Evolutionary algorithm for automated formation of decision-making models for predicting the safety of opioid therapy

L V Lipinskiy¹, O D Melnikova¹, A S Polyakova¹, S A Evseeva¹, O P Bobrova², N A Shnayed³,³, S K Zyryanov⁴, M M Petrova¹, Y A Dychno, S N Zobova² and A V Bobrov¹

¹Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy av., Krasnoyarsk, Russia
²Prof. V. F. Voino-Yasenetsky Krasnoyarsk State Medical University, 1, Partizan Zheleznyak Str., Krasnoyarsk, Russia
³Bekhterev National Medical Research Center of Psychiatry and Neurology, Ministry of Health of Russia, Bekhterev St., Saint Petersburg, Russia
⁴Peoples’ Friendship University of Russia, Podolskoye Sh., 8/5, Moscow, Russia

E-mail: lipinskiyl@mail.ru, melok325@gmail.com, polyakova_nasty@mail.ru, sveta9227@mail.ru, bop_351971@mail.ru, naschnaider@yandex.ru

Abstract. In this paper, an evolutionary algorithm for solving the problem of predicting the safety of opioid therapy for patients with pancreatic cancer is proposed. Opioid analgesics such as fentanyl and morphine are used as a therapy for pain syndromes. Using the patient database, based on the results of the therapy applied to them, it is determined whether there is a correlation between the outcome and the combination of input data taken into account. To find a set of informative features, it is proposed to use the genetic algorithm for multi-criterion optimization, in which two criteria are reduced to one generalized criterion using the method of "additive convolution". The formed combination of the selected input features, which affects the outcome, is used to build a decision support model and to evaluate it afterwards.

1. Introduction

Anesthesia is one of the most important tasks in a doctor's work, since pain is the main cause of suffering and the patient's referral to doctors. Thus, according to estimates by the World Health Organization (WHO), 90% of all diseases are associated with pain, and patients with chronic pain are five times more likely to seek medical care than the rest of the population.

The most severe and acute pain, causing shock reactions, or moderate, but chronic, debilitating pain for the patient can be relieved primarily with the help of opioid analgesics [1]. An alternative to them in severe cases can be only general anesthetics (nitrous oxide, ketamine) and in some situations - clonidine or local anesthetics [2]. In ambulatory practice, the main cause of suffering for patients is chronic pain syndrome with incurable cancer in the terminal stage. In this situation, opioids have absolutely no alternative therapy means [3]. For complete pain relief and increased comfort in the last weeks or months of the life of such patients, the doctor must be well versed in modern medicines and be able to correctly apply them.
In this paper opioid therapy for patients with pancreatic cancer, which is one of the most severe types of cancer is considered. According to statistics, the percentage of 5-year survival of such patients after the moment of diagnosis is one of the lowest. Therapy is further complicated by the fact that malignant neoplasm is rarely detected at an early stage.

2. Problem description
The database represents two groups of cancer patients with different types of pain suppression therapy [4]. The first group includes 45 patients with Morphine therapy, the second group - 45 patients with Fentanyl therapy. The genotype chain is a collection of 13 types of genes.

The input variables are the 57 features (patient data).

The outputs are:

- resistance of patients with a certain genotype to therapy, (binary indicator, where 0 - effective treatment, 1 - resistant treatment);
- the appearance of adverse drug reactions (ADR) (binary indicator, where 0 - no reaction, 1 - presence of a reaction).

The following indicators are presented as ADR: "Constipation", "Nausea / Vomiting", "Sedation", "Weakness", "Dry mouth", "Dizziness", "Disorientation", "Skin itching", "Difficulty urinating", "Local reactions".

Using the patient database, based on the results of the therapy applied to them, it is necessary to determine whether there is a correlation between the outcome and the combination of input data taken into account.

Based on the presented database, 7 problems were solved with the following conditions:

