Investigation of PNN Optimization Methods to Improve Classification Performance in Transplantation Medicine

Myroslav Havryliuk[a], Nazarii Hovdysh[a], Yaroslav Tolstyak[b,c], Valentyna Chopyak[b] , Natalya Kustra[a]

[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 problem of predicting the success of organ transplantation is critical in the field of medicine. The use of a probabilistic neural network is of considerable interest in this context. In this study, the authors compared the speed of work of four popular methods for optimizing the parameter of a probabilistic neural network in the case of analyzing a short medical dataset collected by Lviv Regional Clinical Hospital. All three algorithms have demonstrated efficiency, reaching the optimum performance point. The use of optimizers provided a significant saving of time and computing resources compared to grid search.

Keywords [1]

Probabilistic neural network, Optimization, Small data, Classification

1. Introduction

The problem of predicting the success of organ transplantation is critical in the field of medicine. Currently, there are no models capable of accurately describing the patient's condition after transplantation. Therefore, the use of methods of intelligent data analysis has gained wide popularity.

However, insufficient data is often an obstacle to building adequate machine learning models. Classical models of artificial intelligence do not demonstrate sufficient efficiency in the case of processing small medical datasets. This is due to a number of reasons, the main of which is the problem of overfitting.

The use of a probabilistic neural network in such cases can improve performance compared to traditional models. However, the selection of the optimal network parameter by brute force method requires a lot of time and computing resources. That is why the use of optimization methods is appropriate for this task.

2. State-of-the-arts

New approaches to working with small datasets appear every year. However, this area of research still needs development.

The issue of using a probabilistic neural network for classification problems was analyzed in [1]. The authors found that the number of studies involving the application of probabilistic neural networks had increased over the previous five years. Research concerns various fields of medicine, such as nephrology, cardiology, oncology, pulmonology, endocrinology, neurosurgery, etc. Often use a combination of probabilistic neural network with other machine learning methods, such as SVM in [2], [3] and [4], Naive Bayes in [5], [6] and [7], K-means in [8], [9] and [10].

IDDM'2023: 6th International Conference on Informatics & Data-Driven Medicine, November, 17–19, 2023, Bratislava, Slovakia
EMAIL: myroslav.a.havryliuk@lpnu.ua (A. 1)
ORCID: 0000-0001-5259-7564

In the above-mentioned works, the selection of the optimal parameters of the neural network was carried out using a grid search. Thus, optimizing the parameters of a probabilistic neural network is relevant.

The purpose of this study is to compare the performance of three popular methods for optimizing the parameter of a probabilistic neural network in the case of analyzing a short medical dataset.

2.1. Probabilistic Neural Network

A probabilistic neural network is often used to solve a wide range of tasks, including classification [11]. The training procedure of this neural network is quite simple. The model also has certain disadvantages, the main one of which is the increase in dimensionality of the structure with the increase of the sample [12]. Accordingly, the use of a probabilistic neural network can require the allocation of a large amount of resources.

The work of this neural network in the case of binary classification can be described as follows:

1. Let there be *k* vectors of class 1 and *m* vectors of class 2 in the sample. We denote the j-th component of the i-th vector as $X_{i,j}^1$ for class 1 and as $X_{i,j}^2$ for class 2. The task of the model is to classify the input vector X. Therefore, it is necessary to determine the probability that the vector X belongs to class 1.
2. Canberra distances between the input vector and all sample vectors are calculated:

$$R_i^1 = \sum_{j=1}^{n} \frac{\left|X_{i,j}^1 - X_j\right|}{\left|X_{i,j}^1\right| + \left|X_j\right|},\ R_i^2 = \sum_{j=1}^{n} \frac{\left|X_{i,j}^2 - X_j\right|}{\left|X_{i,j}^2\right| + \left|X_j\right|} \quad (1)$$

