https://doi.org/10.31891/csit-2023-4-9
UDC 004.9; 004.8

Ivan IZONIN
Lviv Polytechnic National University

AN UNSUPERVISED-SUPERVISED ENSEMBLE TECHNOLOGY WITH NON-ITERATIVE TRAINING ALGORITHM FOR SMALL BIOMEDICAL DATA ANALYSIS

Improving the accuracy of intelligent data analysis is an important task in various application areas. Existing machine learning methods do not always provide a sufficient level of classification accuracy for their use in practice. That is why, in recent years, hybrid ensemble methods of intellectual data analysis have begun to develop. They are based on the combined use of clustering and classification procedures. This approach provides an increase in the accuracy of the classifier based on machine learning due to the expansion of the space of the input data of the task by the results of the clustering.

In this paper, the tasks of modification and improvement of such technology for small data analysis are considered. The basis of the modification is the use of clustering with output at the first step of the method to increase the accuracy of the entire technology. Despite the high accuracy of the work, this approach requires a significant expansion of the inputs of the final linear classifier (labels of the obtained clusters are added to the initial inputs). To avoid this shortcoming, the paper proposes an improvement based on the introduction of a new classification procedure at the first step of the method and replaces all the initial inputs of the task with the results of its work. In parallel with it, clustering is performed taking into account the original attribute, the results of which are added to the output of the classifier of the first step. In this way, the formation of an expanded set of data of significantly lower dimensionality in comparison with the existing method takes place (here there is no longer a large number of initial features, which is characteristic of biomedical engineering tasks). This reduces the training time of the method and increases its generalization properties.

Modeling of the method was based on the use of a short dataset contained in an open repository. After the preprocessing procedures, the dataset has only 294 vectors, each of which was characterized by 18 attributes. Data classification was done using an SGTM-based neural-like structure classifier. This linear classifier provides high accuracy of work. In addition, it does not provide for the implementation of an iterative training procedure and additional adjustment of work parameters. Data clustering was performed using the k-means method. This choice is due to both the simplicity and speed of its work.

The search for the optimal number of k-means clusters was carried out using 4 different methods. They all showed different results. That is why, some experiments were conducted to assess the influence of different numbers of clusters (from 3 to 7) on the accuracy of all 4 algorithms of the developed technology. The accuracy of the proposed technology has been established experimentally in comparison with the linear classifier and the existing hybrid method. In addition, by reducing the inputs of the final classifier, the developed technology reduces the duration of the training procedure compared to the basic method. All this ensures the possibility of using the proposed technology when solving various applied problems of medical diagnostics, in particular, based on the analysis of small data.

Keywords: small data approach, non-iterative training, ensemble learning, unsupervised-supervised technology, biomedical engineering.

Іван ІЗОНІН
Національний університет «Львівська політехніка»

АНСАМБЛЕВА ТЕХНОЛОГІЯ БЕЗ ВЧИТЕЛЯ-З ВЧИТЕЛЕМ З НЕІТЕРАТИВНИМ АЛГОРИТМОМ НАВЧАННЯ ДЛЯ АНАЛІЗУ КОРОТКИХ НАБОРІВ БІОМЕДИЧНИХ ДАНИХ

Підвищення точності інтелектуального аналізу даних є важливою задачею в різних прикладних областях. Існуючі методи машинного навчання не завжди забезпечують достатній рівень точності класифікації для їх використання на практиці. Саме тому, в останні роки почали розвиватися гібридні ансамблеві методи інтелектуального аналізу. В їх основі покладено сумісне використання процедур кластеризації та класифікації. Такий підхід забезпечує підвищення точності класифікатора на основі машинного навчання за рахунок розширення простору вхідних даних задачі результатами роботи кластеризатора.

У цій роботі розглядаються завдання модифікації та удосконалення подібної технології аналізу коротких даних. В основі модифікації покладено використання кластеризації з виходом на першому кроці методу для підвищення точності роботи усього методу. Незважаючи на високу точність роботи, такий підхід вимагає суттєвого розширення входів фінального лінійного класифікатора (до початкових входів додаються мітки отриманих кластерів). З метою уникнення цього недоліку, у статті запропоновано удосконалення, в основі якого покладено введеня нової процедури класифікації на першому кроці методу та використання результатів її рботи для заміни усіх початкових входів задачі. Паралельно з нею виконується кластеризації з врахуванням вихідного атрибуту, результати якої додаються до вихода класифікатора першого кроку. Таким чином відбувається формування розширеного набору даних суттєво меншої вимірності в порівнянні із існуючим методом (тут вже немає великої кількості початкових ознак, що характерно для задач біомедичної нженерії). Це зменшує час роботи методу і збільшує його генералізаційні властивості.