- Therapy was performed with Morphine. The entire set of 57 features was used as an input. The binary value of the resistance of patients with a certain genotype to therapy was used as an output.
- The therapy was performed with Morphine. The entire set of features was used as an input, but with the exclusion of the indicator “Daily dose of morphine sulfate at the start of the study” (56 features). The binary value of the resistance of patients with a certain genotype to therapy was also used as an output.
- The therapy was carried out with morphine. The entire set of 57 features was used as input. As a conclusion, we used the binary value of the indicator of the occurrence of ADR in the form of constipation.
- The therapy was performed with Morphine. The entire set of 57 features was used as input. As an output, we used the combination of the outputs of all ADR indicators for each patient into a binary sign with a value of 0 - in the absence of any ADR, 1 - there is at least one ADR.
- The therapy was performed with Fentanyl. The entire set of 57 features was used as an input. The binary value of the resistance of patients with a certain genotype to therapy was used as an output.
- The therapy was performed with Fentanyl. The entire set of 57 features was used as an input. As an output, the binary value of the indicator of the appearance of ADR in the form of sedation was used.
- The therapy was performed with Fentanyl. The entire set of 57 features was used as an input. As an output, we used the combination of the outputs of all ADR indicators for each patient into a binary sign that has a value of 0 - in the absence of any ADR, 1 - at least one ADR is present.
3. Proposed approach

3.1. General scheme of the procedure for selecting informative features

In this paper, it is proposed to use the genetic algorithm for multi-criterion optimization (Multi-Objective Evolutionary Genetic Algorithm, MOEGA) to search for a set of informative features, in which two criteria are reduced to one generalized criterion using the method of "additive convolution" [5, 6]. The formed combination of the selected input features, which affects the outcome, is used to construct a classifier model and to evaluate it afterwards. Figure 1 shows a general outline of the approach described above (figure 1).

To evaluate the MOEGA model, an approach called "Stratified k-fold cross-validation" was used. The original sample is split into k = 5 equal blocks / parts (stratified proportional splitting). In turn, each block is considered as a test sample, and the remaining k-1 blocks are considered as a training sample. The model used in the evaluation of fitness of the individual (current characteristics) in MOEGA, trained on k-1 units and checked on a test block. The result of the model is estimated using the indicator F1-measure. The process is repeated k times, and we get k scores, for which the average value is calculated, which is the final estimation of the model.

At the end of the work of the MOEGA algorithm, it forms the best set of features, which is then used to train the classifier model and validate it. To train this model, the initial sample is divided into training (Train: 70% of the initial sample size) and test (Test: 30% of the initial sample size).

The model is trained on the entire training set. The next step is to check the quality of the model and its generalizing ability on real data that is not involved in training, that is, on a test set. Thus, the model is evaluated on the test set.

![Figure 1. The general scheme of approach to solve the problem of research of influence on the outcome of a procedure for the selection of informative features.](image-url)
Figure 1 also shows the estimation of the model with the best set of features, where \( P_{\text{train}} \) is the accuracy of the model using the best set of features obtained on the basis of MOEGA on a training set, \( P_{\text{test}} \) - on a test set.

Further, the approach "Stratified k-block cross-validation" was also used, but only for the training set. The training sample was divided into 5 equal blocks (parts), 4 of which were intended for training the model and 1 block was intended for its validation. The results obtained were evaluated and averaged to ensure that the quality of our model is independent of the dataset. In figure 1, \( P_{\text{Cross}} \) is the accuracy at the k-th block when it is a test block. Then the accuracy of the model is estimated on the test set. The criterion for the effectiveness of the model is also the F1-measure.

3.2. Scheme of the multi-objective evolutionary genetic algorithm

Genetic algorithms are adaptive search methods that are widely used to solve optimization problems. They use both an analogue of the mechanism of genetic inheritance and an analogue of natural selection [7, 8].

Terms such as gene, genotype, chromosome in this article are found in two senses - genes as the meaning of the genes of a particular patient, and genes as a traditionally established terminology within the framework of genetic algorithms. Genetic algorithms belong to the field of artificial intelligence and are indirectly related to biology. Therefore, in the future, these terms will be considered as artificial (symbolic) analogs using methods of natural evolution, such as inheritance, mutation, selection and crossing.

MOEGA is used to search for a set of informative features. Figure 2 shows the MOEGA chromosome coding (feature set), which is performed in 2 stages (figure 2).

3.2.1. Presentation of the original sample (data table) in the feature space. The first and second lines in figure 2 are not part of the chromosome, but represent a mask that allows decoding the chromosome and presenting it as a ready-made set of features.

The first row is a row in the original table, the length of which is equal to the original number of features (57). Possible variable values are indicated for each feature. Figure 2 shows that some of the variables are quantitative, and some are categorical.

![Figure 2. Method of coding the combination of genes for determining the outcome in MOEGA.](image)
The second line is a line in the changed attribute space, in which the quantitative variables of the table with the original data do not change, and the categorical variables that take only two values are coded 0 and 1. Otherwise, as many bits are allocated to encode categorical features as the number of values in a certain indicator.