3. Gaussian distances are calculated based on the obtained values:

$$D_i^{1,2} = \exp\left(-\frac{(R_i^{1,2})^2}{\sigma^2}\right) \quad (2)$$

4. The probability that the input vector belongs to class 1 is calculated by the formula:

$$P_1 = \frac{\sum_{i=1}^{k} D_i^1}{k} \quad (3)$$

5. Similarly for class 2:

$$P_2 = \frac{\sum_{i=1}^{k} D_i^2}{k} \quad (4)$$

6. So probabilistic neural network predicts a class of the input vector using the following rule:

$$y^{pred} = \begin{Bmatrix} 0, if\ \max\{P_c\} = P_1 \\ 1, if\ \max\{P_c\} = P_2 \end{Bmatrix}, c = 1,2 \quad (5)$$

2.2. Optimization problem formulation

For models that are built using unbalanced datasets, the F1 score is an appropriate measure of performance [13].

Let y_i, $i \in 1..N$ denote belonging to a certain class in the test sample of size *N*, then y_i^{pred}, $i \in 1..N$ is the value predicted by the model.

Precision of the model will be equal to:

$$Precision = \frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i^{pred}} \quad (6)$$

Accordingly, recall is equal to:

$$Recall = \frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i} \quad (7)$$

According to the definition of the F1-score metric, it can be expressed as follows:

$$f1\_score = 2 * \frac{\frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i^{pred}} * \frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i}}{\frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i^{pred}} + \frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i}} \quad (8)$$

Thus, the problem of maximizing F1-score can be presented in the following form:

$$2 * \frac{\frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i^{pred}} * \frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i}}{\frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i^{pred}} + \frac{\sum_{i=1}^{N}\left(y_i^{pred} * y_i\right)}{\sum_{i=1}^{N} y_i}} \rightarrow max \quad (9)$$

with the restrictions 0.001>sigma>10.

2.3. Methods for solving the optimization task

We applied three popular optimization algorithms:

- Bayesian optimization
- Differential evolution
- Dual annealing

These methods do not require the calculation of derivatives and can perform optimization in case the objective function is a “black box” [14].

Bayesian optimization uses Gaussian process to model the black-box objective function [15].

We defined the upper confidence bounds function as acquisition function to balance exploitation and exploration. Also we used the following optimization parameters:

- number of initial points – 5;
- number of iterations – 10.

Differential evolution is a stochastic method. It applies the key concepts of genetic algorithms [16]. The first step of the algorithm is to create a generation of candidates that are the objective function arguments. At each iteration, a new generation is created by mixing with other candidates.

We applied a "best1bin" strategy for creating trial candidates. According to it:

- the difference between two randomly chosen candidates is used to provide a mutation of the best member of the population;
- a binomial distribution is applied for recombination.

We defined following key algorithm parameters:

- population size – 10;
- mutation – [0.5;1);
- recombination – 0.7;
- maximum number of generations – 10.

Dual annealing is also a stochastic approach. It combines the generalization of Fast Simulated Annealing and Classical Simulated Annealing coupled to a strategy for carrying out a local search on accepted locations [17]. This approach describes an advanced method to improve the solution that was found by the generalized annealing process. A distorted Cauchy-Lorentz visiting distribution is used in this optimization algorithm.

We used the following optimization parameters:

- parameter for visiting distribution – 2.62;
- parameter for acceptance distribution – -5.0;
- maximum number of global search iterations – 10.

For all algorithms, the optimization was performed on the interval $\sigma \in [0.00001;10]$.

3. Modeling and results

3.1. Dataset descriptions

The imbalanced dataset collected by Lviv Regional Clinical Hospital (Department Hospital Nephrology and Dialysis) was used in this study. It contains data on 164 patients who received HLA-matched renal allografts between 1992 and 2020 by 42 attributes (such as age, sex, glucose level, etc.). Among them, 64 (42.1%) were women and 88 (57.9%) were men. The age of the patients at the time of transplantation was 32.6 ± 8.7 (in the range of 18–60) years. 152 patients were transplanted for the first time, 12 (5 women and 7 men) were transplanted again.