Моделювання роботи методу відбувалося на основі використання корткого набору даних, який міститься у відкритому репозиторії. Після процедур попереднього опрацювання, набір даних налічував лише 294 вектори кожен з яких характеризувався 18 атрибутами. Класифікація даних відбувалася із використанням SGTM neural-like structure. Цей лінійний класифікатор забезпечує високу точність роботи. Окрім цього він не передбачає виконання ітераційної процедури навчання та додаткового налаштування параметрів роботи. Кластеризація даних відбувалася із використанням методу k-means. Такий вибір обумовлено як простотою так і швидкодією його роботи.

Пошук оптимальної кількості кластерів методу k-means відбувався із використанням 4 різних методів. Усі вони

продемонстрували різні результати. Саме тому, у статті проведено експерименти щодо оцінки впливу різної кількості кластерів (від 3 до 7) на точність роботи усіх 4 алгоритмів розробленої технології. Експериментальним шляхом встановлено підвищення точності роботи запропонованої технолоігї у порівнянні з лінійним класифікатором та існучим гібридним методом. Окрім цього, за рахунок зменшення входів фінального класифікатора, розроблена технологія зменшує тривалість процедури навчання в порівнянні з базовим методм. Все це забезпечує можливість використання запропонованої технології під час розвязання різноманітних прикладних задач медичної діагностики, зокрема на основі аналізу коротких наборів даних..

Ключові слова: підхід до малих даних, неітеративне навчання, ансамблеве навчання, технологія без вчителя-з вчителем, біомедична інженерія.

Introduction

The topicality of the task of increasing the classification accuracy in the biomedical engineering area is extremely important due to many factors [1] that affect the quality of diagnosis, prognosis, and treatment of various diseases.

Correct classification of biomedical data, such as biomarkers or genetic data, can help detect diseases such as cancer, heart disease, and neurological disorders faster and more accurately [2]. In addition, thanks to accurate classification, it is possible to develop personalized approaches to treatment. Knowledge about the nature of the disease at the individual level allows us to choose the optimal methods of treatment and the dose of medicines [3,4]. Here it should be taken into account that incorrect classification can lead to serious consequences for patients.

The high accuracy of classification models based on machine learning helps to reduce the risk of diagnostic errors and incorrect treatment assignments. In addition, accurate classification models allow rational use of medical resources, such as the time of medical professionals, equipment, and material resources, reducing the number of unnecessary diagnostic procedures [5]. However, the effectiveness of machine learning methods largely depends on the data set for analysis.

Biomedical sets of tabular data have their characteristics due to the specificity of the source and purpose of these data in biomedical research and medical practice. In particular, they have a multidimensional nature [4]. This is explained by the need to take into account during the analysis a large number of parameters measured as part of medical research. Such data may include clinical, laboratory, and other medical parameters that require a specific understanding of the medical context for effective analysis and interpretation.

Tabular sets of biomedical data can contain different types of features, including both numeric and categorical features. Effective analysis may require the use of different methods for different types of features [6]. In addition, such data may be collected from a sample of patients or studies, which may affect their representativeness.

The problem discussed above deepens in the case of the analysis of short biomedical datasets. In particular, there may be a limited number of examples for each or one of the classes. This leads to insufficient representativeness of the data, which leads to many problems when building machine learning models, especially for complex classes. In particular, in the conditions of analysis of a limited amount of data, the risk of overfitting increases significantly, when the model can correctly generalize to new examples. On the other hand, underfitting can occur here if the model is not complex enough to solve complex tasks.

Among other problems of classification of short sets of biomedical data using machine learning methods, the problem of automatic selection/extraction of important features for classification should be highlighted, taking into account the multidimensionality of such data [7]. Additionally, models trained on short datasets can be more vulnerable to outliers, noise, and anomalies in the input data. [8].

All this requires specialized approaches, methods, and an understanding of the specifics of the medical context to achieve reliable and valuable results when analyzing biomedical data sets of small volumes. That is why, to increase the accuracy of classification in biomedical engineering, new hybrid methods of data analysis are constantly being developed, which include machine learning, fuzzy logic [9], deep learning [10], kernel methods [11], and statistical approaches [12].

