Preprints, 4th IFAC Conference on
Advances in Proportional-Integral-Derivative Control
Almería, Spain | June 12-14, 2024

Multi-objective Machine Learning for control performance assessment in PID control loops ⋆

Gilberto Reynoso-Meza [*,**] Jesús Carrillo-Ahumada [***]
Tainara Marques [*]

[*] *Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifícia Universidade Católica do Paraná (PUCPR), Rua Imaculada Conceição 1155, Curitiba, Brazil. (e-mail: g.reynosomeza@pucpr.br).*
[**] *Control Systems Optimization Laboratory (LOSC), Pontifícia Universidade Católica do Paraná (PUCPR).*
[***] *Ingeniería en Alimentos. Universidad del Papaloapan. Circuito Central 200, Col. Parque Industrial, 68301 Tuxtepec, Oaxaca, Mex (e-mail: jcarrillo@unpa.edu.mx)*

Abstract: Control Performance Assessment (CPA) is a critical endeavor in industrial processes, ensuring optimal functioning of control systems. Traditionally, CPA has been addressed through solutions using some control performance indicators. Nowadays, the integration of data science and machine learning has emerged as a viable alternative, particularly in classification tasks related to CPA. That is, in a binary classification scheme, the goal is to predict whether incoming data from the control loop belongs to class 0 or 1, representing the absence or presence of an anomaly (performance degradation). In such a case, a trade-off between false positives and false negatives should be obtained, via the training phase of a given supervised machine learning structure for example. Usually, this is a conflicting trade-off, where multi-objective optimization techniques in the training phase of such learners could bring interesting results. In this paper, we explore the usability of multi-objective optimization training in machine learning, for control performance assessment classification. A database describing 30 control performance indicators (features) in a PID control loop is used. The obtained results indicate that the proposed approach could bring interesting applications to improve the performance of CPA classification systems.

Keywords: Control performance assessment; machine learning; multi-objective optimization; PID control; multi-criteria analysis.

1. INTRODUCTION

Control Performance Assessment (CPA) in the industry is a crucial challenge involving the evaluation of the effectiveness of control systems. The need to measure and enhance control performance in industrial settings has been a persistent issue due to its direct impact on process quality and operational efficiency (Jelali, 2006).

Traditionally, CPA has been addressed through solutions that might face significant challenges. Conventional methodologies often focus on specific metrics to assess control performance but might find difficulties in dealing with the complexity and variability of modern industrial systems. In response to these challenges, data science and machine learning have emerged as promising alternatives. Their ability to process large datasets and extract complex patterns offers new insights for addressing CPA more efficiently and effectively (Grelewicz et al., 2023a,b).

For example, the goal of classification tasks within the machine learning context is to minimize misclassifications. Nevertheless, misclassifications can be tagged as false positives or false negatives; minimizing both simultaneously can be seen as a multi-objective problem with conflicting objectives. To overcome such an instance, multi-objective optimization techniques have proven valuable in dealing with conflicting objectives. This idea could be effectively incorporated into the training phase of machine learning, enabling a more holistic assessment of control performance.

In this context, we propose a multi-objective training approach to address the CPA problem. We present an application of this approach on a dataset, describing 30 control performance indicators (features) in a PID control loop. The remainder of this paper is as follows: in Section 2, a brief theoretical background is presented; in Section 3, methods, tools, and experimental framework for this paper are described; in Section 4, results are discussed; and in Section 5, some conclusions and future work are given.

2. THEORETICAL BACKGROUND

Next, theoretical background on CPA, machine learning, and multi-objective optimization are discussed.

⋆ This study was financed in part by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) via the projects PQ-2/310195/2022-5 & Universal/408164/2021-2.