For example, for the variable “rs1799971”, which takes one of the three values AA, AG or GG, three bits will be allocated. Accordingly, this feature will be presented in a row of the attribute space table in one of three combinations: 100 - AA, 010 - AG, 001 - GG.

Consider the variable “Stage”, which takes one of four values: 2a, 2b, 3, 4. Then, four bits will be allocated for this feature in the feature space and one of four combinations is expressed in the table row: 1000 - 2a, 0100 - 2b, 0010 - 3, 0001 - 4.

The length of the second line is 82 bits in the case of using all input features, and 81 in the case of excluding the indicator “Daily dose of morphine sulfate at the start of inclusion in the study”.

3.2.2. Representation of features in the MOEGA search space. Since the second line indicates the number of features, for each of which one bit is allocated (in the new feature space), therefore, in the third line this is reflected in such a way that the value 1 means that the feature is taken into account when building the model, and 0 - when it is not taken into account. The chromosome is 82 (81) bits long.

Evaluation of the effectiveness of MOEGA is carried out based on the following fitness function:

\[ f = \text{Accuracy} - 0.001 \cdot n_{\text{feat}} \] (1)

where Accuracy is the accuracy of the model according to the F1 criterion, and n_feat is the number of characters obtained in the MOEGA chromosome. This criterion is due to the fact that it is necessary to ensure the accuracy of the model, for which 1 term is responsible, and also to minimize the number of features, for which 2 term is responsible. The criterion reaches the greatest value while simultaneously achieving high model accuracy with a minimum set of features.

The following classification algorithms, built on the basis of the Scikit-learn library (Python), were used as models: logistic regression (LR), k-nearest neighbors algorithm for the classification problem (KNC), decision trees ensemble using the “random forest” method for classification problems (RFC), decision trees ensemble by the gradient boosting method for a classification problem (GBC), decision trees for a classification problem (DTC), an artificial neural network (multilayer perceptron) for a classification problem (MLPC), a linear support vector machine for a classification problem (LSVC), support vector machine for classification problem (SVC).

4. Experimental results

GA for the selection of informative features was launched with the following parameters: uniform crossing, tournament selection (tournament size = 5), average mutation, elitism, number of generations - 50, number of individuals - 70. All results were averaged.

Let us consider the results of solving problem No. 1. Table 1 presents the results of calculating the following characteristics: the mean over 50 runs (mean) and the standard deviation over 50 runs (std), determined from samples for each variable by a separate model:

- n_feature_mean is the average number of selected significant features as a result of the algorithm’s operation for 50 runs;
- n_feature_std is the standard deviation of the number of selected significant features as a result of the algorithm’s operation for 50 runs.
- Ptrain_mean is the average value of the accuracy of the algorithm for selecting significant features, calculated on a training sample for 50 runs.
- Ptrain_std is the standard deviation of the value of the accuracy of the algorithm for selecting significant features, calculated on a training sample of 50 runs.
• Ptest_mean - the average value of the accuracy of the algorithm for selecting significant features, calculated on a test sample of 50 runs.
• Ptest_std is the standard deviation of the value of the accuracy of the algorithm for selecting significant features, calculated on a test sample of 50 runs.
• P_Cross_mean - average value for 5-block cross-validation over 50 runs.
• P_Cross_std - standard deviation of the mean of 5-block cross-validation over 50 runs.

| Table 1. Descriptive Statistics (task No.1). |
|---------------------------------------------|
| Model  | n_feature_mean | n_feature_std | Ptrain_mean | Ptrain_std | Ptest_mean | Ptest_std | P_Cross_mean | P_Cross_std |
|--------|----------------|--------------|-------------|------------|------------|-----------|--------------|-------------|
| DTC    | 1              | 0            | 1           | 0          | 1          | 0         | 1            | 0           |
| GBC    | 1              | 0            | 1           | 0          | 1          | 0         | 1            | 0           |
| KNC    | 1.2            | 1.41         | 0.91        | 0.01       | 0.91       | 0         | 0.92         | 0           |
| LR     | 3.78           | 0.93         | 1           | 0          | 0.99       | 0.05      | 1            | 0           |
| LSVC   | 3.34           | 0.94         | 1           | 0          | 0.99       | 0.04      | 1            | 0.01        |
| MLPC   | 24.74          | 3.86         | 0.83        | 0.25       | 0.83       | 0.23      | 0.82         | 0.1         |
| RFC    | 12.2           | 4.65         | 1           | 0.01       | 0.89       | 0.14      | 0.93         | 0.06        |
| SVC    | 1              | 0            | 1           | 0          | 1          | 0         | 1            | 0           |

