DOI: 10.15276/hait.03.2020.2
UDC 004.67: 616.1

Setting up the genetic algorithm for the individualized treatment strategy searching

Anastasiia P. Dydyk
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
ORCID: https://orcid.org/0000-0003-2978-434X
Olena K. Nosovets
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
ORCID: https://orcid.org/0000-0003-1288-3528
Vitalii O. Babenko
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
ORCID: https://orcid.org/0000-0002-8433-3878

ABSTRACT

The genetic algorithm is a verified mechanism for optimization task solutions. Being a heuristic algorithm, it allows speeding up the task solving by the biological evolution principle use. Recently, this algorithm was offered as a method of an individualized treatment strategy search, where it was necessary to optimize the patient's state in the distant period, going through various treatment combinations. In that research as an optimization function, the additive convolution function of the patient's state indicators in the distant period was used, obtained with the help of Saaty analytic hierarchy process, which is one of the multi-criteria decision making methods. Despite showing good results, the genetic algorithm was set with standard parameters. Taking into consideration a big quantity of the parameters, the present study has the aim to find the optimum parameters for the algorithm. First of all, it is necessary for those, who incorporate this algorithm in their work, namely doctors, when they need to prescribe a treatment for a patient. The study describes the analysis of various genetic algorithm parameters and their use in algorithm test launches for individualized treatment strategy search. Also, the optimal patient's input parameter subsets were selected, using the correlation feature selection criterion. The selected parameters were necessary for modeling indicators of the patient's state after treatment. Modeling was performed via random forest classifier with preliminary divided total sample into training (eighty percent) and testing (twenty percent) ones. Two different databases of patients with congenital heart diseases were used for the study, allowing the optimal parameters being more reliable for their future use. In the end, it all allows finding the parameters, which are first of all exclusively recommendatory to the doctors before using the algorithm.

Keywords: genetic algorithm; individualized treatment strategy; correlation feature selection; random forest classifier

For citation: **Dydyk A. P., Nosovets O. K., Babenko V. O. Setting up the genetic algorithm for the individualized treatment strategy searching.** ***Herald of Advanced Information Technology.*** **2020;Vol.3, No.3: 125–135. DOI: 10.15276/hait.03.2020.2**

INTRODUCTION

It was not so long time ago that the algorithm of patient's treatment individualized strategy search was presented [1–2], which would allow the patient's state optimization in the distant period. This algorithm is able to find a necessary solution quickly via genetic algorithm principles [3]. The additive convolution [4] of the patient's state indicators in the distant period is used as an estimation criterion of treatment strategy.

Such an approach has the aim to give a doctor the possibility to look at the problem from a different perspective. When it is necessary to prescribe a certain treatment for a patient to treat the symptoms or eliminate complications after surgery, the doctors usually try to follow the protocols. However, there may be cases, when the majority of symptoms or complications cannot be treated, even by following the protocols. The cases with the patient's state aggravation are also possible. The algorithm goes through the maximum number of available treatment options with their probable outcomes. Treatment outcomes, in particular, patient's state indicators in the distant period, are preliminary modeled on the basis of historical data. Thus, the selection of treatment options is based on the patient's initial condition.

This method can theoretically suggest to the doctor, which treatment suits the patient best without deviation from the protocol, because the algorithm goes through the options from the preliminary prepared sample, determined by the very same protocol.

Nevertheless, there is a problem of which parameters are needed to provide the algorithm to find the necessary variants quickly. Since the genetic algorithm is the basis, there are a considerable number of parameters [5], each of which has a significant influence on the algorithm function

process. Each parameter has its advantages and disadvantages, so it is impossible to determine which of them is necessary. Since the algorithm for individualized treatment strategy search is designated for the use by doctors in their job, which means the immediate decision making of a treatment strategy for the patient, it is necessary to determine the optimal settings of the genetic algorithm first, so that the doctors will not waste their time.

ANALYSIS OF THE LATEST RESEARCH AND PUBLICATIONS

Genetic algorithm is an evolutional search algorithm, which is used for optimization tasks [6–9]. It uses the mechanisms that resemble biological evolution, which allows finding the necessary task solution quicker. The genetic algorithm suits optimization tasks well when the global extremum of the sought function is unknown. The use of genetic operators allows approaching a necessary point in *N* space with every new iteration. An example os such a task is finding an individualized treatment strategy. In Ukraine, the first echo of the solution to this problem can be found in [10], where linear programming was used as a solution method. This method was finding the extremum of a certain patient indicator after treatment while driving the rest into restrictions (which is essentially a conditional optimization). The use of the genetic algorithm in a similar problem is described in [1–2]. Unlike linear programming, the genetic algorithm is not able to solve conditional optimization problem tasks. Therefore, to not using restrictions, all patient indicators were reduced to one additive convolution function obtained by the Saaty analytic hierarchy process.

Even though the genetic algorithm is a powerful tool for solving optimization problems, it is still unclear which parameters are to be set for the algorithm to find a necessary point quicker. There are at least two types for each genetic operator, and their combination with each other can give various results. An attempt to pick better parameters was made in the [11], where the authors created their own program, but the result can differ depending on the assigned task. The present study aims to determine the optimal parameters of the genetic algorithm for doctors in the task of an individualized treatment strategy search. It is necessary to admit that these determined parameters will become a recommendation in the future and not a constant.

PARAMETERS OF THE GENETIC ALGORITHM

The common flowchart of the genetic algorithm is shown on *Fig*. 1.