2.1 Control performance assessment

Control Performance Assessment is a crucial aspect in evaluating the effectiveness of control systems, ensuring their optimal functioning within industrial processes (Jelali, 2012; Domański et al., 2020). It involves the quantitative analysis of how well a control system meets the desired performance criteria. Mathematically, CPA is often expressed through performance indices and metrics, such as the Integral of Time-weighted Absolute Error (ITAE), Integral of Squared Error (ISE), or other relevant measures such as the Harris index or minimum variance. These metrics provide a quantitative basis for assessing the system's ability to maintain stability, reduce oscillations, and respond accurately to setpoint changes or disturbances. Several works propose improvements or variations of such indexes (Horch and Isaksson, 1999; Khosroshahi et al., 2022). In any case, the formalism of CPA aims to capture the dynamic behavior of the controlled system, offering insights into its overall efficiency and robustness.

2.2 Machine learning, supervised learning, and binary classification

Machine learning refers to computer algorithms that enhance their capabilities autonomously through exposure to data. This learning process can be supervised, unsupervised, or via reinforcement (Mitchell et al., 1997). In supervised learning, instances ($\boldsymbol{I}$) with features ($\boldsymbol{X}$) are used to train a learner, employing reliable target information ($\boldsymbol{T}$) for each instance. The primary objective is to establish a relationship that yields an output (target prediction) for new instances.

In binary classification, the task is to predict whether incoming data belongs to class 0 (C_0) or class 1 (C_1). For anomaly detection, these classes signify the presence or absence of an anomaly. Class 0 denotes a situation without anomalies, while class 1 indicates an anomalous situation. The learner undergoes a training phase using a dataset to adjust its parameters $\boldsymbol{\beta}$ through optimization, utilizing an evaluation criteria or cost function. That is, given a loss function statement $J(\boldsymbol{\beta})$ to minimize misclassifications for a given hypothesis $h_\beta(\boldsymbol{x})$ (learner structure) based on the tuneable $\boldsymbol{\beta}$ parameters for a data vector $\boldsymbol{x}$.

Several machine learning structures have been proposed over the years. To name few basic approaches: k-Nearest Neighbors (Cunningham and Delany, 2021), logistic regression (Mitchell et al., 1997), Naive Bayes (Bielza and Larranaga, 2014), artificial neural networks (Meireles et al., 2003), support vector machines (Cortes and Vapnik, 1995), decision trees (Hernández et al., 2021), ensembles (Gomes et al., 2017). Each one of them provides a particular structure that must be trained to adjust its tuneable parameters $\boldsymbol{\beta}$. Several optimization tools and techniques are used for such a training stage; among them, multi-objective optimization has brought interesting results.

2.3 Multi-objective optimization

As outlined in Miettinen (1998), a multi-objective problem (MOP) with m objectives [1], can be formulated as follows:

$$\min_{\boldsymbol{\theta}} \boldsymbol{J}(\boldsymbol{\theta}) = [J_1(\boldsymbol{\theta}), \ldots, J_m(\boldsymbol{\theta})] \tag{1}$$

subject to:

$$\boldsymbol{K}(\boldsymbol{\theta}) \leq 0 \tag{2}$$
$$\boldsymbol{L}(\boldsymbol{\theta}) = 0 \tag{3}$$
$$\underline{\theta_i} \leq \theta_i \leq \overline{\theta_i}, i = [1, \ldots, n] \tag{4}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]$ is defined as the decision vector with $\dim(\boldsymbol{\theta}) = n$; $\boldsymbol{J}(\boldsymbol{\theta})$ as the objective vector and $\boldsymbol{K}(\boldsymbol{\theta})$, $\boldsymbol{L}(\boldsymbol{\theta})$ as the inequality and equality constraint vectors respectively; $\underline{\theta_i}, \overline{\theta_i}$ are the lower and the upper bounds in the decision space.

It is important to note that in MOPs, there is not a single solution due to the absence of a universally superior solution across all objectives. Instead, a set of solutions, the Pareto set $\boldsymbol{\Theta_P}$, is defined. Each solution in the Pareto set corresponds to an objective vector in the Pareto front $\boldsymbol{J_P}$. Typically, the focus is approximating a Pareto front and set, $\boldsymbol{J_P^*}$,$\boldsymbol{\Theta_P^*}$. All solutions on the Pareto front represent a set of Pareto optimal and non-dominated solutions.