Based on the results presented in table 1, we can conclude that the DTC, GBC, SVC models solve the problem with an accuracy of 1 and use only one feature. As shown in table 1, there are no statistically significant differences between these models. However, the distinguishing feature is the dose of morphine. This effect is not obtained from a direct connection that the resistance is determined by the dose, but, on the contrary, that the doctor prescribes the dose taking into account the signs of resistance. Therefore, this model is not useful. To exclude this effect, we will remove from the sample the attribute "Dose of morphine" (the results are presented in task No. 2).

Let us consider the results of solving problem No. 2, in which therapy was performed with morphine, presented in tables 2 and 3. The entire set of signs was used as an input, but with the exception of the indicator "Daily dose of morphine sulfate at the start of the study" (56 signs). The binary value of resistance with a specific genotype to therapy was also used as an output.

| Table 2. Descriptive Statistics (task No.2). |
|---------------------------------------------|
| Model  | n_feature_mean | n_feature_std | Ptrain_mean | Ptrain_std | Ptest_mean | Ptest_std | P_Cross_mean | P_Cross_std |
|--------|----------------|--------------|-------------|------------|------------|-----------|--------------|-------------|
| DTC    | 16.58          | 6.37         | 1           | 0          | 0.75       | 0.15      | 0.82         | 0.05        |
| GBC    | 9.0            | 2.59         | 1           | 0          | 0.62       | 0.11      | 0.86         | 0.03        |
| KNC    | 5.24           | 1.59         | 0.87        | 0.05       | 0.63       | 0.09      | 0.84         | 0.01        |
| LR     | 14.12          | 2.75         | 0.99        | 0.02       | 0.66       | 0.16      | 0.91         | 0.03        |
| LSVC   | 31.96          | 3.75         | 0.83        | 0.1        | 0.62       | 0.15      | 0.73         | 0.09        |
| MLPC   | 32.3           | 4.16         | 0.83        | 0.25       | 0.55       | 0.14      | 0.63         | 0.11        |
| RFC    | 29.86          | 5.81         | 0.98        | 0.03       | 0.49       | 0.1       | 0.63         | 0.09        |
| SVC    | 5.12           | 3.54         | 0.91        | 0.05       | 0.64       | 0.16      | 0.76         | 0.07        |

To assess the statistical significance of the results obtained, each model uses a nonparametric statistical test - the Mann-Whitney U-test (significance level - 0.05). Table 3 shows the results of a comparative analysis of the Mann-Whitney U-test. The criterion returns the value p, which tells how much the sample distributions of the variable Ptest differ. The closer p is to one, the less the difference is. A dash in the table means that there is no scatter in the obtained model outputs.
Table 3. Checking the statistical significance of the results between models (task No. 2).

| Mann–Whitney U-test (p) | RFC | DTC | SVC | L SVC | GBC | LR | MLPC | KNC |
|-------------------------|-----|-----|-----|-------|-----|----|------|-----|
| 29.86                   | 1.0 | 0.0 | 0.0 | 0.0   | 0.0 | 0.0| 0.18 | 0.0 |
| 16.58                   | 0.0 | 1.0 | 0.0 | 0.0   | 0.0 | 0.0| 0.0   | 0.0 |
| 5.12                    | 0.0 | 0.0 | 1.0 | 0.19  | 0.05| 0.49| 0.06  | 0.09|
| 31.96                   | 0.0 | 0.0 | 0.19| 1.0   | 0.27| 0.25| 0.01  | 0.31|
| 9.0                     | 0.0 | 0.0 | 0.05| 0.27  | 1.0 | 0.09| 0.01  | 0.17|
| 14.12                   | 0.0 | 0.0 | 0.49| 0.25  | 0.09| 1.0 | 0.01  | 0.14|
| 32.3                    | 0.18| 0.0 | 0.0 | 0.01  | 0.01| 1.0 | 0.0   | 0.0 |
| 5.24                    | 0.0 | 0.0 | 0.09| 0.31  | 0.17| 0.14| 0.0   | 0.0 |