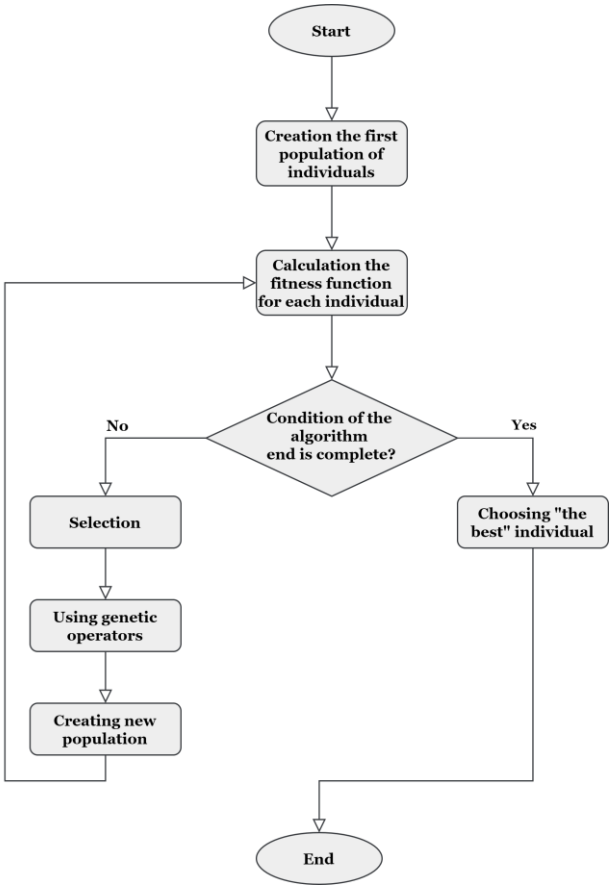


Fig. 1. **Flowchart of genetic algorithm**

If talking in more detail:

1. The first “population” of *N* “individuals” is formed. Individuals are randomly generated number arrays. Respectively, a population is an array, which consists of these generated arrays.

2. The fitness function [12] is calculated for every individual, which is a preliminarily determined criterion. That allows an evaluation of every individual.

3. After the fitness function calculation, the algorithm stop condition fulfillment is checked. It can be either finding a necessary fitness function, or reaching the algorithm iteration limit, or reaching the time limit. There are only 2 possible variants after checking the fulfillment of the condition:

3.1. The condition is fulfilled and the algorithm pulls the best individual out.

3.2. Condition is not fulfilled and the generation of a new population starts. As opposed to the first generation, it is generated not randomly, but via the transformation of the current population. In this case, genetic operators are used [13].

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

3.2.1. Individual picking is done from the current population, namely the selection of individuals (one of the genetic operators). To get new individuals, it is necessary to crossover the existing ones as it is in a real life. In order to create a better new individual generation, it is necessary to select "strong" individuals for the crossover. The strength is objectively determined via fitness function: the bigger it is, the more chances there are for an individual to take part in the crossover. There are two most popular individual selection methods [10]. The first one is a tournament selection when t (usually equals 2) individuals are randomly chosen from the population and the one with the biggest fitness function is selected among them. The second popular selection method is a roulette-wheel selection when the individuals are chosen with the help of N rotations of the roulette wheel. The roulette wheel has one sector for each individual and the sector size is directly proportional to the selection probability, which is calculated with the help of the fitness function. The selection and its methods are schematically shown on *Fig.* 2.

3.2.2. After the selection of individuals, their crossover is done, which means that the so-called genes in each pair are being mixed. As a result, a new individual is obtained, which contains information about its ancestors. There are many types of individual crossover [11].

The most popular of them are:

• single-point crossover, where a crossover point (point of a discontinuity) is randomly chosen, which allows to mix a pair of individuals. The genes of the first parent are before the point of discontinuity, and after it – the genes of the second parent.

That creates a new individual:

• two-point crossover, which is the same as above, but there are two crossover points;

• uniform crossover, where a random number from 0 to 1 is set in order to define which gene to use – from the first or the second parent. If the number is less than 0.5, the gene of the first parent is used, if otherwise, the gene from the second parent is used.

The working principle of each crossover type is shown on *Fig. 3* in more detail (g_i – individual gene, r_i – random number).

