

A Hybrid Two-ML-based Classifier to Predict the Survival of Kidney Transplants One Month after Transplantation

Ivan Izonin^a, Yaroslav Tolstyak^{b,c}, Veronika Kachmar^a, Roman Tkachenko^a, Mariia Rashkevych^a, Valentyna Chopyak^b

^a Lviv Polytechnic National University, S. Bandera str., 12, Lviv, 79013, Ukraine

^b Danylo Halytsky Lviv National Medical University, Pekarska str., 69, Lviv, 79010, Ukraine

^c Lviv Regional Clinical Hospital, Chernihivska str., 7, Lviv, 79010, Ukraine

Abstract

The paper deals with the task of predicting the survival of kidney transplants one month after transplantation using artificial intelligence tools. The use of machine learning to solve it can provide additional information to the doctor for timely and correct diagnosis and the possibility of adjusting therapy schemes. The authors used a real-world dataset obtained from the Lviv Regional Clinical Hospital, Ukraine. Taking into account the imbalance of the dataset, the effectiveness of using different approaches to balancing was investigated in this paper. It has been established that the class weights approach provides the best results for solving the stated problem. The authors developed the new hybrid two-ML-based classifier based on the sequential use of two machine learning algorithms. The results of the first of them, the Naive Bayes classifier, in the form of a set of probabilities belonging to each of the defined classes of the task, replace all the initial attributes with the found set of probabilities. The second machine learning method, the Random Forest algorithm, uses a new dataset of significantly reduced dimensionality to predict the result. The optimal parameters of both algorithms, which are the basis of the method, are selected. The efficiency of the method compared with several known approaches. The highest accuracy of the hybrid two-ML-based classifier, its high generalization properties, and the satisfactory duration of the training procedure were determined experimentally. All this ensures the possibility of its use when solving the real problems of transplantation medicine.

Keywords ¹

small data approach, kidney transplant survival, classification, machine learning, class weights, hybrid approach, Random forest, Naive Bayes, medical diagnosis, non-linear extensions

1. Introduction

The modern development of medicine is accompanied by the appearance of a large number of complex diagnosis and prediction tasks, which require the use of artificial intelligence for their effective solution [1], [2]. Today, a wide range of existing artificial intelligence tools provides the possibility of solving a large number of similar tasks.

If we talk about the tasks of analyzing tabular datasets in medicine, then it should be noted a large number of factors that affect the effectiveness of the selected machine learning methods. First, it is the multi-parameter nature of tabular datasets, which includes attributes from clinical, laboratory, and other types of research. It can reduce the generalization properties of a particular machine learning method. In addition, in such data, there is a large number of complex nonlinear interconnections between a large number of attributes of each data vector, which are very difficult for the clinician to detect.

IDDM'2023: 6th International Conference on Informatics & Data-Driven Medicine, November 17 – 19, 2023, Bratislava, Slovakia

EMAIL: ivanizonin@gmail.com (A. 1); tolstyakyaroslav@gmail.com (A. 2); veronika.kachmar.knm.2020@lpnu.ua (A. 3); roman.tkachenko@gmail.com (A. 4); marikupchak@gmail.com (A. 5); chopyakv@gmail.com (A. 6)

ORCID: 0000-0002-9761-0096 (A. 1); 0000-0002-5990-5977 (A. 2); 0009-0006-1512-9183 (A. 3); 0000-0002-9761-0096 (A. 4); 0000-0001-5490-1750 (A. 5); 0000-0003-3127-2028 (A. 6)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Medical data can be noisy or contain errors for various reasons, such as improper collection or registration [3]. They contain many omissions and outliers. If we are talking about classification tasks, then in the vast majority such data is unbalanced.

All these factors significantly affect the effectiveness of intellectual data analysis and, as a result, can lead to incorrect conclusions and diagnoses. In addition, if the medical dataset intended for analysis is critically small, then the impact of all of the above increases many times.

This paper aims to develop a hybrid two-ML-based classifier to improve the accuracy of medical diagnostics in the conditions of processing an unbalanced multi-parameter short dataset. The applied task solved in this paper consists of predicting the survival of kidney transplants one month after transplantation.