3.2. Results

Three optimization algorithms described above were used to optimize the parameter. The implementation of optimizers from the scipy.optimize and bayesian optimization libraries of the Python programming language was used. The optimization results are shown in Table 1.

Table 1
Optimization results

Optimizer	Accuracy	Precision	Recall	F1 score	Number of evaluations of F1-score	σ	Time, sec
Differential evolution	0.896	0.727	1	0,842	62	7.232	0.377
Bayesian	0.896	0.727	1	0,842	15	5.472	1.624
Dual annealing	0.896	0.727	1	0,842	29	6.678	0.189

4. Comparison and discussion

As can be seen, all three optimizers have reached the point from the intervals where the value of F1-score is maximal. The precision value indicates a quite large proportion of false-positive results, while the recall is 100%.

All algorithms showed quite good optimization speed. The shortest execution time was demonstrated by dual annealing. A visualization of the optimization duration can be seen at Fig. 1.

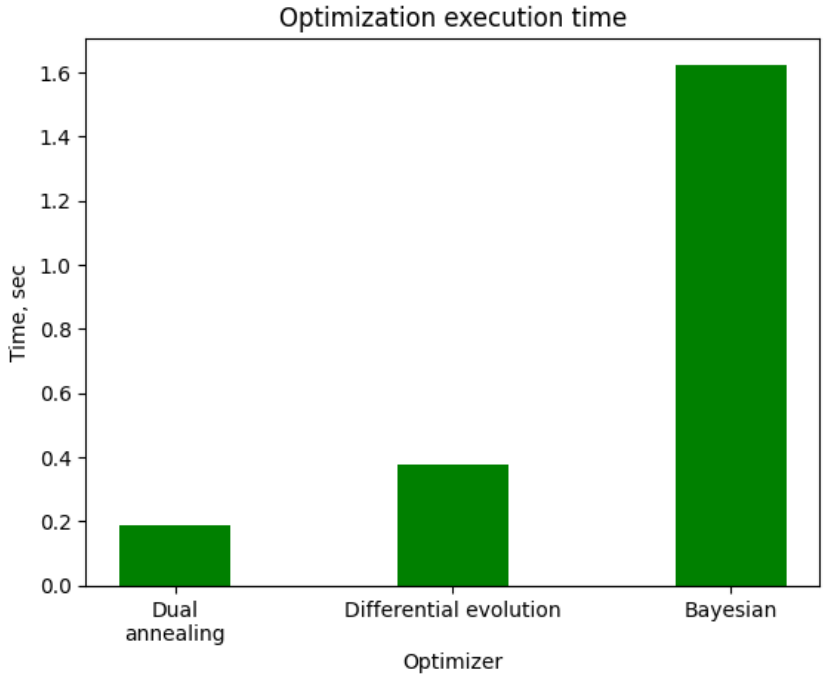

Figure 1: Optimization execution time

In terms of the number of evaluations of the objective function, Bayesian optimization is the most effective (a visualization can be seen at Fig. 2). However, other steps of this algorithm also cause computational costs, which is reflected in the duration of execution.

