor.2008.07.015> (2009).
- Marhauser, F., Wehreter, E., Dykes, D. & McIntosh, P. in *PACS2001. Proceedings of the 2001 Particle Accelerator Conference (Cat. No. 01CH37268)*. 846–848 (IEEE).
- Lippmann, R. An introduction to computing with neural nets. *IEEE ASSP Mag.* **4**, 4–22. <https://doi.org/10.1109/MASSP.1987.1165576> (1987).
- Widrow, B. & Lehr, M. A. 30 years of adaptive neural networks: Perceptron, madaline, and backpropagation. *Proc. IEEE* **78**, 1415–1442. <https://doi.org/10.1109/5.58323> (1990).
- Girosi, F. & Poggio, T. Networks and the best approximation property. *Biol. Cybern.* **63**, 169–176. <https://doi.org/10.1007/BF00195855> (1990).
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536. <https://doi.org/10.1038/323533a0> (1986).
- Khan, S. *et al.* Transformers in vision: A survey. *ACM Comput. Surv.* <https://doi.org/10.1145/3505244> (2021).
- Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems*, Vol. 30 (eds. Guyon, I. *et al.*) (Curran Associates, Inc., 2017).
- Wang, Y., Li, J. P., Xue, X. H. & Wang, B. C. Utilizing the correlation between constraints and objective function for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.* **24**, 29–43. <https://doi.org/10.1109/tevc.2019.2904900> (2020).
- Tessema, B. & Yen, G. G. An adaptive penalty formulation for constrained evolutionary optimization. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **39**, 565–578. <https://doi.org/10.1109/tsmca.2009.2013333> (2009).
- Liu, J. J., Teo, K. L., Wang, X. Y. & Wu, C. Z. An exact penalty function-based differential search algorithm for constrained global optimization. *Soft. Comput.* **20**, 1305–1313. <https://doi.org/10.1007/s00500-015-1588-6> (2016).
- Deb, K. An efficient constraint handling method for genetic algorithms. *Comput. Meth. Appl. Mech. Eng.* **186**, 311–338. [https://doi.org/10.1016/S0045-7825\(99\)00389-8](https://doi.org/10.1016/S0045-7825(99)00389-8) (2000).

30. Runarsson, T. P. & Yao, X. Stochastic ranking for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.* **4**, 284–294. <https://doi.org/10.1109/4235.873238> (2000).
31. Cai, Z. X. & Wang, Y. A multiobjective optimization-based evolutionary algorithm for constrained optimization. *IEEE Trans. Evol. Comput.* **10**, 658–675. <https://doi.org/10.1109/tevc.2006.872344> (2006).
32. Wang, Y., Cai, Z. X., Guo, G. Q. & Zhou, Y. R. Multiobjective optimization and hybrid evolutionary algorithm to solve constrained optimization problems. *IEEE Trans. Syst. Man Cybern. Part B-Cybern.* **37**, 560–575. <https://doi.org/10.1109/tsmcb.2006.886164> (2007).
33. Jiao, L. C., Li, L., Shang, R. H., Liu, F. & Stolkin, R. A novel selection evolutionary strategy for constrained optimization. *Inf. Sci.* **239**, 122–141. <https://doi.org/10.1016/j.ins.2013.03.002> (2013).
34. Peng, C. D., Liu, H. L. & Gu, F. Q. A novel constraint-handling technique based on dynamic weights for constrained optimization problems. *Soft. Comput.* **22**, 3919–3935. <https://doi.org/10.1007/s00500-017-2603-x> (2018).
35. Wang, Y., Cai, Z., Zhou, Y. & Zeng, W. An adaptive tradeoff model for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.* **12**, 80–92. <https://doi.org/10.1109/tevc.2007.902851> (2008).
36. Fan, Q. & Yan, X. Differential evolution algorithm with co-evolution of control parameters and penalty factors for constrained optimization problems. *Asia-Pac. J. Chem. Eng.* **7**, 227–235 (2012).
37. Hou, Z. L., He, C. & Cheng, R. Reformulating preferences into constraints for evolutionary multi- and many-objective optimization. *Inf. Sci.* **541**, 1–15. <https://doi.org/10.1016/j.ins.2020.05.103> (2020).
38. Hansen, W. G. How accessibility shapes land use. *J. Am. Inst. Plann.* **25**, 73–76. <https://doi.org/10.1080/01944365908978307> (1959).
39. McKay, M. D., Beckman, R. J. & Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **42**, 55–61 (2000).
40. Feng, H. *et al.* Proposed design and optimization of a higher harmonic cavity for ALS-U. *Rev. Sci. Instrum.* **91**, 014712 (2020).
41. Yamada, S. in *Proc. 1981 Linac Conf.* (International Atomic Energy Agency).
42. Lewis, A., Weis, G., Randall, M., Galehdar, A. & Thiel, D. in *2009 IEEE Congress on Evolutionary Computation*. 1486–1492 (IEEE).
43. Todoroki, A. & Sekishiro, M. in *AIAA Infotech@ Aerospace 2007 Conference and Exhibit*. 2880 (Aerospace Research Central).
44. Elham, A. & van Tooren, M. J. Weight indexing for wing-shape multi-objective optimization. *AIAA J.* **52**, 320–337. <https://doi.org/10.2514/1.J052406> (2014).
45. Studios, C. M. & CST, M. CST Microwave studio. *CST Studio Suite* (2008).

Author contributions

P.W. executed the experiments and drafted of the manuscript. K.Y. contributed to the interface between optimizer and CST. X.H. directed the construction of the RF cavity in CST. J.W. contributed to the study conception and critical revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-27478-7>.

Correspondence and requests for materials should be addressed to J.W.

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

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



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

© The Author(s) 2023