In particular, in [13] the authors considered a problem of classification using a hybrid, hierarchical approach. The authors proposed the use of clustering at the first step of the hierarchical method to select clusters in a given data set. In the second step of the method, the classifier is used within each separate separated cluster. This approach provides a significant increase in classification accuracy. However, in the case of analyzing short data sets, the selected clusters may be very small. This will make it impossible to use classifiers based on machine learning

To overcome this limitation, another approach was developed in [14]. The dataset according to [14] is not divided into clusters here, which is a significant advantage in the case of analysis of short data sets. In this case, clustering is also performed, but its results are used to expand the space of the input data of the problem by the observation belonging to each of the selected clusters.

This approach provides the possibility of intellectual analysis of small volumes of data, and significantly increases the accuracy of classification. However, it also expands the already multidimensional feature space of each vector in a biomedical dataset. This causes several problems when analyzing short sets of biomedical data using machine learning tools.

That is why this paper aims to improve the unsupervised-supervised classification technology for the case of the analysis of multidimensional short sets of biomedical data.

The main contribution of this paper can be summarized as follow:

1. The two-step method of intelligent data analysis [14] was modified by performing a data clustering procedure taking into account the original attribute, which allowed to increase the accuracy of classification in the field of biomedical engineering;
2. The two-step classification method has been improved due to the parallel execution of the data classification and data clustering procedures in the first step of the method, taking into account the original attribute, and the formation of a new dataset of significantly reduced size from their output signals for training the final classifier. This provides a significant reduction in the duration of the training procedure of a non-iterative classifier while increasing the accuracy of its work in the case of analysis of short sets of biomedical data.

Materials and methods

This paper is devoted to the development of unsupervised-supervised ensemble technology for data classification in medicine. It is based on several modifications and improvements of the hybrid method of data classification based on the consistent use of clustering and classification to improve the accuracy of the latter.

Let us consider the adaptation (Basic method, algorithm 1) and modification (Basic method, algorithm 2) of the method from [14] to increase the accuracy of solving classification problems.

Let a training sample of medical data be given in the form of a set of vectors of the form $x_1, x_2, ..., x_n \to y$. The task consists in assigning an observation with an unknown output to one of the K known classes

So, the main steps of the Basic method's (algorithm 1) training procedure are as follows:

1. We perform clustering of the training data sample using the selected method for selection C clusters. For this, we use vectors of the form $x_1, x_2, ..., x_n$. It should be noted that unlike [14], the next condition must be met here: $C > K$.
2. We calculate the centers for each of the found clusters C: $x_1^c, x_2^c, ..., x_n^c$
3. We form a new, expanded training dataset by adding to each observation the labels, belonging to each of the obtained clusters $m_l, l = 1,...C$. As a result, we obtained new vectors in the form $x_1, x_2, ..., x_n, m_1, ..., m_C \to y$. It should be noted that in the case when the current vector belongs to a cluster l, $l = 1,...C$, then $m_l = 1$. In all other cases $m_l = 0$.
4. We train the final classifier on extended vectors of the species $x_1, x_2, ..., x_n; m_1, ..., m_c \to y$ using the selected machine learning method.

The procedure for applying the Basic method (algorithm 1) requires the following steps:

1. We assign the current vector of the test sample $u_1, u_2, ..., u_n$ with an unknown output to one of the defined clusters C. To do this, we calculate the Euclidean distance between the current vector and each of the cluster centers found in step 2 of the training procedure. The smallest value of the Euclidean distance will determine whether the observation belongs to the corresponding cluster.
2. We form a new, expanded data vector by adding the membership labels of each of the obtained clusters to the current vector $u_1, u_2, ..., u_n, m_1, ..., m_C$.
3. We apply the final classifier to the formed vectors $u_1, u_2, ..., u_n, m_1, ..., m_C$ using the selected machine learning method for searching y^{final_pred}.

The modification of the basic method (algorithm 2) consists in using the data clustering procedure taking into account the initial value. In this case, we enter the initial value known for the training sample as an additional feature and perform step 2 of the training procedure of the previous method.

In the case of solving a classification task, the output value is the class to which the observation belongs. Therefore, each input vector of the training sample is expanded by a set of observation membership labels $m_q, q = 1,...K$ to each of the classes K defined by the stated task ($m_q = 1$, if the vector belongs to q-th class, and $m_q = 0$ in all other cases). Thus, clustering according to algorithm 2 will be performed on the vectors of the training set of the species $x_1, x_2, ..., x_n; m_1, ..., m_K$.