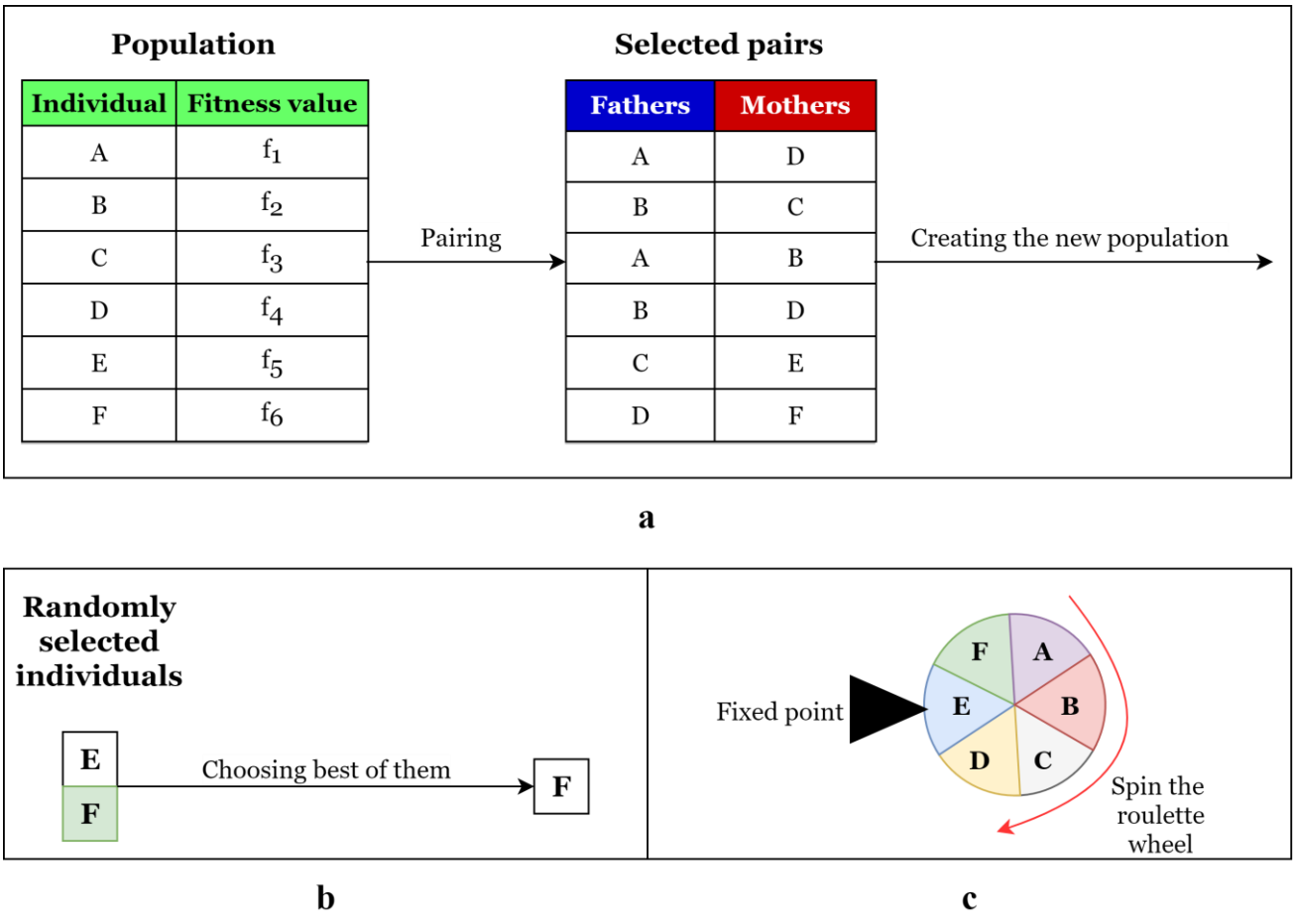


Fig. 2. **Selection operator:**
a – principle of selection; b – tournament type of selection; c – roulette-wheel type of selection

However, as practice shows, the constant use of the crossover leads to all individuals becoming the same, and the further work of the algorithm becomes meaningless. To maintain individual diversity until the algorithm work stops, the "mutation" is implemented [16]. This process replaces randomly selected genes by completely new ones. There are 2 types of mutation use: either it happens with 5 % chance instead of the crossover or as a result of an "incest", which means that new individual mutation takes place if the parents with considerably similar properties are used for crossover.

3.2.3. Getting a new population and going back to point 2. In this way, the algorithm can work long enough and ends only when the condition from point 3 is fulfilled.

It is important to underline some points:

• The roulette-wheel selection method can use normed and ranged fitness functions as a selection probability. The normed function is an individual fitness function, which is divided on the fitness function sum of all population individuals. The ranged function [5] is calculated by the following formula:

$$p_i = \frac{1}{N}(a - (2a - 2)\frac{i-1}{N-1}),$$

where: $1 \leq a \leq 2$ and is generated randomly; N – population size; i – individual's number in the sorted fitness function list.

• There is also a possibility to use so-called "elitism" selection [3]. This means that a certain number (by default it is 10 % of the entire population) of the best individuals ("elite") are chosen after fitness function calculation, which is guaranteed to get into the next generation without any changes for sure, and genetic operators take place with those individuals, which are left.