- **Pareto optimality**: An objective vector $J(\theta^1)$ is Pareto optimal if there is no other objective $J(\theta^2)$, such that:

$$\begin{aligned} iff \quad & J_i(\theta^2) \leq J_i(\theta^1) \; \forall \; i \in [1, 2, ..., n] \quad and \\ & J_j(\theta^2) < J_j(\theta^1) \; \exists \; j \in [1, 2, ..., n] \end{aligned} \tag{5}$$

- **Dominance**: Given two objective vectors $J(\theta^1)$, $J(\theta^2)$, the objective vector $J(\theta^1)$ is dominated by the objective vector $J(\theta^2)$ iff:

$$\begin{aligned} iff \quad & J_i(\theta^2) \leq J_i(\theta^1) \; \forall \; i \in [1, 2, ..., n] \quad and \\ & J_j(\theta^2) < J_j(\theta^1) \; \exists \; j \in [1, 2, ..., n] \end{aligned} \tag{6}$$

  This is denoted as $J(\theta^2) \preceq J(\theta^1)$.

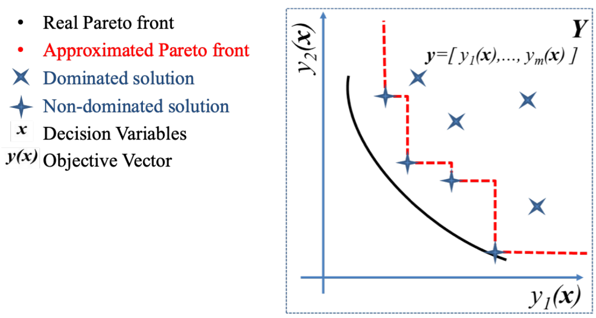

Figure 1. Pareto optimality and dominance concepts for design objective $y_1(\boldsymbol{x}), y_2(\boldsymbol{x})$ for decision variables $\boldsymbol{x}$.

[1] A maximization problem can be converted to a minimization problem. For each of the objectives that have to be maximized, the transformation: $\max J_i(\boldsymbol{\theta}) = -min(-J_i(\boldsymbol{\theta}))$ could be applied.

2.4 Related Works and Contextual Framework

On the one hand, it is not new to deal with the CPA problem from the machine learning perspective. For example, in Grelewicz et al. (2023a,b) a novel machine learning derived CPA classification system is proposed and implemented for a wide class of PID-based control industrial loops of reduced order. On the other hand, the idea of using multi-objective optimization statements in machine learning applications has been largely analyzed for machine learning ensembles (Onan et al., 2016; Ribeiro and Reynoso-Meza, 2020), dynamic ensemble selection (Ribeiro et al., 2020), feature selection and classification(Ribeiro and Reynoso-Meza, 2021), and for the learner structure training itself (Ryu and Baik, 2016; Reynoso-Meza et al., 2022; Nag and Pal, 2015). In spite of the promising results of using multi-objective optimization techniques in machine learning training, to the best of our knowledge, such an idea has yet to be explored for CPA. Therefore, we move forward in this direction, addressing the following questions:

Q1 Could machine learning techniques benefit from multi-objective training for supervised CPA classification systems?
Q2 If so, which structures benefit the most from this approach?

3. EXPERIMENTAL FRAMEWORK, TOOLS AND METHODS

A supervised learning approach using multi-objective optimization training is proposed to answer the above-commented questions.