The application procedure of the basic method (algorithm 2) corresponds to the application procedure according to algorithm 1. It should be noted that the first step of the application procedure of algorithm 1 will be performed for cluster centers $u_1^c, u_2^c, ..., u_n^c$, without use $m_1, ..., m_K$.

The advantage of using clustering with output is the possibility of increasing the accuracy of this procedure, in particular with linear methods of machine learning, and, as a result, increasing the accuracy of the entire method. From a theoretical point of view, this effect can be explained by Cover's theorem [15]. However, the main drawback of both algorithms is a significant increase in the number of features of each data vector, which depends on the optimal number of clusters of the training sample. In addition, biomedical datasets are characterized by a large

number of independent attributes that must be considered during analysis. All this together will increase the training time of the selected classifier. In addition, this approach can be accompanied by a deterioration of the generalization properties of the classifier and even cause overfitting.

To avoid such shortcomings, this paper also proposes a technology for improving the basic method (both algorithms). It is based on the need to reduce the dimensionality of the space of extended vectors while ensuring high classification accuracy. For this, at the first step of the method, an additional classification procedure is introduced, the results of which will replace all the initial independent attributes of the initial dataset. In parallel with it, clustering will be performed, the results of which will expand each vector as in the basic method. Thus, the final classifier of the second step of the technology will receive a significantly smaller number of inputs, which will reduce the duration of the training procedure.

Let's take a closer look at the main steps of the Proposed technology (algorithm 3)**.**

1. We train the linear classifier on the vectors $x_1, x_2, ..., x_n \to y$.
2. We apply the training data sample to the previously trained classifier to obtain $y^{pred^{(1)}}$ for each data vector.
3. We perform clustering of the training data sample using the selected method for selection C clusters. For this, we use vectors of the form $x_1, x_2, ..., x_n$. It should be noted that unlike [14], the next condition must be met here: $C > K$.
5. We calculate the centers for each of the found clusters C: $x_1^c, x_2^c, ..., x_n^c$
4. We form a new, expanded training dataset, replacing all initial independent attributes with new ones. This happens by adding to $y^{pred^{(1)}}$ labels belonging to each of the obtained clusters $m_l, l = 1,...C$. As a result, we will receive a new training data set for the final classifier in the form of a set of vectors in the form $y^{pred^{(1)}}, m_1, ..., m_C \to y$. It should be noted that in the case when the current vector belongs to a cluster l, $l = 1,...C$, then $m_l = 1$. In all other cases $m_l = 0$.
5. We train the final classifier on extended vectors $y^{pred^{(1)}}, m_1, ..., m_C$ using the selected machine learning using the selected method of machine learning.

The structural diagram of the developed technology is shown in Fig. 1.