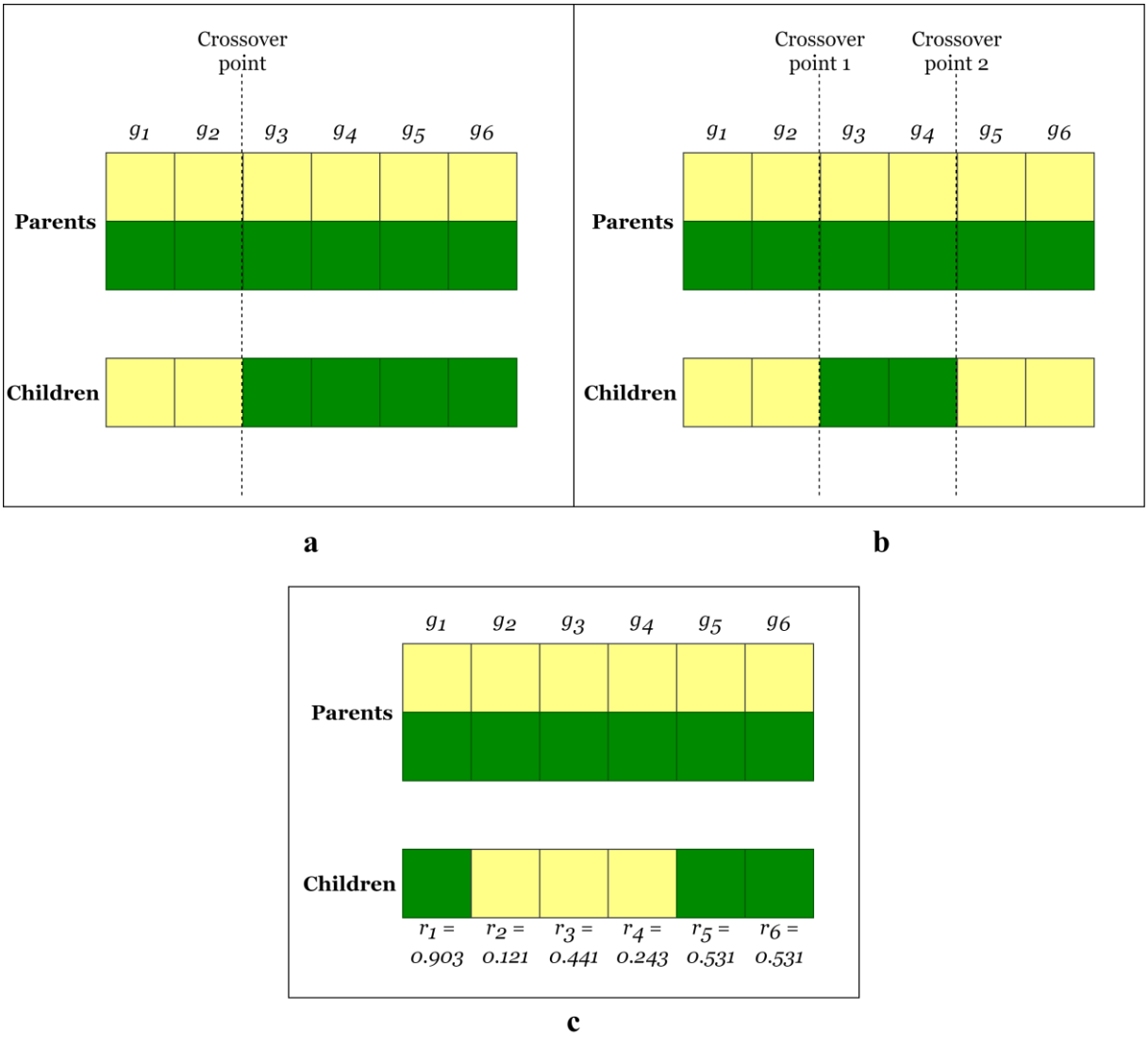


Fig. 3. **Crossover operator types:**
a – single-point crossover; b – two-point crossover; c – uniform crossover

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Thus, the following parameters of the genetic algorithm can be distinguished, which are shown on *Fig. 4* (all of them were used in the study; a total of 36 parameters' different combinations came out).

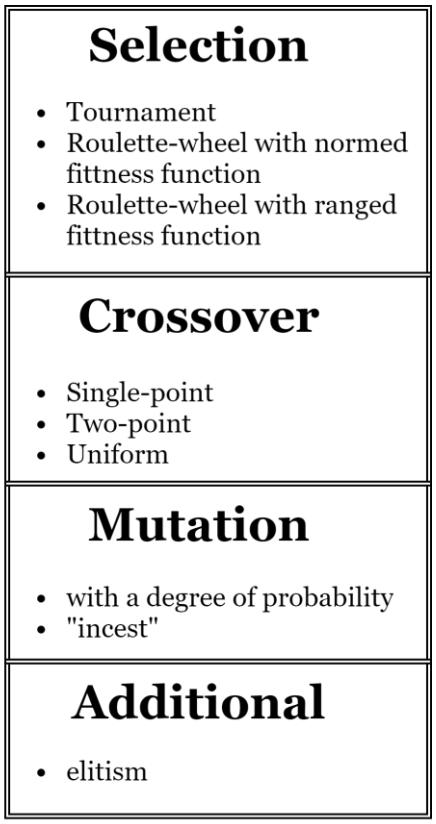


Fig. 4. **Parameters of genetic algorithm**

CLINICAL DATA DESCRIPTION

For the research, two databases of patients with congenital heart diseases provided by the Amosov National Institute of Cardiovascular Surgery were selected:

• The first database ("Base A") consists of 128 patients of the age from 3 to 28, who were first treated with the help of the surgery and afterward were used medication to eliminate post-surgery complications. For the study, only medication is taken into consideration. In total, the database contains 181 indicators. Each patient had from 4 to 18 complications (11 in average) from a maximum of 38 possible ones after treatment.

• The second database ("Base B") contains 144 patients of the age from 1 to 18. They were also treated in two ways (the difference was only in the methods). In total, the database contains 86 values. Each patient had from 1 to 8 complications (4 in averages) from a maximum of 9 possible ones after treatment and only 15 patients didn't have any complications.

It is worth emphasizing that for the study, 3 types of features were identified from the above-mentioned medical databases: patient indicators before treatment (patient's input state), treatment indicators, and patient indicators after treatment (patient's output state). The features' number of each type is shown in *Table.*

Table. **Distribution of features in databases**

Database	Input state indicators number	Treatment indicators number	Output state indicators number
"Base A"	121	22	38
"Base B"	67	10	9

Features, which describe the final output state of patients after applying a specific treatment strategy, are the indicators of certain complications presence. This state must be modeled in order to be optimized by use of the genetic algorithm. More precisely, it is necessary to build predictive models of patient indicators after treatment, using the patient's input state indicators and indicators of treatment. Since the patient's input state indicators are set by default, the genetic algorithm only needs to go through the indicators of treatment, thus finding the optimal values of output state indicators. That optimizes the final state of a patient, which is the search of an individualized treatment strategy since treatment indicators are gone through based on the individual patient's input state indicators.

RESEARCH OBJECTIVE

The aim of the study is to determine the recommendatory parameters for the genetic algorithm in the task of individualized patient's treatment strategy search.

In order to reach the aim, the following tasks were assigned:

1. Finding the optimal subset of input parameters according to the correlation feature selection criterion [17–18].

2. Using the determined subset for modeling the patient's condition indicators after treatment.

The tasks will be performed separately for each selected database. After that, the results will be compared at the end. This will allow achieving the main goal, namely, to obtain the optimal settings for the genetic algorithm, which the doctor can use in working with the algorithm.

RESEARCH RESULTS

1. Optimal subset of input parameters

Correlation features selection correlation [17–18] is calculated according to the below formula:

$$S_k = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}, \qquad (1)$$

where: $\overline{r_{cf}}$ – mean of absolute values of correlations between all independent features and dependent variable; $\overline{r_{ff}}$ – mean of absolute values of correlations of all independent features with each other; k – number of independent features in the subset.

This criterion evaluates a subset of independent features based on the following hypothesis: "Good subsets of features have features that strongly correlate with the dependent variable but do not correlate with each other". This hypothesis solves the problem of multicollinearity.