The significant scientific and practical results of this paper can be summarized as follows:

- We investigated the influence of the data balancing methods from different classes on the accuracy of ML-based classifiers when solving the task of predicting the survival of kidney transplants one month after transplantation in the conditions of processing an unbalanced short dataset;
- We proposed a procedure for data pre-processing by the Naive Bayes classifier to obtain a set of probabilities belonging to each data class and a procedure for replacing all independent initial attributes with the found set of probabilities to increase the accuracy and speed of the next classifier;
- We developed a hybrid two-ML-based classifier based on the consistent use of the Naive Bayes classifier and Random Forest algorithm to increase the classification accuracy of prediction of the kidney transplants survival one month after transplantation
- We selected the optimal parameters of the developed method and established the highest accuracy of its operation according to various indicators by comparing it with several existing methods.

2. State-of-the-arts

The task of predicting the risk of kidney graft failure after transplantation is not new. The application of artificial intelligence tools to its solution provides the possibility of taking into account a large number of hidden dependencies in a multi-parameter dataset, which provides the possibility of increasing the accuracy of solving this task [4], [5].

In particular, in [6] the authors investigated this task using one of the most widely used methods in statistical survival and time-to-event analysis, the Cox Proportional-Hazards Model (CPHM). It provides the ability to assess the impact of various factors on the risk of an event on a time scale. In [7] the CPHM was also used, however, for the assessment of the primary immunosuppressive therapy's influence on kidney transplant survival. Experimental studies of both tasks were conducted on a short set of data. Despite the sufficient accuracy of the obtained results, the CPHM assumes that the influence of risk factors should remain constant over time. In some cases, this assumption may be violated, and then the model may produce inaccurate or incorrect results.

In [8] the authors solved the kidney allograft failure prediction task using an artificial intelligence tool. The peculiarity of this paper is that the authors used a dataset, parts of which were obtained from 18 academic transplant centers around the world. Thus, the data sample consisted of more than 13 thousand vectors. As the chosen approach, the authors investigated the effectiveness of Bayesian joint models. The proposed approach's efficiency was evaluated using only the AUC curve. The obtained accuracy isn't at such a level that this approach can be used in practice.

In [9] the authors tried to predict the risk of kidney graft failure one year and five years after transplantation. The peculiarity of this paper is that they used a large dataset (more than 50 thousand observations). The classification was done using several well-known machine learning methods. The authors established the highest classification accuracy of 83% using SVM. Other methods showed significantly lower accuracy. The authors also developed a smart system for analyzing the patient's activity using the principles before such development [10], [11]. However, the low classification accuracy using the existing large dataset (relative to other studies) is not satisfactory.

In [12] the above task was also solved. The authors investigated the effectiveness of the application of the stacking strategy to increase the accuracy of solving the stated task. In this case, they worked

with only 513 observations. In connection with the imbalance of the dataset, the paper uses several balancing methods. Performance was evaluated using various indicators. The authors established the highest prediction accuracy of 87% using the C5.0 algorithm. The use of artificial neural networks in this case demonstrated significantly lower classification accuracy.

In [13] a stacking strategy was also considered to increase the classification accuracy of prediction of the risk of kidney graft failure one month after transplantation. The data set contained about 200 observations. The authors performed data pre-processing procedures (missing data recovery, feature selection), applied classic machine learning methods, and stacked the best of them in one method. The maximum obtained accuracy reached 91% (f1-score). However, considering that the dataset is not balanced, the accuracy of the stacking proposed by the authors could be improved using data balancing methods at the algorithmic level [14], and the speed of operation due to the parallelization [15].

In general, the above-mentioned approaches are quite complex, require qualification from the user, and do not provide high classification accuracy. In this paper, we tried to correct the above shortcomings.

3. Materials and methods

This section describes the dataset to predict the survival of kidney transplants one month after transplantation. The problem of analyzing an unbalanced dataset and the main classes of methods for its solution are considered. The analysis of two machine learning methods, which are the basis of the approach proposed in this paper, is presented. A description of the composition of the proposed approach and algorithms for its training and application are presented.

3.1. Dataset description

In this study, the outpatient medical histories of 146 patients who received HLA-compatible renal allografts between 1992 and 2020 were retrospectively analyzed: 146 patients were transplanted for the first time: 55 (37.6%) women and 91 (62.3%) men with an average age of 32.8 years \pm 8.4 years (range = 18–60 years) at the time of transplantation. All patients were receiving outpatient treatment in the nephrology and dialysis department of the Lviv Regional Clinical Hospital (LRKL), One Month after kidney transplantation [13].