The results of tables 2 and 3 can be structured as follows. The best model in terms of accuracy on a test sample is DTC. Moreover, the result of the DTC model has a statistically significant difference from the results of all other models. Next comes the group of models GBC, KNC, LR, L SVC, SVC, the accuracy of which is approximately the same. Based on table 3, we can conclude that the differences between these models are statistically insignificant and are explained by the action of random factors. MLP and RFC models did the worst. These models differ significantly from all the others and are not significantly different from each other. The DTC model is taken as the resulting model. In addition to high accuracy, this model has one more significant advantage - the work of this model is easily interpreted in the language of an expert's reasoning and can be manually evaluated and corrected by an expert. The SVC and KNC models deserve special attention. These models use the least number of features. At the same time, the spread in the number of features for KNC is less. At the same time, KNC also lends itself to interpretation in the natural language of an expert's reasoning, since it builds its conclusion based on similar cases from the training set. This model also deserves attention and can be used for quick (preliminary) diagnosis of a patient.

Consider the results of solving problem No. 3.

Table 4. Descriptive statistics (task No.3).

| Model | n feature mean | n feature std | Ptrain mean | Ptrain std | Ptest mean | Ptest std | P_Cross mean | P_Cross std |
|-------|---------------|---------------|-------------|------------|------------|-----------|---------------|--------------|
| DTC   | 18.62         | 6.68          | 1           | 0          | 0.77       | 0.11      | 0.84          | 0.06         |
| GBC   | 21.86         | 4.71          | 1           | 0          | 0.79       | 0.1       | 0.88          | 0.03         |
| KNC   | 4.98          | 2.95          | 0.76        | 0.07       | 0.67       | 0.09      | 0.77          | 0.05         |
| LR    | 16.04         | 4.49          | 0.98        | 0.03       | 0.67       | 0.1       | 0.91          | 0.02         |
| L SVC | 34.36         | 3.26          | 0.9         | 0.09       | 0.65       | 0.09      | 0.72          | 0.08         |
| MLPC  | 29.86         | 4.58          | 0.95        | 0.15       | 0.7        | 0.1       | 0.75          | 0.07         |
| RFC   | 28.9          | 4.64          | 0.98        | 0.03       | 0.68       | 0.14      | 0.66          | 0.07         |
| SVC   | 12.28         | 4.79          | 0.94        | 0.06       | 0.63       | 0.12      | 0.77          | 0.04         |

Table 5 shows the results of checking the significance of features on the resulting models.

Table 5. Checking the statistical significance of the results between models (task No. 3).

| Mann–Whitney U-test (p) | RFC | DTC | SVC | L SVC | GBC | LR | MLPC | KNC |
|-------------------------|-----|-----|-----|-------|-----|----|------|-----|
| 28.9                    | 1.0 | 0.0 | 0.01| 0.06  | 0.0 | 0.35| 0.15 | 0.21|
| 12.28                   | 34.36| 21.86| 16.04| 29.86 | 4.98| 28.9| RFC  |     |

7
The results of solving problem No. 4 can also be grouped. The best model here is the LR. It gives a fairly high accuracy and is statistically significantly different from all the others (table 7). A good solution is also given by the DTC, GBC, MLPC models. These models differ significantly from all the

\[
\begin{array}{cccccccccc}
18.62 & DTC & 0.0 & 1.0 & 0.0 & 0.0 & 0.34 & 0.0 & 0.0 & 0.0 \\
12.28 & SVC & 0.01 & 0.0 & 1.0 & 0.06 & 0.0 & 0.0 & 0.0 & 0.0 \\
34.36 & LSVC & 0.06 & 0.0 & 0.06 & 1.0 & 0.0 & 0.16 & 0.0 & 0.1 \\
21.86 & GBC & 0.0 & 0.34 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\
16.04 & LR & 0.35 & 0.0 & 0.0 & 0.16 & 0.0 & 1.0 & 0.01 & 0.5 \\
29.86 & MLPC & 0.15 & 0.0 & 0.0 & 0.0 & 0.01 & 1.0 & 0.0 & 0.0 \\
4.98 & KNC & 0.21 & 0.0 & 0.0 & 0.1 & 0.0 & 0.5 & 0.0 & 1.0 \\
\end{array}
\]

The results of the models presented in tables 4 and 5 can be divided into three groups:

- The DTC and GBC models coped well. The results of these models differ statistically insignificantly and both models can be used as the resultant.
- The second group of models with an average result KNC, LR, LSVC, SVC. The differences between them are not statistically significant, but they differ significantly from other models. As in the previous task, KNC uses the least number of signs and can be used for quick diagnostics.
- The worst result was again given by MLPC, RFC models.