Since a subset can have both quantitive and qualitative features, Spearman correlation was used to calculate criterion (1). As a dependent variable, the additive convolution function of patient's condition after treatment was used, which is obtained with the help of Best-Worst multi-criteria decision making method [19] (it was used instead of the Saaty analytic hierarchy process as it is easier to interpret for doctors). This function was also used as a fitness function in the genetic algorithm.

The consultation with the doctors has made it clear that each treatment indicator is important and all of them are used for the treatment of a bunch of complications (38 complications total in "Base A", 9 complications total in "Base B"). This is the reason why the sought feature subset must contain all treatment indicators. Consequently, k varied from 23 to 50 for "Base A" (since it has 22 treatment indicators) and from 11 to 50 for "Base B" (it has 10 treatment indicators). On each k, the best feature subset was found according to the S criterion. The calculation results of S criterion are shown on *Fig. 5*.

Each found subset was used for modeling the indicators of the patient's condition after treatment. Since these indicators, which are the indicators of various complications' presence, are binary variables (1 – indicator is normal, 2 – indicator is abnormal), the random forest classifier [20] (which is one of the best machine learning algorithms), was used as the modeling algorithm.

The final subsets for each base were chosen according to the mean accuracy criterion of random forest classifiers on test samples (which comprise 20 % of all).

This criterion were calculated by next formula:

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{m_i},$$

where: $\overline{A}$ – mean accuracy of random forests, n – number of random forests; p – number of true predictions; m – total number of predictions.

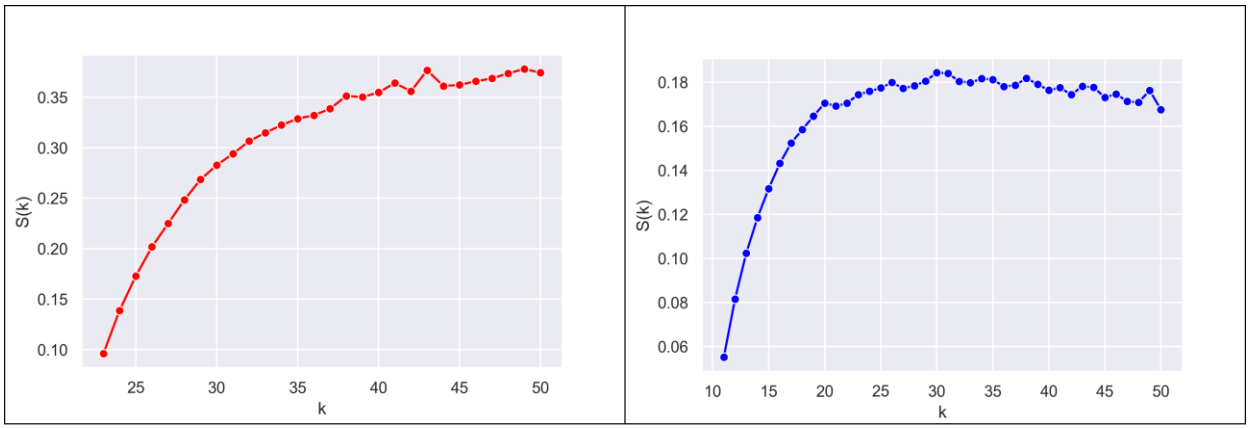
Accuracy values for each k are shown on *Fig. 6*.

It is clear that there is no dependence between S criterion and obtained random forest accuracy.

For "Base A", the best mean accuracy of 38 random forest classifiers was obtained on the subset of 29 values. Consequently, this so-called group was used for further research.

It consists of 22 treatment indicators and 7 following chosen indicators of the patient's condition before treatment:

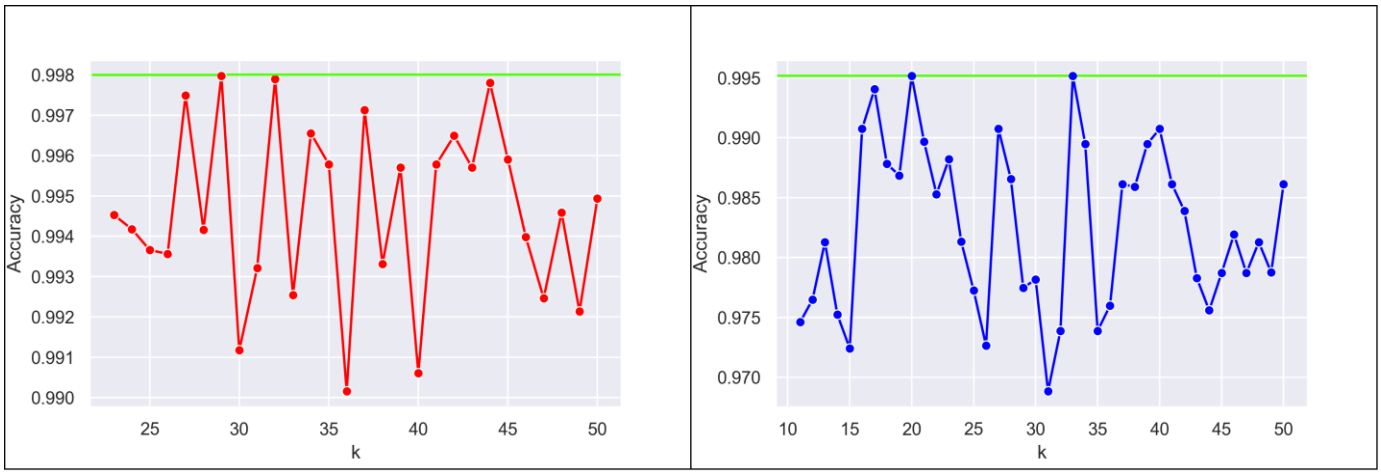
- thrombocytes quantity;
- aortic valve sinotabular junction size;
- aortic valve atresia presence (1 – no, 2 – yes);
- great vessel transposition presence (1 – no, 2 – yes);



a **b**

Fig. 5. **Criterion *S* values of various *k*: a – for "Base A"; b – for "Base B"**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)



a **b**

Fig. 6. **Mean accuracy values of classification random forests: a – for "Base A"; b – for "Base B"**

• left system ventricle morphology presence (1 – no, 2 – yes);
• right system ventricle morphology presence (1 – no, 2 – yes);
• surgical fenestration was performed (1 – no, 2 – yes).