The study used a group of patients who had transplantation for the first time, lived One Month after kidney transplantation, and received supportive immunosuppressive therapy. Kidney transplantation was performed in Ukraine (Kyiv, Zaporizhzhia, Kharkiv, Lviv, Odesa) in 79.8% of patients, and in other countries (Italy, Belarus, Pakistan, Poland, Turkey, China) - 20.2% of patients. 100% of patients from Ukraine received a kidney from a family donor, and 14.5% of patients, all from abroad, received a kidney from a deceased donor. The dataset consists of 40 features and one target attribute [13].

Clinical examinations of patients (collection of complaints, medical and life anamnesis, objective examination of patients) were carried out in the department of nephrology and dialysis of the Lviv Regional Clinical Hospital One Month after kidney transplantation. Laboratory research of general clinical indicators (general analysis of blood and urine and biochemical analysis of blood) was carried out in the central laboratory of the Lviv Regional Clinical Hospital according to the instructions for the use of laboratory research kits. In this retrospective analysis, a history of renal dysfunction and other data were obtained from personal history [13].

Viral infections were determined by the immunoenzymatic method and standard methods of polymerase chain reaction in private immunological laboratories.

3.2. A balancing approach for small data

In the medical diagnostics field, often arise the tasks, when one dataset's class (for example, patients with a rare disease) may be significantly smaller than another (healthy patients) [2]. As a result, we get unbalanced data samples that should be analyzed by machine learning methods [16]. A similar situation can lead to a decrease in the effectiveness of the selected classifier since machine learning methods

usually tend to learn a larger class and underestimate a smaller class [14], [17]. This leads to false conclusions about the work of the chosen method, which can, have serious consequences for the life and health of people.

The two main classes of methods for dealing with the imbalance of a stated set of medical data are data-level balancing and algorithmic-level balancing. The first class of methods includes the intuitive techniques of oversampling and undersampling. The underdamping technique removes redundant observations from more numerous classes to compare them with minor classes. However, reducing the data sample can lead to the loss of important information and reduce the accuracy of the classifier [18]. The over-sampling technique consists of adding additional observations to less represented classes. This can be done by duplicating existing samples or creating new, synthetic data, for example using SMOTE. In the case of analysis of short datasets, it is appropriate to solve the balancing problem using the latter approach.

The second class of methods is based on the balancing at the algorithmic level. In particular, methods based on changing the decision threshold can help balance decisions between classes. Typically, the model makes decisions based on the probability of class membership, and changing the threshold can affect positive and negative predictions. However, this requires an expert. Another technique is a method based on the use of class weights. It uses a higher weight when calculating the loss function specifically for smaller classes. This allows the machine learning method to pay more attention to smaller classes.

Given the need to analyze extremely short datasets, in this paper, we will investigate two balancing methods from different classes: SMOTE and class weights. Details of their work and practical implementation can be found in [18], and [19].

3.3. Naive Bayes classifier

The Naive Bayes classifier is based on Bayes' theorem, which originated in the 18th century, but the naive Bayes classifier as a machine learning method was formulated much later when computer technologies became more accessible [21].

The Naive Bayesian classifier is a machine learning method used to solve classification tasks. It is based on the assumption that all characteristics (or variables) of the sample are independent of each other, although this is rarely true.

The principle of operation of the naive Bayesian classifier includes the following steps [21]:

1. A collection of datasets in which objects have labels or classes, as well as a set of features that describe each object.
2. Calculation of posterior probabilities for each possible class (probability that this object belongs to a specific class) using Bayes' theorem.
3. Classification procedure. The object is assigned to the class for which the posterior probability is the highest.

The Naive Bayes classifier has its advantages and disadvantages, especially when applied to the analysis of short datasets.

Among the advantages, it should be noted the high speed of its operation. All calculations are based on calculating probabilities for individual features, so it is well suited for large datasets. However, the naive assumption of feature independence can increase classification performance when the training dataset is small.

Disadvantages of this machine learning method include the naive assumption of independence: This assumption rarely holds, which can lead to poor accuracy. In addition, if a priori assumptions about the class probabilities are incorrect or unreliable, this can significantly affect the classification results. In addition, to achieve high classification accuracy, it is necessary to have a large amount of training data, which can be a limitation in some domain areas.

3.4. Random Forest algorithm

Random Forest is a powerful machine learning method that uses an ensemble of trees to solve classification and regression tasks [18]. The main idea of the method is that instead of using one big

tree, several trees are created with random subsamples of data and random subsamples of features, and the results of their outputs are combined to provide more stable and accurate results. The main steps of implementing this algorithm are as follows [18]:

1. Construction of random subsamples of data with repetition (bootstrap sampling). This means that each tree will be built on its dataset, which may include duplicate objects.
2. Construction of each branch in the tree due to the selection of a subset of random features (Random Feature Selection). This helps to make the trees less correlated and diverse.
3. Building each tree randomly by choosing the best branch from a subset of features at each step.
4. Combining the results of the classification of each tree by voting and forming the result based on the largest number of votes.