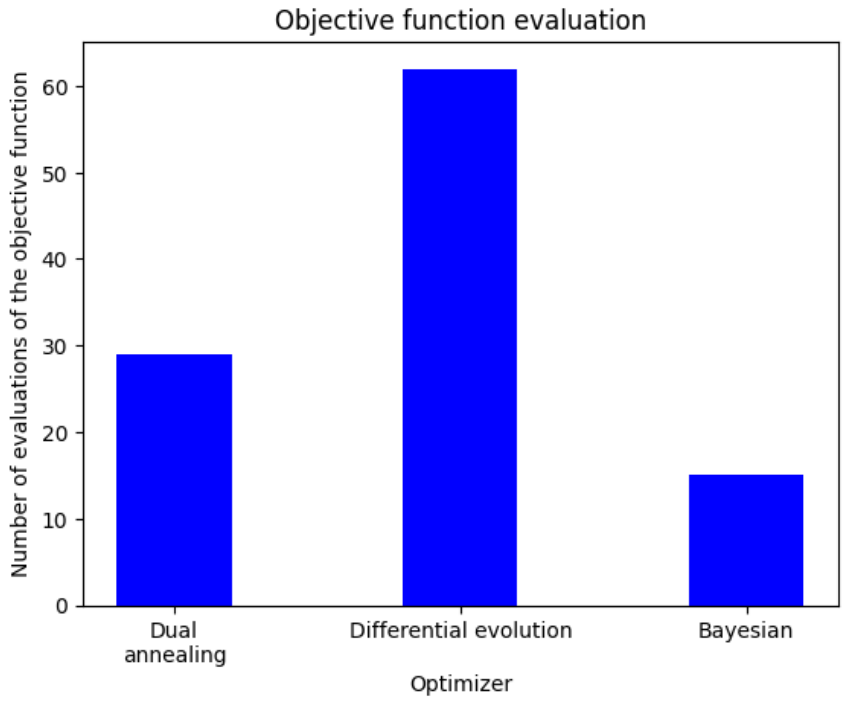

Figure 2: Number of evaluations of F1 score

The selection of the optimal value of the parameter was also carried out using a grid search on the interval $\sigma \in [0.00001;10]$ with a step $\Delta=0.001$. The execution time was 41.852 seconds. The number of objective function calculations was 10000. As a result of the experiment, it was found that there are two intervals on which F1 score reaches a maximum of 0.842 (Fig. 3).

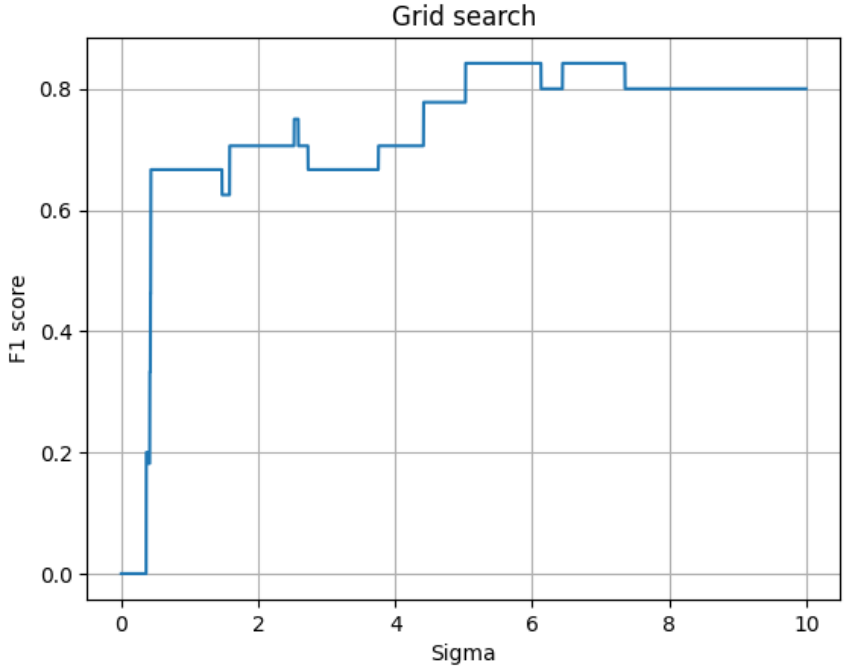

Figure 3: Grid search optimization results

So the use of each of the optimizers provides a significant reduction in execution time and computational costs, compared to the grid search.

5. Conclusions

The problem of predicting the success of organ transplantation is critical in the field of medicine. The use of a probabilistic neural network is of considerable interest in this context. In this study, the authors compared the performance of three popular methods for optimizing the parameters of a probabilistic neural network in the case of analyzing a short set of medical data collected by Lviv Regional Clinical Hospital. All three algorithms have demonstrated efficiency, reaching the optimum performance point. The use of optimizers provided a significant saving of time and computing resources compared to a grid search.