For "Base B", the best mean accuracy of 9 classification random forests was obtained on the subsets of 20 and 33 values. Since they gave the same accuracy in the same conditions, it is more profitable to choose the group of a smaller number of values [21], which consists of 20.

It consists of 10 treatment indicators and 10 following indicators of the patient's condition before treatment:

• ventricle posterior wall diameter;
• calcinosis level;
• sternum stenosis presence (1 – no, 2 – yes);
• artificial blood circulation presence (1 – no, 2 – yes);
• intensive care unit (1 – no, 2 – yes);
• aortic compression duration;
• non-coronary cusp enhancement (1 – no, 2 – yes);
• exudate volume;
• aortic valve annulus z-score before treatment;
• aortic valve sinus z-score before treatment.

2. Genetic algorithm optimal parameters

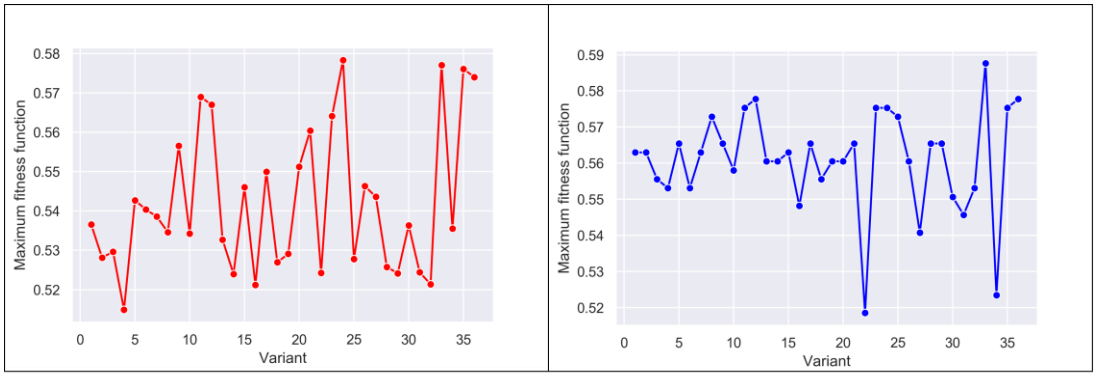
The algorithm with the same limit for all the parameters was used several items for each combination. The results are shown on *Fig. 7.*

As the result, the following was obtained:

1. The best five variants for "Base A" are:

• V24 – tournament selection type, two-point crossover, mutation with a degree of probability, and elitism principle presence.

• V33 – tournament selection type, uniform crossover, incest, mutation, and elitism principle absence.



a **b**

Fig. 7. **Results of the genetic algorithm: a – for "Base A"; b – for "Base B"**

• V35 – tournament selection type, uniform crossover, mutation with a degree of probability, and elitism principle absence.

• V36 – tournament selection type, uniform crossover, mutation with a degree of probability, and elitism principle presence.

• V11 – tournament selection type, single-point crossover, mutation with a degree of probability, and elitism principle absence.

2. The best five variants for "Base B" are:

• V33.

• V12 – tournament selection type, single-point crossover, mutation with a degree of probability, and elitism principle presence.

• V36.

• V11.

• V23 – tournament selection type, two-point crossover, mutation with a degree of probability, and elitism principle absence.

DISCUSSION OF THE OBTAINED RESULT

The conducted study provides the following results:

1. Comparison of the best value subsets, according to the *S* criterion, obtained from various *k*, showed that the accuracy of random forests, based on these subsets, does not depend on *S* criterion.

2. Comparison of genetic algorithm parameter results, obtained on various databases, gave the following "recommendatory" parameters:

• selection type – tournament;

• crossover type – uniform;

• mutation type – with a degree of probability;

• elitism principle use – absent.

These parameters are not the ultimate truth, but just give the possibility to a doctor to save time on looking through various parameters of the genetic algorithm, searching for the best-individualized treatment strategy.

CONCLUSIONS AND FURTHER RESEARCH PERSPECTIVE

As a result of the present study, the following tasks were completed:

1. The optimal subsets of input parameters were found for each of the databases according to the correlation feature selection. The additive convolution value of the patient's parameters after treatment was used as a dependent variable, obtained with the help of the Best-Worst method.

2. According to the obtained subsets, the random forest classifiers were built for each parameter of the patients in the distant period. The forests were built on the training samples (80 %), which were then evaluated on test samples (20 %). Referring to the random forest mean accuracy value, the best input parameter subsets were chosen for each base.

3. Obtained classification forests were used in test launches of the genetic algorithm to find the optimal individualized treatment strategy with sorting through 36 combinations of various algorithm setting parameters. As a result, the parameters were found, which performed in the best way for both first and second databases. Consecutively, these parameters can be used as recommendations for the algorithm usage by the doctors.

The obtained results can help the future studies significantly, connected to the individualized treatment strategy search, in particular to the direct use of the algorithm in practice. This is necessary in order to determine the real advantage of such an approach in practical task solutions.

REFERENCES

1. Babenko, V. "Sistema analizu rizikiv pislia hirurgichnogo likuvannia u ranniomu pisliaoperatsiynomu periodi". *International Scientific Journal "Internauka"*. 2019; Vol.8:18–22 (in Ukrainian). DOI: 10.25313/2520-2057-2019-8.