3.1 Learner structures under consideration

In the same way that PID control, a common, simple, and reliable technique for automation, can benefit from a multi-objective optimization tuning approach, we intend to verify if simple machine learning structures could benefit in the same way for the CPA problem; future work will focus on more complex techniques. Therefore, two basic structures will be posed in this work: Logistic Regression (7) and Naive Bayes (8).

$$h_\beta(\boldsymbol{x}) = \frac{1}{1 + \exp^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}} \tag{7}$$

$$h(\boldsymbol{x}) = p(C_k|\boldsymbol{x}) = \frac{1}{Z} p(C_0) \prod_{i=1}^{N} p(x_i|C_k) \tag{8}$$

Naive Bayes (8), in particular, is known for assuming that the features are conditionally independent, given the target class C_k. It is a probabilistic model and, therefore, without tuneable $\boldsymbol{\beta}$ parameters *a priori.*

3.2 Data set description

The data set proposed by Grelewicz et al. (2023a) will be used. It consists of a data set of 30 control performance indicators computed from the rejection response when a disturbance step change is applied in a PID control loop. A large and representative set of different second-order plus deadtime processes was defined to build the database. For each process, a PID tuning was derived via an optimization problem, and several controllers were proposed in the proximities of them via random perturbation of the PID parameters. The labeling process was performed in the following way: if the disturbance rejection response was similar to the reference one, the instance is labeled as OK; otherwise, as *NOK*. This similarity was calculated in terms of gain margin, phase margin, and normalized error. The training dataset consists of 60,000 instances, while the test database has 10,000 instances, both of them with equally balanced classes.

3.3 Multi-objective problem statement

Instead of using an aggregation function for FP (false positives) and FN (false negatives), classification performance will be evaluated simultaneously via multi-criteria analysis. Therefore, a multi-objective problem is considered as shown in Equation (9) and previously proposed in Reynoso-Meza et al. (2022):

$$\min_{\boldsymbol{\beta}} \boldsymbol{J}(\boldsymbol{\beta}) = [FP + CEm(\boldsymbol{\beta}), FN + CEm(\boldsymbol{\beta})] \tag{9}$$

With:

$$CEm(\boldsymbol{\beta}) = -\frac{[y \log(h_\beta(\boldsymbol{x})) + (1-y)\log(1 - h_\beta(\boldsymbol{x}))]}{M \cdot \log(\epsilon)} - 1 \tag{10}$$

subject to:

$$\underline{\beta_i} \le \beta_i \le \overline{\beta_i}, i = [0, \ldots, n] \tag{11}$$

While it is straightforward for the Logistic regression case to use the above-commented optimization statement, it is not the case for Naive Bayes. Therefore, here we use the modified structure of equation (12), where $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_N]$ are the weighting coefficients, that are adjusted given M observations or instances and $\beta_0 = p(C_0)$as proposed by Ryu and Baik (2016).

$$p(C_k|\boldsymbol{x}) = \frac{1}{Z}\beta_0 \prod_{i=1}^{N} p(x_i|C_k)^{\beta_i} \tag{12}$$

3.4 Multi-objective optimization process

As an initial population, a set of learners trained with the classical approach is calculated. Such a set of learners are trained using a different number of features, using the Minimum Redundancy Maximum Relevance (mRMR) criteria (Radovic et al., 2017). That is, a total of 30 learners are adjusted, ranging from 1 to 30 features, according to the mRMR criteria.

For the experiments presented here, the spMODEx [2] algorithm will be used. It is a multi-objective evolutionary algorithm based on Differential Evolution (Storn and Price, 1997; Pant et al., 2020), using as a diversity mechanism a spherical pruning (Reynoso-Meza et al., 2014). The

[2] https://www.mathworks.com/matlabcentral/fileexchange/65145

following hyper-parameters are used: binomial mutation; scaling factor $F = 0.5$; crossover rate $Cr = 0.9$; population $Pop = 50$; function evaluations $FEs = 2e4$ and 100 arcs in the spherical grid.

For pertinency improvement, the physical programming matrix of Table 1 is used, where $\boldsymbol{\beta_0}$ is the baseline solution using all available features for the learner structure via classical training phase (*i.e.* single objective criteria). Pertinency (Reynoso-Meza et al., 2014) is a well-documented technique useful to approximate a Pareto front with useful solutions. In this case, just learners dominating the baseline $\boldsymbol{\beta_0}$ are allowed.