The Random Forest algorithm has many advantages. In particular, due to bootstrap sampling and random feature selection, Random Forest is resistant to overfitting. In addition, it often shows high accuracy on a variety of tasks, particularly on small data with a large number of features. In addition, Random Forest can effectively work with large volumes of data without significant loss of performance.

However, this method also has disadvantages. Compared to the single trees that form it, a Random Forest can be difficult to interpret due to its complexity. In addition, building multiple trees can be time-consuming, especially on large datasets. However, these shortcomings do not reduce the relevance of using this machine learning method to solve the task.

3.5. Combined two-ML-based classifier

This paper proposes a combined approach to classification. It is based on the consistent use of the Naive Bayes classifier and Random Forest algorithm.

The Naive Bayes classifier is used for pre-processing the data to modify the features that will be submitted to the Random Forest algorithm. Naive Bayes classifier provides the possibility of forming the result in two ways: as a label belonging to a certain class of the task, and a set of probabilities of observation belonging to each of the classes of the task. In this paper, we use the last option. We use the obtained set of probabilities for each observation of the studied dataset to form a new dataset for the classifier based on the Random Forest algorithm. That is, we replace all the initial attributes with the output signals of the Random Forest algorithm.

The Random Forest algorithm uses a new dataset. The results of classification using this machine learning algorithm are the final desired result of the work of the hybrid two-ML-based classifier proposed in this paper. The flow-chart of the proposed approach is shown in Fig. 1

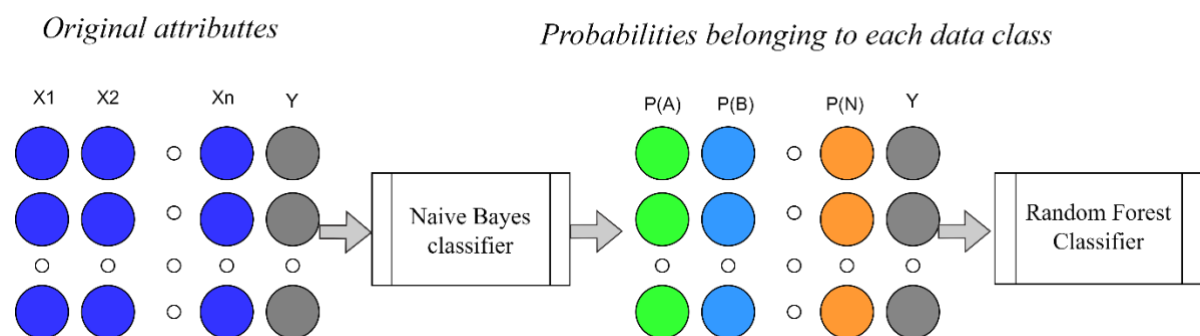


Figure 1: Flow-chart of the proposed hybrid two-ML-based classifier

The algorithmic implementation of the training procedure of the hybrid two-ML-based classifier involves the following steps:

1. Division of the data sample into training and test
2. Data normalization according to the selected method
3. Naive Bayes classifier training
4. Application of training and test samples on the previously trained Naive Bayes classifier to form a new dataset

5. Formation of a new dataset by replacing the initial features with a set of probabilities obtained by the Naive Bayes classifier for each observation from the dataset.
6. Training the Random Forest algorithm.

The algorithmic implementation of the application procedure of the hybrid two-ML-based classifier involves the following steps:

1. Normalization of the input vector, which should be assigned to one of the classes defined by the task.
2. Application of the normalized vector on the pre-trained Naive Bayes classifier to obtain its output signals.
3. Application of the set of probabilities obtained in the previous step to the previously trained Random Forest algorithm.
4. Obtaining the desired value.

4. Modeling and results

This section describes the modeling procedure. The optimal operating parameters of the developed method were selected and the obtained results were presented.

Modeling took place using a real-world dataset obtained from various hospitals in Ukraine. The data set had 146 observations and it is unbalanced. It was randomly split 80% to 20%. Data was normalized using MaxAbsScaler.