2. Dydyk, A. "Sistema analizu rizikiv pislia konservativnogo likuvannia u pizniomu pisliaoperatsiynomu periodi". *International Scientific Journal "Internauka"*. 2019; Vol.9: 29–35 (in Ukrainian). DOI: 10.25313/2520-2057-2019-9.

3. Ghaheri, A., Shoar, S., Naderan, M. & Hoseini, S. S. "The Applications of Genetic Algorithms in Medicine". *Oman Medical Journal.* 2015; Vol.30 No.6: 406–416. DOI: 10.5001/omj.2015.82.

4. Triantaphyllou, E. "Multi-criteria decision making methods". *Dordrecht: Kluwer Acad. Publ.* 2000.

5. Goldberg, D. E. "Genetic Algorithms in Search, Optimization & Machine Learning". *Addison-Wesley Publishing Company.* Inc. 1989. 432 p.

6. McCall, J. "Genetic algorithms for modelling and optimization". *Journal of Computational and Applied Mathematics.* 2005; Vol.184 No.1: 205–222. DOI: 10.1016/j.cam.2004.07.034.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

7. Holland, J. H. “Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence”. *A Bradford Book.* 1992. 232 p.

8. García, J. M., Acosta, C. A. & Mesa, M. J. “Genetic algorithms for mathematical optimization”. *Journal of Physics: Conference Series.* 2020; 1448. 012020. 5 p. DOI: 10.1088/1742-6596/1448/1/012020.

9. Maatouk, I., Jarkass, I., Châtelet, E. & Chebbo, N. “Preventive Maintenance Optimization and Comparison of Genetic Algorithm Models in a Series–Parallel Multi-State System”. *Journal of Intelligent Systems.* 2017. 12 p. DOI: 10.1515/jisys-2017-0096.

10. Nastenko, I., Pavlov, V., Nosovets, O., Zelensky, K., Davidko, O. & Pavlov, O. “Solving the Individual Control Strategy Tasks Using the Optimal Complexity Models Built on the Class of Similar Objects”. In: Shakhovska, N., Medykovskyy, M. (Eds.). “Advances in Intelligent Systems and Computing IV”. *CCSIT 2019. Advances in Intelligent Systems and Computing, Springer*. Cham. 2020; Vol.1080: 535–546. DOI: 10.1007/978-3-030-33695-0_36.

11. Villa Acuna, Y. P. & Sun, Y. “An efficiency improved genetic algorithm and its application on multimodal functions and a 2D common reflection surface stacking problem”. *Geophysical Prospecting.* 2020. 22 p. DOI: 10.1111/1365-2478.12920.

12. Jennings, P. C., Lysgaard, S., Hummelshøj, J. S., Vegge, T. & Bligaard, T. “Genetic algorithms for computational materials discovery accelerated by machine learning”. *NPJ Computational Materials*. 2019; Vol.5 No.1: 6 p. DOI: 10.1038/s41524-019-0181-4.

13. Whitley, D. “A genetic algorithm tutorial. Statistics and Computing”. 1994; Vol.4 No.2: 64–85. DOI: 10.1007/BF00175354.

14. Saini, N. “Review of Selection Methods in Genetic Algorithms”. *International Journal Of Engineering And Computer Science*. 2017; Vol.6 No.12: 22261–22263. DOI: 18535/ijecs/v6i12.04.

15. Kora, P. & Yadlapalli, P. “Crossover Operators in Genetic Algorithms: A Review”. *International Journal of Computer Applications.* 2017; Vol.162 No.10: 34–36. DOI: 10.5120/ijca2017913370.

16. Chebbi, O. & Chaouachi, J. “Effective parameter tuning for genetic algorithm to solve a real world transportation problem”. *20th International Conference on Methods and Models in Automation and Robotics.* 2015. p. 370–375. DOI: 10.1109/MMAR.2015.7283904.

17. Xuan, P., Guo, M. Z., Wang, J., Wang, C. Y., Liu, X. Y. & Y. Liu. “Genetic Algorithm-Based Efficient Feature Selection for Classification of Pre-MiRNAs”. *Genetics and Molecular Research.* 2011; Vol.10 No.2: 588–603. DOI: 10.4238/vol10-2gmr969.

18. Hall, M. A. “Correlation-Based Feature Selection for Machine Learning”. 1999. 109 p.

19. Dydyk, A., Nosovets, O. & Babenko, V. “Feature Selection for predicting the patient state in distant postoperative period”. *Herald of Advanced Information Technology*. *Publ. Nauka i Tekhnika.* Odessa, Ukraine: 2020; Vol.3 No.2: 34–41. DOI: 10.15276/hait.02.2020.3.

20. Belgiu, M. & Drăguţ, L. “Random forest in remote sensing: A review of applications and future directions”. *ISPRS Journal of Photogrammetry and Remote Sensing.* 2016; Vol.114: 24–31. DOI: 10.1016/j.isprsjprs.2016.01.011.

21. Stańczyk, U., Zielosko, B. & Jain, L. C. “Advances in Feature Selection for Data and Pattern Recognition”. *Publ. Springer.* 2017. 328 p.