3.5 Multi-criteria Analysis and performance measures

As two design objectives will be discussed, a simple 2D visualization is enough to interpret the approximated Pareto fronts. With the Dominance criteria, it will be possible to analyze the impact of Multi-objective optimization training when compared with the initial population and the baseline learner.

To evaluate the performance of a given binary classifier, the confusion matrix could provide valuable information in terms of false positives and false negatives. Nevertheless, as we are approximating a Pareto front with several learners, it will be impractical to include a confusion matrix for each of them. Therefore, we propose a different visualization approach, named Pareto Confusion Matrix, based on the classical array of a Confusion matrix, but with violin plots [3] to depict the distribution of false positives and negatives of the approximated Pareto front.

4. RESULTS AND DISCUSSION

Tests were performed in a Desktop DELL precision 3561, Intel(R) 11th. generation i7-11800H, 2.30 GHz, and 16GB RAM running Matlab© R2021b. In Figures 2 and 3, results from the training phase for the Logistic regression and the Naive Bayes are depicted. Red diamonds represent the approximated Pareto front, blue squares the initial population, and the black hex star is the baseline solution. In the first case, it is possible to appreciate that the approximated Pareto front dominates the baseline solution and also the initial population. This shows the potential [4] viability of using the multi-objective training to adjust the logistic regression structure. In particular, it is possible to perceive that the logistic regression structure benefits from using all available features, given that the baseline solution also dominates several learners from the initial population. In the second case, contrasting with Logistic regression, the naive Bayes structure benefits the most from using fewer features; this can be observed given that several members of the initial population (using less than 30 features) dominate the baseline solution. One such solution is almost Pareto-optimal; nevertheless, the approximated Pareto front can dominate all but three solutions from the initial population; this was mainly in part because they are outside of the boundary box defined by the baseline solution. Therefore, we can conclude that the

[3] `https://github.com/bastibe/Violinplot-Matlab.git`

[4] We say potential, because we need to discard an over-fitting, which will be discussed next.

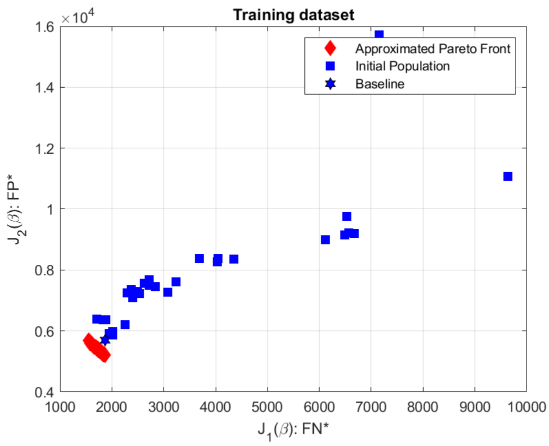

Figure 2. Approximated Pareto set performance on the training data set for Logistic Regression.

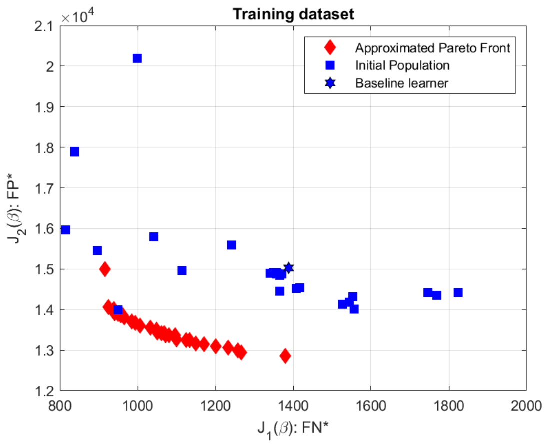

Figure 3. Approximated Pareto set performance on the training data set for Naive Bayes.

Naive Bayes also benefit from the multi-objective training. Figures 4 and 5 depict the performance of the Pareto set approximation with test data. As we can observe that most of the trade-off relations are preserved, we can conclude that we do not have over-fitted learners; besides, as we can observe that solutions from the initial population are still being dominated, the performance improvement of the multi-objective training is validated and thus, answering Q1.