4.1. Results on data balancing

First, we investigated the effectiveness of applying two balancing methods from different classes. We used data-level balancing by applying the SMOTE algorithm. In this case, we increased the number of representatives of the smaller class from 37 to 109 using the above approach. Another method, class weights, provides balancing at the algorithm level by penalizing errors in a smaller class. It uses a set of weights that can be formed as a ratio of a larger class to a smaller one and uses them in the process of the selected machine learning algorithm. Due to this, the smaller class is given more attention, which increases the accuracy of its processing.

Fig. 2 summarizes the results of the work of both classifiers on original, unbalanced data, and using both balancing methods.

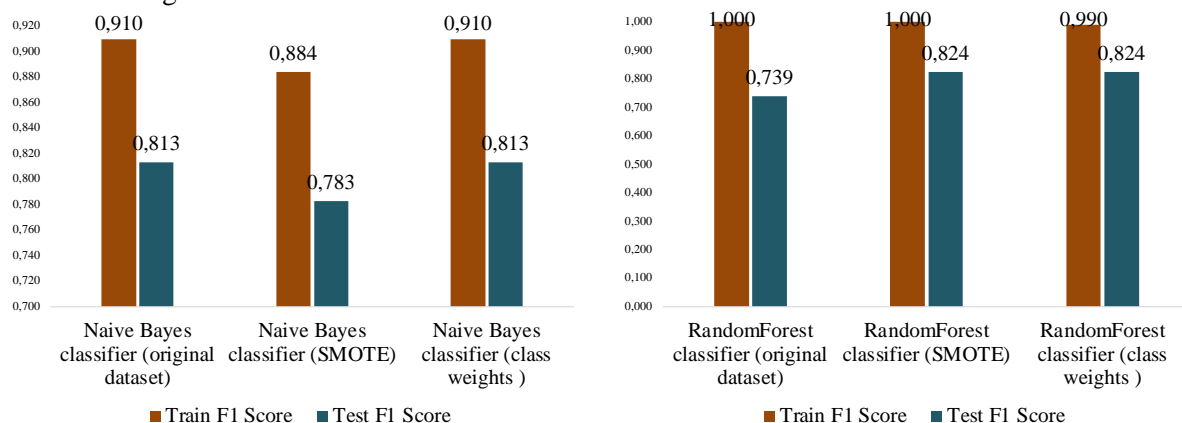


Figure 2: Data balancing results

The following conclusions can be drawn using Fig. 2:

- Naive Bayes classifier using the original, unbalanced dataset provides fairly high classification accuracy, taking into account the large number of input attributes and the limited volume of the studied dataset;
- Application of the SMOTE algorithm significantly reduced the classification accuracy based on the F1-score indicator;

- The use of the class weights approach for dataset balancing provides the same level of accuracy as the classification by the Naive Bayes algorithm on the original data;
- Random Forest classifier using the original, unbalanced dataset shows significantly lower classification accuracy compared to Naive Bayes classifier;
- Applying the SMOTE algorithm to balance the dataset for the Random Forest classifier increased the classification accuracy based on the F1-score, but there is observed the overfitting problem;
- The use of the class weights approach for dataset balancing provides the same level of accuracy in the test mode as the use of the SMOTE method, but an overfitting problem is not observed in this case. Considering everything described above, further research will use the class weights method for balancing the studied dataset.

4.2. Results

In the paper, the optimal parameters of each of the machine learning methods, which are the basis of the proposed, hybrid two-ML-based classifier, were selected. We implemented the grid search method from the Python library to achieve this goal.

The optimal operating parameters of the Naive Bayes classifier are as follows: *'var_smoothing': 1e-09*.

Accordingly, the optimal parameters of the Random Forest classifier are as follows: *'bootstrap': True, 'max_depth': 25, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 2, 'n_estimators': 100*.

The results of the developed method when using the optimal parameters of its operation based on various performance indicators are presented in Table 1

Table 1

Proposed method results

| Performance indicator | Training mode | Test mode |
|--|---------------|-----------|
| Total accuracy | 0,976 | 0,931 |
| Precision | 0,978 | 0,945 |
| Recall | 0,976 | 0,931 |
| F1-score | 0,976 | 0,933 |
| Matthews's correlation coefficient (MCC) | 0,936 | 0,840 |
| Cohen's Kappa | 0,938 | 0,851 |
| Training time (seconds) | 1,780 | - |

As can be seen from Table 1, the developed method:

- Ensures high classification accuracy based on all studied performance indicators (in particular, F1-score in test mode equal 93%);
- Provides high generalization properties (the difference between the metrics of both training and test modes is small);
- Does not provoke overfitting, which is typical when analyzing small data.