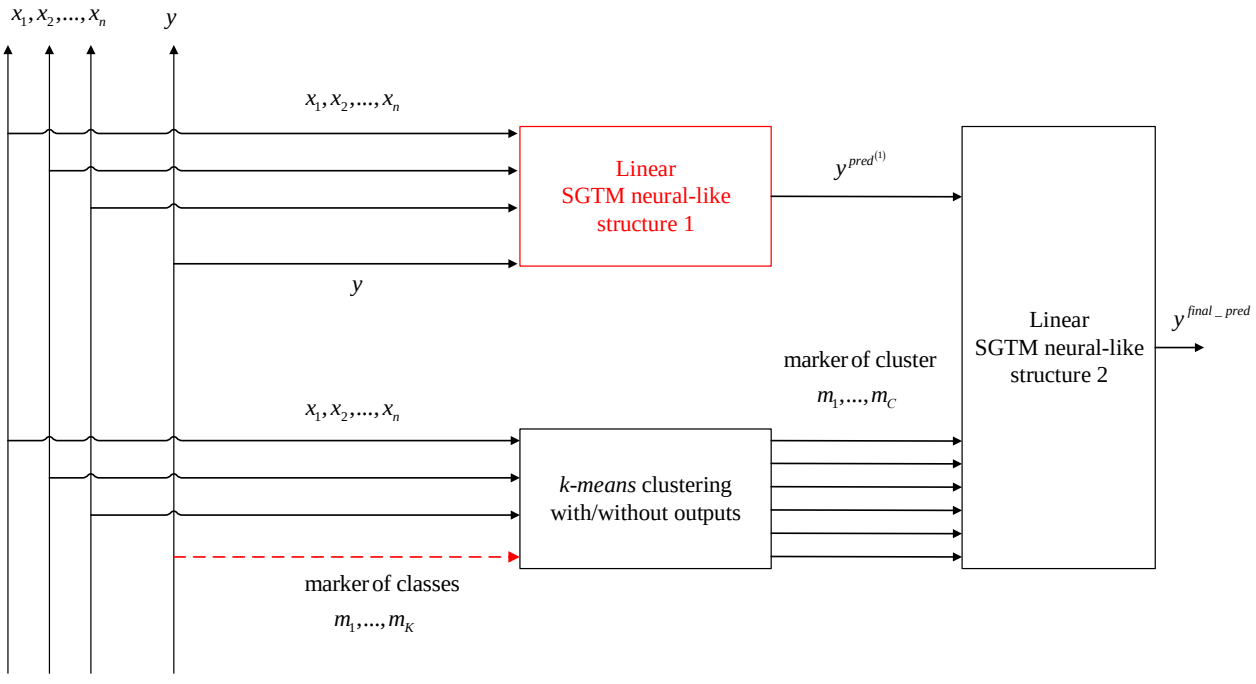


Fig. 1. Flow-chart of the improved unsupervised-supervised technology with non-iterative training algorithms

The application procedure of the Proposed technology (algorithm 3) requires the following steps.

1. We apply the current vector of the test sample to the previously trained first classifier. We get the predicted value $y^{u-pred^{(1)}}$,
2. We find the membership of the current vector of the test sample $u_1, u_2, ..., u_n$ with an unknown output to one of the defined clusters. To do this, we calculate the Euclidean distance between the current

vector and each of the cluster centers found in step 2 of the training procedure. The smallest value of the Euclidean distance will determine whether the observation belongs to the corresponding cluster.

3. We form a new, expanded data vector by adding labels of belonging to each of the obtained clusters to the corresponding $y^{u-pred^{(1)}}$. As a result, we get $y^{u-pred^{(1)}}, m_1, ..., m_C$.
4. We apply the final classifier to the formed vectors $y^{u-pred^{(1)}}, m_1, ..., m_C \to y$ using the selected machine learning method for searching $y^{final-pred}$.

An obvious advantage of the proposed technology would be a significant reduction of inputs for the final classifier. This will ensure the possibility of using the proposed technology for the analysis of short datasets in the tasks of medical diagnostics. However, to increase the accuracy of the technology, this article proposed the Proposed technology (algorithm 4). It consists in performing the clustering procedure with output, with identical steps as in the basic method (algorithm 4)**.**

Modeling, results, and comparison.

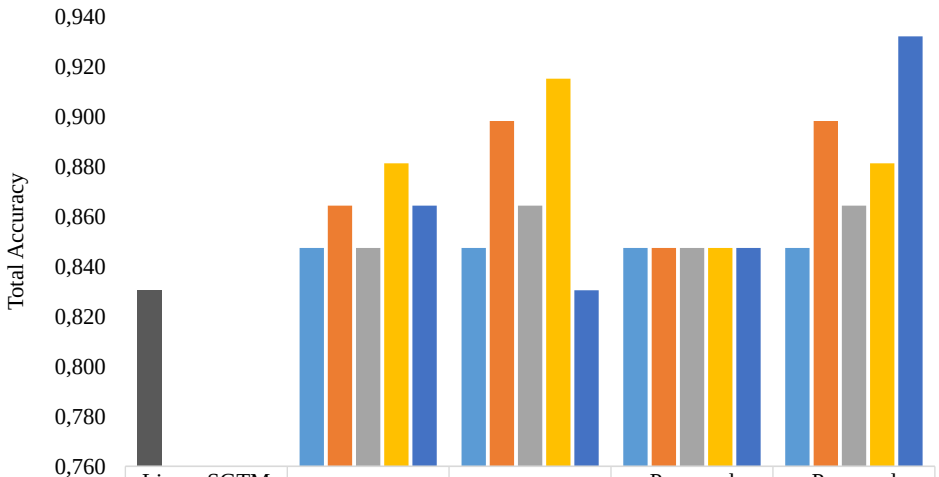
Both methods and their algorithms were modeled using a short set of medical data. An applied binary classification task, considered in this paper is to predict a heart attack based on 14 input attributes of a short dataset. The dataset is contained in an open repository [16].

The selected dataset contains a considerable number of gaps and categorical variables. In addition, the authors of the dataset did not specify which attributes are the most important. That is why the author carried out preliminary processing of the data, which consisted of:

- Removal of columns that contained a large number of omissions;
- Filling of missing values with average values for some observations;
- Conversion of categorical features into numerical ones;
- Performance of feature selection technique to select the most significant features for analysis.