In Figures 6 and 7, we can appreciate the Pareto Confusion matrix for each learning structure, respectively, and its comparison with the baseline learner. In the case of the Logistic regression, it is possible to appreciate that the main advantage of this approach relies on the minimization of the false negatives. In the case of the Naive Bayes, the benefits of this approach are the minimization of false positives and negatives. Finally, we can perform a learner structure comparison (or design concept comparison as defined by Mattson and Messac (2005)) in Figure 8. It is possible to notice that Naive Bayes with the multi-objective training approach minimizes false negatives in exchange for false positives. This could also guide the structure selection for a given problem, thus answering Q2.

Table 1. Preference matrix $\mathfrak{m}$ for MOP statement. Five preference ranges have been defined: highly desirable (HD), desirable (D), tolerable (T), undesirable (U), and highly undesirable (HU).

Preference Matrix						
	← HD →	← D →	← T →	← U →	← HU →	
Objective	J_i^0	J_i^1	J_i^2	J_i^3	J_i^4	J_i^5
$J_1(\boldsymbol{\beta})$	$0.4 * J_1(\boldsymbol{\beta_0})$	$0.6 * J_1(\boldsymbol{\beta_0})$	$0.8 * J_1(\boldsymbol{\beta_0})$	$1.0 * J_1(\boldsymbol{\beta_0})$	$2.0 * J_1(\boldsymbol{\beta_0})$	$5.0 * J_1(\boldsymbol{\beta_0})$
$J_2(\boldsymbol{\beta})$	$0.4 * J_2(\boldsymbol{\beta_0})$	$0.6 * J_2(\boldsymbol{\beta_0})$	$0.8 * J_2(\boldsymbol{\beta_0})$	$1.0 * J_2(\boldsymbol{\beta_0})$	$2.0 * J_2(\boldsymbol{\beta_0})$	$5.0 * J_2(\boldsymbol{\beta_0})$

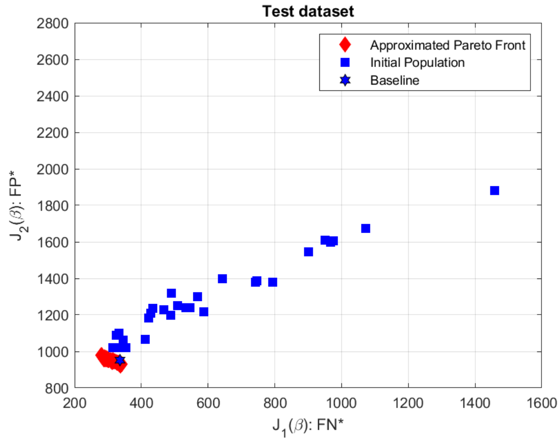

Figure 4. Approximated Pareto set performance on the test data set for Logistic Regression.

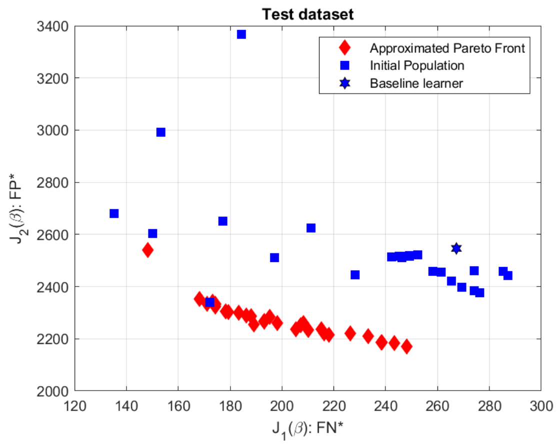

Figure 5. Approximated Pareto set performance on the test data set for Naive Bayes.