5. Comparison and discussion

To evaluate the results of the work of the hybrid two-ML-based classifier, both machine learning algorithms that form the method and several existing methods were used, in particular:

- Probabilistic Neural Network with next optimal parameters: the Canberra distance, smooth factor = 6.51617;
- SVR with RBF kernel with next optimal parameters: 'C': 1, 'coef0': 1.0, 'degree': 2, 'gamma': 1, 'kernel': 'sigmoid', 'max_iter': 10000

The results of such a comparison based on different performance indicators are summarized in Fig. 3.

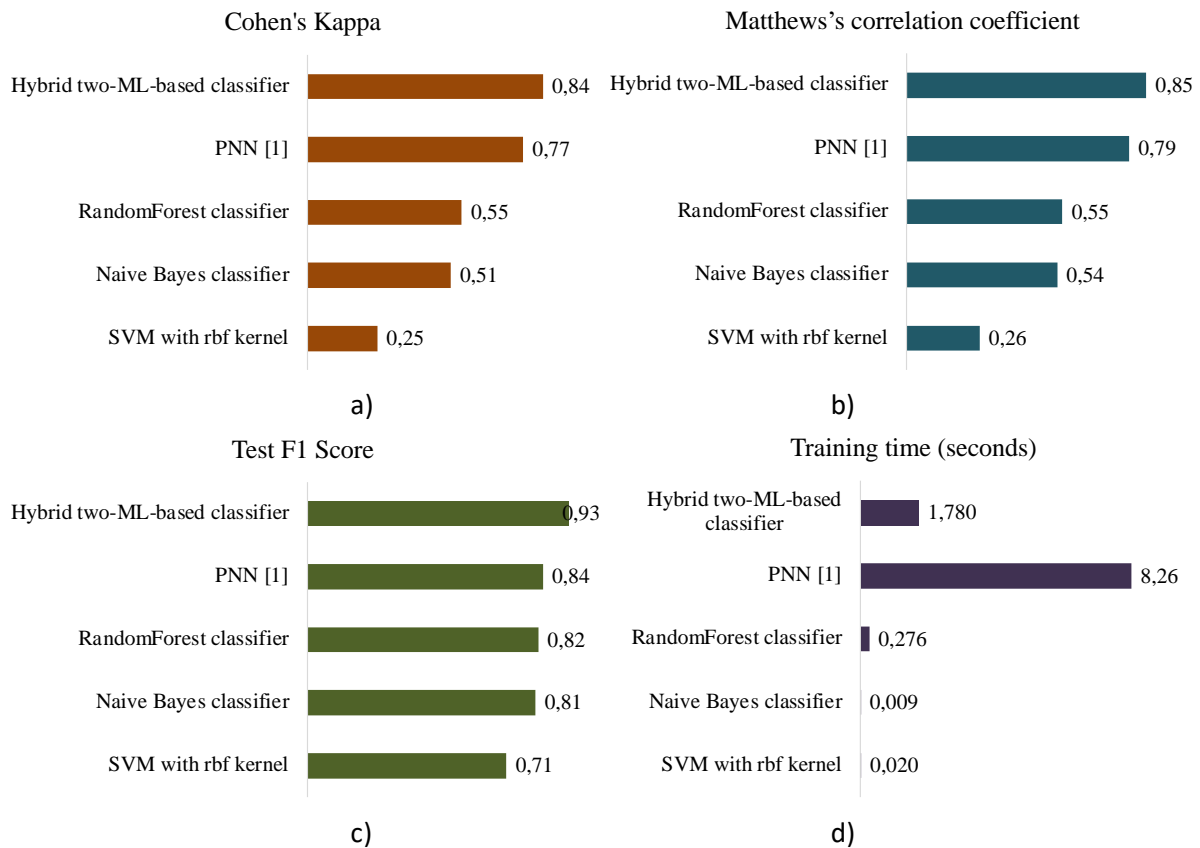


Figure 3: Comparison of all investigated methods based on: a) Cohen's Kappa; b) Matthews's correlation coefficient; c) f1-score; d) training time

The following conclusions can be drawn from the obtained results (Fig. 3):

- The nonlinear method, SVR with RBF kernel, even though it is quite often used during the analysis of short datasets, demonstrates the lowest accuracy of work using all three indicators;
- Both machine learning algorithms, which are the basis of the hybrid two-ML-based classifier, demonstrate somewhat higher accuracy of work compared to SVR;
- Acceptable results, taking into account the complexity of the task, are demonstrated by PNN. In particular, the accuracy of this neural network reaches 84% (f1-score). However, the time of its operation, taking into account the optimization by dual annealing for the optimal parameters selection of its operation, is quite significant.
- The developed method demonstrates the highest accuracy using all performance indicators when solving the task of predicting the survival of kidney transplants one month after transplantation. In addition, it shows high generalization properties (Table 1) and a satisfactory duration of the training procedure. All this ensures the possibility of its use when solving the real-world task of transplantation medicine.