Consider the results of solving problem No. 4.

**Table 6.** Descriptive statistics (task No.4).

| Model | n_feature_mean | n_feature_std | Ptrain_mean | Ptrain_std | Ptest_mean | Ptest_std | P_Cross_mean | P_Cross_std |
|-------|----------------|--------------|-------------|------------|------------|-----------|--------------|-------------|
| DTC   | 15             | 4.47         | 1           | 0          | 0.76       | 0.06      | 0.68         | 0.16        |
| GBC   | 11.4           | 3.59         | 1           | 0          | 0.78       | 0.07      | 0.77         | 0.11        |
| KNC   | 4.9            | 2.68         | 0.6         | 0.1        | 0.71       | 0.11      | 0.31         | 0.18        |
| LR    | 16.5           | 4.63         | 0.94        | 0.04       | 0.81       | 0.1       | 0.55         | 0.19        |
| LSVC  | 34.88          | 4.5          | 0.89        | 0.09       | 0.66       | 0.07      | 0.7          | 0.17        |
| MLPC  | 29.66          | 4.48         | 0.86        | 0.18       | 0.75       | 0.13      | 0.49         | 0.19        |
| RFC   | 28.92          | 5.13         | 0.99        | 0.02       | 0.68       | 0.14      | 0.63         | 0.19        |
| SVC   | 16.46          | 5.71         | 0.99        | 0.04       | 0.65       | 0.09      | 0.51         | 0.18        |

**Table 7.** Checking the statistical significance of the results between models (task No. 4).

| Mann–Whitney U-test (p) | 28.92 | 15 | 16.46 | 34.88 | 11.4 | 16.5 | 29.66 | 4.9 |
|-------------------------|-------|----|-------|-------|------|------|-------|----|
| RFC                     | 1.0   | 0.0| 0.03  | 0.1   | 0.0  | 0.0  | 0.0   | 0.0|
| DTC                     | 0.0   | 1.0| 0.0   | 0.0   | 0.05 | 0.0  | 0.26  | 0.0|
| SVC                     | 0.03  | 0.0| 1.0   | 0.31  | 0.0  | 0.0  | 0.0   | 0.0|
| LSVC                    | 0.1   | 0.0| 0.31  | 1.0   | 0.0  | 0.0  | 0.01  | 0.01|
| GBC                     | 0.0   | 0.05| 0.0 | 0.0  | 1.0  | 0.0  | 0.31  | 0.01|
| LR                      | 0.0   | 0.0| 0.0   | 0.0   | 0.01 | 1.0  | 0.01  | 0.0|
| MLPC                    | 0.0   | 0.26| 0.0 | 0.0  | 0.31 | 0.01 | 1.0   | 0.01|
| KNC                     | 0.19  | 0.0| 0.01  | 0.0   | 0.0  | 0.0  | 0.01  | 1.0|

The results of solving problem No. 4 can also be grouped. The best model here is the LR. It gives a fairly high accuracy and is statistically significantly different from all the others (table 7). A good solution is also given by the DTC, GBC, MLPC models. These models differ significantly from all the
and are not significantly different from each other. The next group of models is RFC, KNC. At the same time, KNC again uses the least of all features.

The worst results are given by the LSVC and SVC models. LR is taken as the resulting model. Consider the results of solving problem No. 5.

Table 8. Descriptive statistics (task No.5).

| Model | n_feature_mean | n_feature_std | Ptrain_mean | Ptrain_std | Ptest_mean | Ptest_std | P_Cross_mean | P_Cross_std |
|-------|----------------|---------------|-------------|------------|------------|-----------|--------------|-------------|
| DTC   | 20.94          | 5.74          | 1           | 0.56       | 0.12       | 0.74      | 0.07         |
| GBC   | 11.76          | 4.5           | 1           | 0.71       | 0.15       | 0.83      | 0.04         |
| KNC   | 2.02           | 0.25          | 0.75        | 0.43       | 0.04       | 0.58      | 0.05         |
| LR    | 13.88          | 4.06          | 0.98        | 0.56       | 0.13       | 0.86      | 0.04         |
| LSVC  | 35.48          | 3.97          | 0.75        | 0.55       | 0.12       | 0.62      | 0.09         |
| MLPC  | 31.7           | 5.33          | 0.75        | 0.57       | 0.18       | 0.63      | 0.10         |
| RFC   | 13.44          | 4.28          | 0.96        | 0.71       | 0.19       | 0.67      | 0.11         |
| SVC   | 5.34           | 1.1           | 0.9         | 0.88       | 0.2        | 0.8       | 0.04         |