5. CONCLUSIONS

In this paper, a multi-objective optimization training process for machine learning structures was presented to tackle the CPA classification problem. Two questions were posed and answered by the experimental results. Regarding the first question, learning structures such as Logistic Regression and Naive Bayes for this problem benefit from multi-objective optimization training. Furthermore, via the Pareto Confusion Matrix visualization approach, it is possible to compare the classification performance between structures, helping to decide which structure to implement and answering the second question. Future work will focus

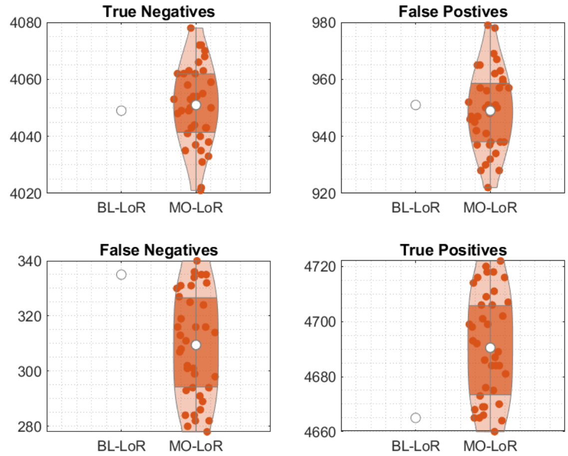

Figure 6. Pareto Confusion matrix of Logistic regression.

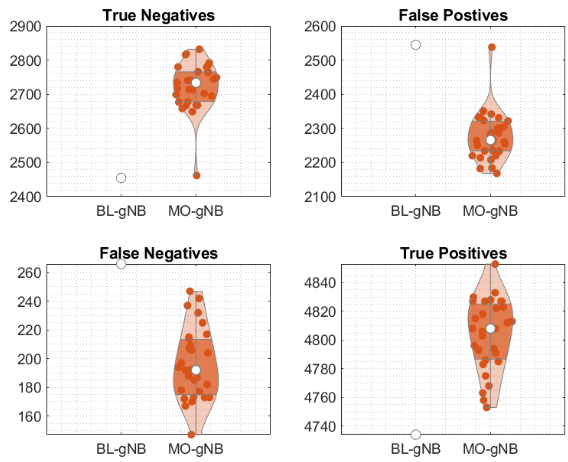

Figure 7. Pareto Confusion matrix of Naive Bayes.

on using this approach for different and more complex structures.

ACKNOWLEDGMENTS

We want to acknowledge the kindness and support of Patryk Grelewicz, Jacek Czeczot, et al. for allowing us to use their data set presented in Grelewicz et al. (2023a,b).

REFERENCES

Bielza, C. and Larranaga, P. (2014). Discrete bayesian network classifiers: A survey. *ACM Computing Surveys*

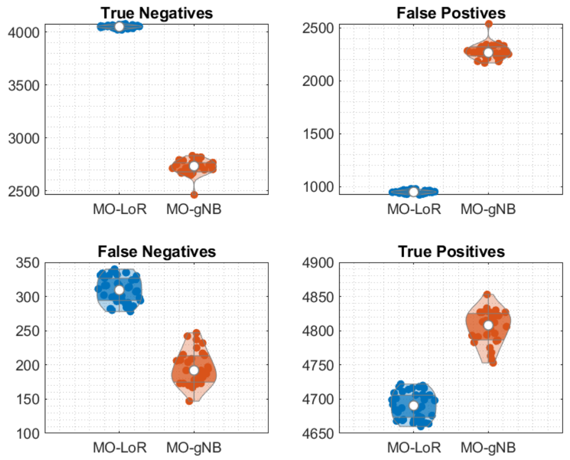

Figure 8. Pareto Confusion matrix comparison of Naive Bayes and Logistic Regression.

(CSUR), 47(1), 1–43.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273–297.

Cunningham, P. and Delany, S.J. (2021). k-nearest neighbour classifiers-a tutorial. *ACM computing surveys (CSUR)*, 54(6), 1–25.

Domański, P.D. et al. (2020). *Control Performance Assessment: Theoretical Analyses and Industrial Practice*, volume 245. Springer.

Gomes, H.M., Barddal, J.P., Enembreck, F., and Bifet, A. (2017). A survey on ensemble learning for data stream classification. *ACM Computing Surveys (CSUR)*, 50(2), 1–36.