Further research may concern the optimization of model parameters, where the probabilistic neural network is used in combination with other machine learning methods.

6. Acknowledgments

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

7. References

[1] Izonin I, et al. Addressing Medical Diagnostics Issues: Essential Aspects of the PNN-based Approach. In: Proceedings of the 3rd International Conference on Informatics & Data-Driven Medicine. Växjö, Sweden, November 19 - 21, 2020

[2] Izonin I, et al. PNN-SVM Approach of Ti-Based Powder’s Properties Evaluation for Biomedical Implants Production. Comput Mater Contin. 2022;71(3):5933–47.

[3] Mochurad L, et al. Classification of X-Ray Images of the Chest Using Convolutional Neural Networks. In: Proceedings of the 4th International Conference on Informatics & Data-Driven Medicine. Valencia, Spain, November 19 - 21, 2021. 269-282.

[4] Tolstyak Y, Havryliuk M. An Assessment of the Transplant's Survival Level for Recipients after Kidney Transplantations using Cox Proportional-Hazards Model. In: Proceedings of the 5th International Conference on Informatics & Data-Driven Medicine. Lyon, France, November 18 - 20, CEUR-WS.org, 2022. pp. 260-265.

[5] Tolstyak Y, et al. The Ensembles of Machine Learning Methods for Survival Predicting after Kidney Transplantation. Appl Sci. 2021 Jan;11(21):10380.

[6] Liaskovska S, et al. Investigation of Anomalous Situations in the Machine-Building Industry Using Phase Trajectories Method. In: Hu Z, Petoukhov S, Yanovsky F, He M, editors. Advances in Computer Science for Engineering and Manufacturing. Cham: Springer International Publishing; 2022. p. 49–59. (Lecture Notes in Networks and Systems).

[7] Basystiuk O, Melnykova N. Multimodal Approaches for Natural Language Processing in Medical Data. In: Proceedings of the 5th International Conference on Informatics & Data-Driven Medicine. Lyon, France, November 18–20, CEUR-WS.org, pp. 246–252 (2022).

[8] Tolstyak Y, et al. An investigation of the primary immunosuppressive therapy’s influence on kidney transplant survival at one month after transplantation. Transpl Immunol. 2023 Jun 1;78:101832.

[9] Kotsovsky V, 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). 2021. p. 13–6.

[10] Mochurad LI. Canny Edge Detection Analysis Based on Parallel Algorithm, Constructed Complexity Scale and CUDA. Comput Inform. 2022 Nov 9;41(4):957–80.

[11] Kotsovsky V, Batyuk A. Feed-forward Neural Network Classifiers with Bithreshold-like Activations. In: 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT). 2022. p. 9–12.

[12] Oleksiv I, 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). 2021. p. 270–5.
[13] Martyn Y, et al. Optimization of Technological's Processes Industry 4.0 Parameters for Details Manufacturing via Stamping: Rules of Queuing Systems. Procedia Comput Sci. 2021 Jan 1;191:290–5.
[14] Ganguli C, et al. Adaptive Artificial Bee Colony Algorithm for Nature-Inspired Cyber Defense. Systems. 2023 Jan;11(1):27.
[15] Ljaskovska S, et al. Optimization of Parameters of Technological Processes Means of the FlexSim Simulation Simulation Program. In: 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP). 2020. p. 391–7.
[16] Mochurad L. Optimization of Regression Analysis by Conducting Parallel Calculations. In: COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine, 982-996 p.
[17] Basystiuk O, et al. Machine Learning Methods and Tools for Facial Recognition Based on Multimodal Approach. In: Proceedings of the Modern Machine Learning Technologies and Data Science Workshop (MoMLeT&DS 2023). Lviv, Ukraine, June 3, 2023. CEUR-WS.org, pp. 161-170 (2023).