After performing all the above procedures, the dataset contained 18 features and 294 observations. It was randomly divided into training and test samples in the ratio of 80% to 20%, respectively.

Data classification in this paper was done using SGTM neural-like structure. This linear classifier ensures high accuracy of work. In addition, it does not provide for the implementation of an iterative training procedure and additional adjustment of work parameters. Details of learning algorithms and the application of this classifier can be found in [17,18]. Data clustering was done using the *k-means* method. This choice is due to both the simplicity and speed of its work.



	Linear SGTM neural-like sctructure	Basic method - algorithm 1	Basic method - algorithm 2	Proposed technology - algorithm 3	Proposed technology - algorithm 4
3 clusters	0,831	0,847	0,847	0,847	0,847
4 clusters	0,831	0,864	0,898	0,847	0,898
5 clusters	0,831	0,847	0,864	0,847	0,864
6 clusters	0,831	0,881	0,915	0,847	0,881
7 clusters	0,831	0,864	0,831	0,847	0,932

Fig. 2. Comparison of the classification accuracy for all investigated methods using different numbers of clusters

The search for the optimal number of *k-means* clusters was carried out using 4 different methods. They all showed different results. That is why, in this paper, experiments were conducted on the assessment of the effect of different numbers of clusters (from 3 to 7) on the accuracy of all 4 algorithms of the developed technology.

Fig. 2 shows the total accuracy indicator for all investigated methods (Linear SGTM-based classifier and 4 algorithms developed in this paper) when using different numbers of clusters.

From the results presented in Fig. 2 the following conclusions can be drawn:

- Linear SGTM-based classifiers provide fairly high accuracy when analyzing a short dataset;
- The adapted method (algorithm 1) demonstrates an increase in the accuracy in comparison with the Linear SGTM-based classifier using any of the studied numbers of clusters;
- The use of clustering with output (algorithm 2) provides both a significant increase in the accuracy of work in particular in comparison with algorithm 1 and a slight deterioration. It depends on the number of clusters of the clustering procedure;
- The proposed technology (algorithm 3) increased the accuracy of the Linear SGTM-based classifier. However, the number of clusters here did not affect the change in this indicator's value;
- The use of the proposed technology (algorithm 4), which uses clustering with output, increased the accuracy of all previous methods. It should be noted that this improvement is characteristic of all values of the clusters studied in the paper;
- The highest classification accuracy was obtained using the proposed technology (algorithm 4) for seven clusters. In particular, it was possible to obtain a 10.1 higher accuracy compared to the Linear SGTM-based classifier.

In addition to the accuracy of work, an important indicator of the effectiveness of all studied algorithms is their training time. Since the training procedures of all four algorithms differ slightly, their training time was calculated as the duration for the final linear classifier at the second step of all methods.

Fig. 3 summarizes the results of this experiment.