6. Conclusions

The paper considers the problem of the prediction of the survival of kidney transplants one month after transplantation using machine learning tools. This approach will provide the doctor with additional information for timely and correct diagnosis and will allow for adjusting the treatment regimen.

The authors used a real-world dataset obtained from the Lviv Regional Clinical Hospital. Taking into account the imbalance of the dataset, the effectiveness of the application of various approaches to

balancing was investigated in the paper and the best one of them was determined for solving the stated problem.

The paper presents the new hybrid two-ML-based classifier, which is based on the sequential use of two machine learning algorithms. In the first step of the method, a Naive Bayes classifier is used, the results of which, in the form of a set of probabilities belonging to each of the data classes, replace all initial attributes with the found set of probabilities. In the second step of the method, the Random Forest algorithm uses a new dataset of significantly reduced dimensionality to predict the result.

Due to the comparison with several known approaches, the high accuracy of the method presented in this paper has been established, which provides high generalization properties and does not provoke overfitting. All this ensures the possibility of its use when solving the real problems of transplantation medicine. This approach can be also used in other areas [3], [22]–[24] as an Interpretable AI solution based on the use of the SGTm neural-like structure as a second element of the proposed cascade structure.

7. Acknowledgments

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

8. References

- [1] I. Izonin, et. al., ‘Addressing Medical Diagnostics Issues: Essential Aspects of the PNN-based Approach’, *CEUR-WS Proc. 3rd Int. Conf. Inform. Data-Driven Med. Växjö Swed. Novemb. 19 - 21 2020*, vol. 2753, pp. 209–218, 2020.
- [2] Y. Bodyanskiy, et. al., ‘Modified generalized neo-fuzzy system with combined online fast learning in medical diagnostic task for situations of information deficit’, *Math. Biosci. Eng.*, vol. 19, no. 8, pp. 8003–8018, 2022, doi: 10.3934/mbe.2022374.
- [3] S. Babichev, et. al., ‘Information Technology of Gene Expression Profiles Processing for Purpose of Gene Regulatory Networks Reconstruction’, in *2018 IEEE Second International Conference on Data Stream Mining Processing (DSMP)*, Aug. 2018, pp. 336–341. doi: 10.1109/DSMP.2018.8478452.
- [4] M. Medykovskyy, et. al., ‘Development of a regional energy efficiency control system on the basis of intelligent components’, in *2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT)*, Lviv, Ukraine: IEEE, Sep. 2016, pp. 18–20. doi: 10.1109/STC-CSIT.2016.7589858.
- [5] D. Chumachenko, et. al., ‘Predictive Model of Lyme Disease Epidemic Process Using Machine Learning Approach’, *Appl. Sci.*, vol. 12, no. 9, p. 4282, Apr. 2022, doi: 10.3390/app12094282.
- [6] Y. Tolstyak and M. Havryliuk, ‘An Assessment of the Transplant’s Survival Level for Recipients after Kidney Transplantations using Cox Proportional-Hazards Model’, *CEUR-WSorg*, vol. 3302, pp. 260–265, 2022.
- [7] Y. Tolstyak, et. al., ‘An investigation of the primary immunosuppressive therapy’s influence on kidney transplant survival at one month after transplantation’, *Transpl. Immunol.*, vol. 78, p. 101832, Jun. 2023, doi: 10.1016/j.trim.2023.101832.
- [8] M. Raynaud *et al.*, ‘Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence: an observational, international, multicohort study’, *Lancet Digit. Health*, vol. 3, no. 12, pp. e795–e805, Dec. 2021, doi: 10.1016/S2589-7500(21)00209-0.
- [9] S. A. A. Naqvi, K et. al., ‘Predicting Kidney Graft Survival Using Machine Learning Methods: Prediction Model Development and Feature Significance Analysis Study’, *J. Med. Internet Res.*, vol. 23, no. 8, p. e26843, Aug. 2021, doi: 10.2196/26843.
- [10] A. Altameem, *et al.*, ‘Patient’s data privacy protection in medical healthcare transmission services using back propagation learning’, *Comput. Electr. Eng.*, vol. 102, p. 108087, Sep. 2022, doi: 10.1016/j.compeleceng.2022.108087.