DOI: 10.15276/hait.03.2020.2
UDC 004.67 : 616.1

Налаштування генетичного алгоритму для пошуку індивідуалізованої стратегії лікування

А. П. Дидик
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
ORCID: https://orcid.org/0000-0003-2978-434X
О. К. Носовець
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
ORCID: https://orcid.org/0000-0003-1288-3528

В. О. Бабенко
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
ORCID: https://orcid.org/0000-0002-8433-3878

АНОТАЦІЯ

Генетичний алгоритм є перевіреним механізмом у вирішенні задачі оптимізації. Будучи евристичним алгоритм, він дозволяє прискорити рішення задачі завдяки використанню принципів біологічної еволюції. Недавно даний алгоритм був запропонований як спосіб знаходження індивідуалізованої стратегії лікування, де необхідно оптимізувати стан пацієнтів у віддаленому періоді, перебираючи різні комбінації лікування. В цьому дослідженні в якості оптимізуючої функції було використано функцію адитивної згортки показників стану пацієнта у віддаленому періоді, отриману за допомогою метода аналізу ієрархій Сааті, який є одним із методів багатокритеріального прийняття рішень. Хоча генетичний алгоритм в поставленій задачі проявив себе непогано, слід відмітити, що були встановлені стандартні параметри алгоритму. Враховуючи, що параметрів немала кількість, в теперішній роботі була поставлена задача знайти оптимальні параметри для алгоритму. Це в першу чергу необхідно для тих, хто в майбутньому буде використовувати алгоритм в безпосередній роботі, а саме для лікарів, коли їм потрібно буде призначити лікування пацієнтові. Робота описує аналіз різних параметрів генетичного алгоритму і їх використання в експериментальних запусках алгоритму для пошуку індивідуалізованої стратегії лікування. Також були відібрані оптимальні підмножини вхідних параметрів пацієнта, використовуючи критерій кореляційного відбору ознак. Відібрані параметрі були необхідні для моделювання показників пацієнтів після лікування. Моделювання було виконано за допомогою випадкового лісу класифікації, попередньо розбивши загальну вибірку на навчальну (вісімдесят відсотків) і тестову (двадцять відсотків). Для дослідження були використані дві різні бази даних хворих з вродженими вадами серця, таким чином це дозволяє оптимальним параметрам бути більш надійними, щоб їх можна було надалі використати. Це все за підсумком дозволило знайти параметри, які насамперед будуть виключно рекомендованими для лікарів перед використанням алгоритму.

Ключові слова: генетичний алгоритм; індивідуалізована стратегія лікування; кореляційний відбір ознак; випадковий ліс класифікації

DOI: 10.15276/hait.03.2020.2
UDC 004.67 : 616.1

Настройка генетического алгоритма для поиска индивидуализированной стратегии лечения

А. П. Дыдык
Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина
ORCID: https://orcid.org/0000-0003-2978-434X
Е. К. Носовец
Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского, Киев, Украина
ORCID: https://orcid.org/0000-0003-1288-3528
В. О. Бабенко
Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина
ORCID: https://orcid.org/0000-0002-8433-3878

АННОТАЦИЯ

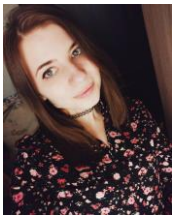
Генетический алгоритм является проверенным механизмом в решении задачи оптимизации. Являясь эвристическим алгоритмом, он позволяет ускорить решение задачи благодаря использованию принципов биологической эволюции. Недавно данный алгоритм был предложен как способ нахождения индивидуализированной стратегии лечения, где необходимо оптимизировать состояние пациента в отдаленном периоде, перебирая различные комбинации лечения. В этом исследовании в качестве оптимизируемой функции было использовано функцию аддитивной свёртки показателей состояния пациента в отдаленном периоде, полученную с помощью метода анализа иерархий Саати, который является одним из методов многокритериального принятия решений. Хоть генетический алгоритм в поставленной задаче проявил себя неплохо, стоит отметить, что были установлены стандартные параметры алгоритма. Учитывая, что параметров немалое количество, в нынешней работе была поставлена задача найти оптимальные параметры для алгоритма. Это в первую очередь необходимо для тех, кто в будущем будет использовать алгоритм в непосредственной работе, а именно для врачей, когда им нужно будет назначить лечение пациенту. Работа описывает анализ различных параметров генетического алгоритма и их использование в экспериментальных запусках алгоритма для поиска индивидуализированной стратегии

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

лечения. Также были отобраны оптимальные подмножества входных параметров пациента, используя критерий корреляционного отбора признаков. Отобранные параметры были необходимы для моделирования показателей пациентов после лечения. Моделирование было выполнено с помощью случайного леса классификации, предварительно разбив общую выборку на обучающую (восемьдесят процентов) и тестовую (двадцать процентов). Для исследования были использованы две разные базы данных больных с врождёнными пороками сердца, таким образом позволяя оптимальным параметрам быть более надёжными, чтобы их можно было использовать в дальнейшем. Это всё по итогу позволило найти параметры, которые прежде всего будут исключительно рекомендованными для врачей перед использованием алгоритма.

Ключевые слова: генетический алгоритм; индивидуализированная стратегия лечения; корреляционный отбор признаков; случайный лес классификации

ABOUT THE AUTHORS



Anastasiia P. Dydyk – student of the Department of Biomedical Cybernetics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
anastasiia.dydyk@gmail.com

Анастасія П. Дидик – студентка каф. біомедичної кібернетики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

Анастасия П. Дыдык – студентка каф. биомедицинской кибернетики, Национальний технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина

Olena K. Nosovets – Cand. Sci. (Eng.), Associate Prof. of the Department of Biomedical Cybernetics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
o.nosovets@gmail.com

Олена К. Носовець – канд. техніч. наук, доц. каф. біомедичної кібернетики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

Елена К. Носовец – канд. технич. наук, доц. каф. биомедицинской кибернетики, Национальний технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», пр. Победы, 37, Киев, Украина

Vitalii O. Babenko – student of the Department of Biomedical Cybernetics, , National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
vbabenko2191@gmail.com

Віталій О. Бабенко – студент каф. біомедичної кібернетики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

Виталий О. Бабенко – студент каф. биомедицинской кибернетики, Национальний технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина

Received 10.08.2020
Received after revision 12.09.2020
Accepted 18.09.2020