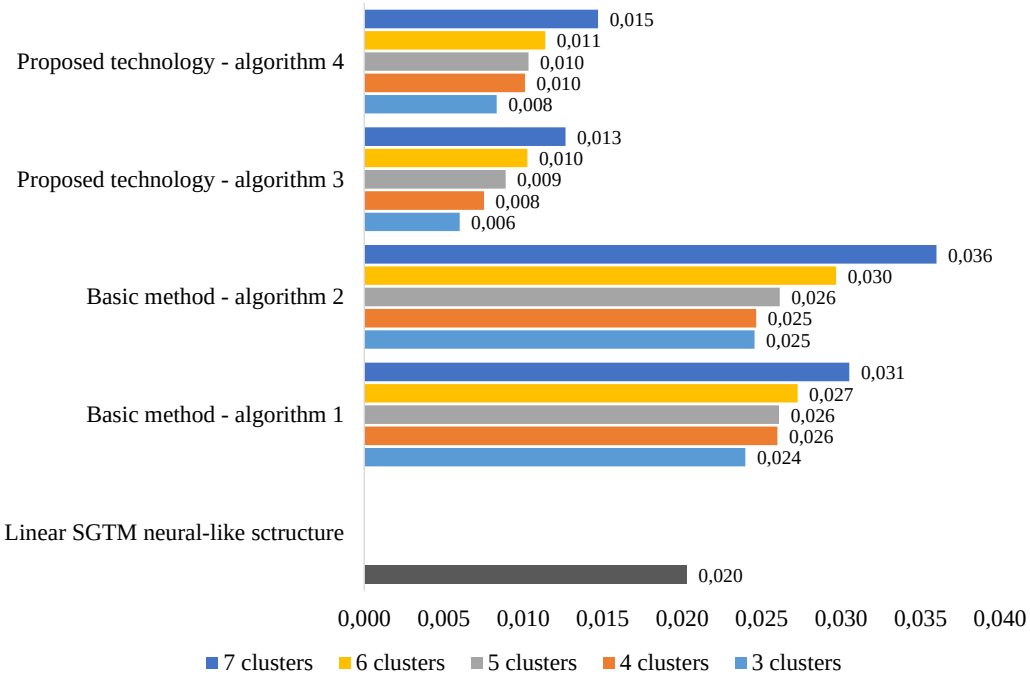


Fig. 3. Comparison of the training time (in seconds) for all investigated methods using different numbers of clusters

From the results presented in Fig. 3 the following conclusions can be drawn:

- Linear SGTM-based classifier provides the highest speed during the analysis of the studied short dataset;
- The adapted method (algorithm 1) demonstrates an increase in the duration of training time, in particular, when the number of clusters of this method increases;
- The use of clustering with output (algorithm 2) is accompanied by a slight increase in the training time of the method in comparison with algorithm 1;
- The proposed technology (algorithm 3) significantly reduced the duration of training time both in comparison with algorithm 1 and algorithm 1. This is explained by a significant reduction in the number of features used for analysis;
- The use of the proposed technology (algorithm 4), which applies clustering with output, does not significantly increase the duration of the training procedure compared to algorithm 3**.**

If we summarize the results presented in both figures, it should be noted that the developed technology (algorithm 4) demonstrated more than 10% higher classification accuracy with an insignificant increase in the

duration of the training procedure, compared to the basic linear classifier. All this provides several advantages for the use of the proposed technology when solving various applied tasks of medical diagnostics, in particular, based on the analysis of small data.

Conclusions

The modern development of the field of biomedical engineering is characterized by the appearance of an increasing number of tasks, including the processing of tabular datasets, with a limited amount of data for the implementation of training procedures by artificial intelligence tools. This imposes several limitations on the application of existing machine learning tools. The problem is deepened by the fact that this area is characterized by the presence of a large number of independent variables necessary for analysis.

To analyze short datasets of large dimensions, the paper proposes a hybrid technology of the combined use of clustering and classification. The author considers modified and improved algorithms that demonstrate a significant increase in classification accuracy during expert diagnostics.

The simulation was carried out on a real dataset to solve the binary classification task. Experimentally, it was established that the improved technology demonstrated more than 10% higher classification accuracy with an insignificant increase in the duration of the training procedure, compared to the basic linear classifier. Summarizing the advantages of improved technology, the following should be highlighted. It provides:

— a significant reduction of the space of input data for the classifier in comparison with the basic method of classification for tabular datasets [14];

— reducing the computational complexity of the improved method due to a significant reduction in the space of input data for further analysis by the selected classifier based on machine learning;

— a significant increase in the accuracy of solving the classification tasks in comparison with the existing method from [1];

— the possibility of processing data of small volumes with a large number of input attributes;

— the possibility of processing tabular datasets with a critically small number of observations in comparison with the hierarchical classifier from [13];

— the possibility of effective processing of datasets of large volumes due to an insignificant increase in the space of the input data of the task, but with a significant increase in the accuracy of the work of this approach;

— the possibility of building different combinations of clustering-classification methods depending on the research area, the task, and the volume and specificity of the available dataset.

All this ensures the possibility of using the proposed technology when solving various applied problems of medical diagnostics, in particular, based on the analysis of short data sets

Acknowledgment. This research is supported by the EURIZON Fellowship Program: “Remote Research Grants for Ukrainian Researchers”, grand № 138.

References

[1] Voronenko MA, Zhunissova UM, Smailova SS, Lytvynenko LN, Savina NB, Mulesa PP, et al. Using Bayesian methods in the task of modeling the patients’ pharmacoresistance development. IAPGOS 2022;12:77–82. https://doi.org/10.35784/iapgos.2968.

[2] Berezsky O, Zarichnyi M, Pitsun O. Development of a metric and the methods for quantitative estimation of the segmentation of biomedical images. EEJET 2017;6:4–11. https://doi.org/10.15587/1729-4061.2017.119493.