Table 9. Checking the statistical significance of the results between models (task No. 5).

| Mann–Whitney U-test (p) | 13.44 | 20.94 | 5.34 | 35.48 | 11.76 | 31.7 | 2.02 |
|-------------------------|-------|-------|------|-------|-------|------|------|
| RFC                     | 1.0   | 0.0   | 0.0  | 0.0   | 0.0   | 0.0  | 0.0  |
| DTC                     | 0.0   | 1.0   | 0.0  | 0.22  | 0.0   | 0.31 | 0.07 |
| SVC                     | 0.0   | 0.0   | 1.0  | 0.0   | 0.0   | 0.0  | 0.0  |
| LSVC                    | 0.0   | 0.22  | 0.0  | 0.1   | 0.0   | 0.39 | 0.45 |
| GBC                     | 0.4   | 0.0   | 0.0  | 0.0   | 1.0   | 0.0  | 0.0  |
| LR                      | 0.0   | 0.31  | 0.0  | 0.39  | 0.0   | 1.0  | 0.15 |
| MLPC                    | 0.0   | 0.07  | 0.0  | 0.45  | 0.0   | 0.15 | 1.0  |
| KNC                     | 0.0   | 0.0   | 0.0  | 0.0   | 0.0   | 0.0  | 1.0  |

The SVC model shows itself best on this task. This model provides a fairly high accuracy, although it has the largest spread in accuracy. At the same time, it uses on average only about five features. Despite the high scatter, this model significantly differs from other models and is accepted as the resulting one. Ensemble methods RFC, GBC cope somewhat worse. At the same time, they have a spread in accuracy close to SVC. KNC does the worst job, although it uses only about 2 features on average.

Consider the results of solving problem No. 6.

Table 10. Descriptive statistics (task No.6).

| Model | n_feature_mean | n_feature_std | Ptrain_mean | Ptrain_std | Ptest_mean | Ptest_std | P_Cross_mean | P_Cross_std |
|-------|----------------|---------------|-------------|------------|------------|-----------|--------------|-------------|
| DTC   | 14.08          | 3.91          | 1           | 0.67       | 0.11       | 0.84      | 0.03         |
| GBC   | 9.2            | 2.92          | 1           | 0.57       | 0.14       | 0.92      | 0.03         |
| KNC   | 3.6            | 2.07          | 0.75        | 0.68       | 0.08       | 0.68      | 0.11         |
| LR    | 13.48          | 4.76          | 0.91        | 0.73       | 0.15       | 0.86      | 0.02         |
| LSVC  | 30.82          | 3.77          | 0.77        | 0.56       | 0.14       | 0.66      | 0.08         |
| MLPC  | 35.1           | 4.56          | 0.83        | 0.63       | 0.16       | 0.6       | 0.09         |
| RFC   | 27.24          | 4.87          | 0.97        | 0.53       | 0.1        | 0.63      | 0.07         |
| SVC   | 7.82           | 2.58          | 0.98        | 0.78       | 0.13       | 0.83      | 0.08         |
Table 11. Checking the statistical significance of the results between models (task No. 6).

| Mann–Whitney U-test (p) | n_feature_mean | 27.24 | 14.08 | 7.82 | 30.82 | 9.2 | 13.48 | 35.1 | 3.6 |
|-------------------------|----------------|-------|-------|------|-------|-----|-------|------|-----|
| RFC                     | 1.0            | 0.0   | 0.0   | 0.1  | 0.07  | 0.0 | 0.0   | 0.0  | 0.0 |
| DTC                     | 0.0            | 1.0   | 0.0   | 0.0  | 0.0   | 0.0 | 0.02  | 0.05 | 0.07|
| SVC                     | 0.0            | 0.0   | 0.0   | 1.0  | 0.0   | 0.0 | 0.0   | 0.0  | 0.0 |
| LSVVC                   | 0.1            | 0.0   | 0.0   | 1.0  | 0.43  | 0.0 | 0.02  | 0.0  | 0.0 |
| GBC                     | 0.07           | 0.0   | 0.0   | 0.43 | 1.0   | 0.0 | 0.13  | 0.0  | 0.0 |
| LR                      | 0.0            | 0.02  | 0.01  | 0.0  | 0.1   | 1.0 | 0.1   | 0.0  | 0.0 |
| MLPC                    | 0.0            | 0.05  | 0.02  | 0.0  | 0.13  | 0.0 | 1.0   | 0.06 | 1.0 |
| KNC                     | 0.0            | 0.07  | 0.0   | 0.0  | 0.0   | 0.0 | 0.06  | 1.0  | 0.0 |

The best results for this task are given by SVC and LR. However, SVC uses only about 8 features on average, and LR on average uses about 13. Therefore, SVC is preferable and is accepted as the resultant. However, LR can be used as an auxiliary model. The worst accuracy is provided by the ensemble GBC, RFC, and LSVVC models.