Grelewicz, P., Khuat, T.T., Czeczot, J., Nowak, P., Klopot, T., and Gabrys, B. (2023a). Application of machine learning to performance assessment for a class of pid-based control systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

Grelewicz, P., Nowak, P., Khuat, T.T., Czeczot, J., Klopot, T., and Gabrys, B. (2023b). Practical implementation of computationally-efficient machine learning-based control performance assessment system for a class of closed loop systems. *Applied Soft Computing*, 146, 110690.

Hernández, V.A.S., Monroy, R., Medina-Pérez, M.A., Loyola-Gonzalez, O., and Herrera, F. (2021). A practical tutorial for decision tree induction: Evaluation measures for candidate splits and opportunities. *ACM Computing Surveys (CSUR)*, 54(1), 1–38.

Horch, A. and Isaksson, A.J. (1999). A modified index for control performance assessment. *Journal of Process Control*, 9(6), 475–483.

Jelali, M. (2006). An overview of control performance assessment technology and industrial applications. *Control engineering practice*, 14(5), 441–466.

Jelali, M. (2012). Control performance management in industrial automation: assessment, diagnosis and improvement of control loop performance.

Khosroshahi, M., Poshtan, J., and Alipouri, Y. (2022). Practical control performance assessment method using hurst exponents and crossover phenomena. *Computers & Chemical Engineering*, 161, 107774.

Mattson, C.A. and Messac, A. (2005). Pareto frontier based concept selection under uncertainty, with visualization. *Optimization and Engineering*, 6(1), 85–115.

Meireles, M.R., Almeida, P.E., and Simões, M.G. (2003). A comprehensive review for industrial applicability of artificial neural networks. *IEEE transactions on industrial electronics*, 50(3), 585–601.

Miettinen, K. (1998). Concepts. In *Nonlinear multiobjective optimization*, 5–36. Springer.

Mitchell, T.M. et al. (1997). *Machine learning*. McGraw-hill New York.

Nag, K. and Pal, N.R. (2015). A multiobjective genetic programming-based ensemble for simultaneous feature selection and classification. *IEEE transactions on cybernetics*, 46(2), 499–510.

Onan, A., Korukoğlu, S., and Bulut, H. (2016). A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Systems with Applications*, 62, 1–16.

Pant, M., Zaheer, H., Garcia-Hernandez, L., Abraham, A., et al. (2020). Differential evolution: a review of more than two decades of research. *Engineering Applications of Artificial Intelligence*, 90, 103479.

Radovic, M., Ghalwash, M., Filipovic, N., and Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC bioinformatics*, 18(1), 1–14.

Reynoso-Meza, G., Hernández, M., and Carreño Alvarado, E.P. (2022). Multi-objective insights and analysis on data driven classifiers for anomaly detection in water distribution systems. In *Proceedings of the 2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry*, 11. Universitat Politecnica de Valencia.

Reynoso-Meza, G., Sanchis, J., Blasco, X., and Garcia-Nieto, S. (2014). Physical programming for preference driven evolutionary multi-objective optimization. *Applied Soft Computing*, 24, 341–362.

Ribeiro, V.H.A., Moritz, S., Rehbach, F., and Reynoso-Meza, G. (2020). A novel dynamic multi-criteria ensemble selection mechanism applied to drinking water quality anomaly detection. *Science of The Total Environment*, 749, 142368.

Ribeiro, V.H.A. and Reynoso-Meza, G. (2020). Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets. *Expert Systems with Applications*, 147, 113232.

Ribeiro, V.H.A. and Reynoso-Meza, G. (2021). Feature selection and regularization of interpretable soft sensors using evolutionary multi-objective optimization design procedures. *Chemometrics and Intelligent Laboratory Systems*, 212, 104278.

Ryu, D. and Baik, J. (2016). Effective multi-objective naïve bayes learning for cross-project defect prediction. *Applied Soft Computing*, 49, 1062–1077.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4), 341–359.