[3] Tolstyak Y, Havryliuk M. An Assessment of the Transplant’s Survival Level for Recipients after Kidney Transplantations using Cox Proportional-Hazards Model. CEUR-WSOrg 2022;3302:260–5.

[4] Tolstyak Y, Chopyak V, Havryliuk M. An investigation of the primary immunosuppressive therapy’s influence on kidney transplant survival at one month after transplantation. Transplant Immunology 2023;78:101832. https://doi.org/10.1016/j.trim.2023.101832.

[5] Hnatchuk Y, Hovorushchenko T, Pavlova O. Methodology for the development and application of clinical decisions support information technologies with consideration of civil-legal grounds. Reks 2023:33–44. https://doi.org/10.32620/reks.2023.1.03.

[6] Subbotin S, Tabunshchyk G, Arras P, Tabunshchyk D, Trotsenko E. Intelligent Data Analysis for Individual Hypertensia Patient’s State Monitoring and Prediction. 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan: IEEE; 2021, p. 1–4. https://doi.org/10.1109/SIST50301.2021.9465989.

[7] Perova I, Bodyanskiy Y. Adaptive human machine interaction approach for feature selection-extraction task in medical data mining. IJC 2018:113–9. https://doi.org/10.47839/ijc.17.2.997.

[8] Manziuk E, Barmak O, Krak I, Mazurets O, Pylypiak O. Method of features analysis on transition data. 2021 IEEE 3rd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine: IEEE; 2021, p. 272–7. https://doi.org/10.1109/ATIT54053.2021.9678787.

[9] Chumachenko D, Sokolov O, Yakovlev S. Fuzzy recurrent mappings in multiagent simulation of population dynamics systems. IJC 2020:290–7. https://doi.org/10.47839/ijc.19.2.1773.

[10] Krak I, Kuznetsov V, Kondratiuk S, Azarova L, Barmak O, Padiuk P. Analysis of Deep Learning Methods in Adaptation to the Small Data Problem Solving. In: Babichev S, Lytvynenko V, editors. Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making, vol. 149, Cham: Springer International Publishing; 2023, p. 333–52. https://doi.org/10.1007/978-3-031-16203-9_20.

[11] Bodyanskiy Y, Chala O, Kasatkina N, Pliss I. Modified generalized neo-fuzzy system with combined online fast learning in medical diagnostic task for situations of information deficit. MBE 2022;19:8003–18. https://doi.org/10.3934/mbe.2022374.

[12] Bisikalo O, Kharchenko V, Kovtun V, Krak I, Pavlov S. Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis. Entropy 2023;25:184. https://doi.org/10.3390/e25020184.

[13] Shakhovska N, Izonin I, Melnykova N. The Hierarchical Classifier for COVID-19 Resistance Evaluation. Data 2021;6:6. https://doi.org/10.3390/data6010006.

[14] Mishchuk O, Tkachenko RO. Neural Network with Combined Approximation of the Surface of the Response. KPI Science News 2018;2:18–24. https://doi.org/10.20535/1810-0546.2018.2.129022.

[15] Samuelson F, Brown DG. Application of Cover's theorem to the evaluation of the performance of CI observers. The 2011 International Joint Conference on Neural Networks, 2011, p. 1020–6. https://doi.org/10.1109/IJCNN.2011.6033334.

[16] Heart Attack Prediction n.d. https://www.kaggle.com/datasets/imnikhilanand/heart-attack-prediction (accessed August 16, 2023).

[17] Tkachenko R, Izonin I. Model and Principles for the Implementation of Neural-Like Structures Based on Geometric Data Transformations. In: Hu Z, Petoukhov S, Dychka I, He M, editors. Advances in Computer Science for Engineering and Education, vol. 754, Cham: Springer International Publishing; 2019, p. 578–87. https://doi.org/10.1007/978-3-319-91008-6_58.

[18] Tkachenko R. An Integral Software Solution of the SGTM Neural-Like Structures Implementation for Solving Different Data Mining Tasks. In: Babichev S, Lytvynenko V, editors. Lecture Notes in Computational Intelligence and Decision Making, Cham: Springer International Publishing; 2022, p. 696–713. https://doi.org/10.1007/978-3-030-82014-5_48.

Ivan Izonin **Іван Ізонін**	Ph.D. (Engineering), Associated Professor at the Department of Artificial Intelligence, Lviv Polytechnic National University https://orcid.org/0000-0002-9761-0096 e-mail: ivanizonin@gmail.com	Кандидат технічних наук, доцент, доцент кафедри систем штучного інтелекту Національного університету «Львівська політехніка»