- [11] O. Basystiuk, *et al.*, ‘Machine Learning Methods and Tools for Facial Recognition Based on Multimodal Approach’, *Proc. Mod. Mach. Learn. Technol. Data Sci. Workshop MoMLLeTDS 2023 Lviv Ukr. June 3 2023*, vol. 3426, pp. 161–170.
- [12] L. Shahmoradi, *et al.*, ‘Predicting the survival of kidney transplantation: design and evaluation of a smartphone-based application’, *BMC Nephrol.*, vol. 23, no. 1, p. 219, Dec. 2022, doi: 10.1186/s12882-022-02841-4.
- [13] Y. Tolstyak *et al.*, ‘The Ensembles of Machine Learning Methods for Survival Predicting after Kidney Transplantation’, *Appl. Sci.*, vol. 11, no. 21, p. 10380, Nov. 2021, doi: 10.3390/app112110380.
- [14] V. Kotsovsky, *et al.*, ‘On the Size of Weights for Bithreshold Neurons and Networks’, in *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, LVIV, Ukraine: IEEE, Sep. 2021, pp. 13–16. doi: 10.1109/CSIT52700.2021.9648833.
- [15] L. Mochurad and R. Panto, ‘A Parallel Algorithm for the Detection of Eye Disease’, in *Advances in Intelligent Systems, Computer Science and Digital Economics IV*, vol. 158, Z. Hu, Y. Wang, and M. He, Eds., in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 158, Cham: Springer Nature Switzerland, 2023, pp. 111–125. doi: 10.1007/978-3-031-24475-9_10.
- [16] I. Krak, *et al.*, ‘Data Classification Based on the Features Reduction and Piecewise Linear Separation’, in *Intelligent Computing and Optimization*, vol. 1072, P. Vasant, I. Zelinka, and G.-W. Weber, Eds., in *Advances in Intelligent Systems and Computing*, vol. 1072, Cham: Springer International Publishing, 2020, pp. 282–289. doi: 10.1007/978-3-030-33585-4_28.
- [17] V. Kotsovsky and A. Batyuk, ‘Feed-forward Neural Network Classifiers with Bithreshold-like Activations’, in *2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, Ukraine: IEEE, Nov. 2022, pp. 9–12. doi: 10.1109/CSIT56902.2022.10000739.
- [18] A. Trostianchyn, Z. *et al.*, ‘Sm-Co alloys coercivity prediction using stacking heterogeneous ensemble model’, *Acta Metall. Slovaca*, vol. 27, no. 4, pp. 195–202, Dec. 2021, doi: 10.36547/ams.27.4.1173.
- [19] ‘sklearn.utils.class_weight.compute_class_weight’, scikit-learn. Accessed: Aug. 28, 2023. [Online]. Available: https://scikit-learn/stable/modules/generated/sklearn.utils.class_weight.compute_class_weight.html
- [20] ‘SMOTE — Version 0.11.0’. Accessed: Aug. 28, 2023. [Online]. Available: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html
- [21] G. I. Webb, E. *et al.*, ‘Naïve Bayes’, in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds., Boston, MA: Springer US, 2011, pp. 713–714. doi: 10.1007/978-0-387-30164-8_576.
- [22] I. Oleksiv, *et al.*, ‘Quality of Student Support at IT Educational Programmes: Case of Lviv Polytechnic National University’, in *2021 11th International Conference on Advanced Computer Information Technologies (ACIT)*, Deggendorf, Germany: IEEE, Sep. 2021, pp. 270–275. doi: 10.1109/ACIT52158.2021.9548648.
- [23] W. Auzinger, *et al.*, ‘A Continuous Model for States in CSMA/CA-Based Wireless Local Networks Derived from State Transition Diagrams’, in *Proceedings of International Conference on Data Science and Applications*, vol. 287, M. Saraswat, S. Roy, C. Chowdhury, and A. H. Gandomi, Eds., in *Lecture Notes in Networks and Systems*, vol. 287, Singapore: Springer Singapore, 2022, pp. 571–579. doi: 10.1007/978-981-16-5348-3_45.
- [24] J. Wasilczuk, *et al.*, ‘Entrepreneurial competencies and intentions among students of technical universities’, *Probl. Perspect. Manag.*, vol. 19, no. 3, pp. 10–21, Jul. 2021, doi: 10.21511/ppm.19(3).2021.02.