Consider the results of solving problem No. 7.

Table 12. Descriptive statistics (task No.7).

| Model | n_feature_mean | n_feature_std | Ptrain_mean | Ptrain_std | Ptest_mean | Ptest_std | P_Cross_mean | P_Cross_std |
|-------|----------------|---------------|-------------|------------|------------|-----------|--------------|-------------|
| DTC   | 9.28           | 6.44          | 1           | 0          | 0.81       | 0.1       | 0.87         | 0.04        |
| GBC   | 4.6            | 1.7           | 1           | 0          | 0.98       | 0.05      | 0.94         | 0.01        |
| KNC   | 6.7            | 1.89          | 0.64        | 0.05       | 0.61       | 0.14      | 0.65         | 0.04        |
| LR    | 15.02          | 3.42          | 0.89        | 0.04       | 0.86       | 0.1       | 0.89         | 0.03        |
| LSVVC | 29.72          | 5.08          | 0.77        | 0.17       | 0.7        | 0.15      | 0.69         | 0.14        |
| MLPC  | 32.9           | 5.09          | 0.93        | 0.17       | 0.66       | 0.17      | 0.64         | 0.08        |
| RFC   | 28.5           | 5.88          | 0.99        | 0.02       | 0.73       | 0.12      | 0.71         | 0.08        |
| SVC   | 9.06           | 3.35          | 0.97        | 0.05       | 0.75       | 0.07      | 0.8          | 0.06        |

The GBC model worked best for this task. The accuracy of the model is close to 1, while the spread in accuracy is less than that of other models, and the GBC model uses on average only about 5 features to make a forecast. It is this model that is accepted as the resulting one for the given task.

5. Conclusion

The problem of predicting the occurrence of adverse side effects after opioid therapy is considered. In the course of the research, the influence of various clinical and pathogenetic factors in patients with pancreatic cancer on the efficacy and safety of analgesic therapy for chronic pain syndrome was studied. A comparative analysis of the efficiency of mining methods for solving these problems is carried out. A procedure for the formation of an effective set of informative features using the genetic algorithm for multi-criterion optimization NSGA-II is proposed.

Acknowledgements

Results were obtained in the framework of the state task No. 056-00119-18-00 of the Ministry of Science and Higher Education of the Russian Federation «Development of personalized approaches to safe and effective treatment of chronic pain in patients with malignant neoplasms». 
References

[1] Swegle J M and Logemann C D 2006 Management of common opioid-induced adverse effects 
American family physician 74(8) 1347-54
[2] Roberge C W and McEwen M 1998 The effects of local anesthetics on postoperative pain Aorn 
Journal 68(6) 1003-12
[3] Bobrova O P, Shnayder N A, Petrova M M and Gavrilyuk O A 2017 Genetic Polymorphism of 
Cytokines as a Predictor of Phenotypic Development of Chronic Pain Syndrome in Cancer 
Patients. International Journal of Biomedicine 7(4) 277-81
[4] Bobrova O P, Shnayder N A, Dychno Y A, Zyryanov S K and Petrova M M 2019 Chronic pain 
syndrome in patients with pancreatic cancer: individual therapy and its pathogenetic 
background Russian Open Medical Journal 8(2) 205
[5] Coello C A C, Lamont G B and Van Veldhuizen D A 2007 Evolutionary algorithms for solving 
multi-objective problems New York: Springer 5 79-104
[6] Fernandez A, Lopez V, del Jesus M J and Herrera F 2015 Revisiting evolutionary fuzzy systems: 
Taxonomy, applications, new trends and challenges Knowledge-Based Systems 80 109-121
[7] Back T 1996 Evolutionary algorithms in theory and practice: evolution strategies, evolutionary 
programming genetic algorithms (Oxford university press)
[8] Goldberg D E 1989 Genetic Algorithms in Search Optimization and Machine Learning (Addison 
- Wesley Reading MA)