2nd International Joint Conference on Water Distribution
Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14761

MULTI-OBJECTIVE INSIGHTS AND ANALYSIS ON DATA DRIVEN CLASSIFIERS FOR ANOMALY DETECTION IN WATER DISTRIBUTION SYSTEMS

Elizabeth Pauline Carreño-Alvarado[1], Mayra Hernández-Alba[2] and Gilberto Reynoso-Meza[3]

[1]Dpto. de Ingeniería Geomática e Hidráulica, Universidad de Guanajuato, Guanajuato Guanajuato, México.
[2] Universidad de Guanajuato, Guanajuato Guanajuato, México.
[3]Pontifícia Universidade Católica do Paraná (PUCPR) Programa de Pós-graduação em Engenharia de Produção e Sistemas (PPGEPS), Curitiba, Paraná, Brazil.

[1] ep.carreno@ugto.mx, [2] m.hernandezalba@ugto.mx, [3] g.reynosomeza@pucpr.br

Abstract

Machine learning techniques have shown to be a powerful tool for extracting and/or inferring complex patterns from data. In the case of the so-called supervised learning, a given learner representation could learn such patterns using labeled data. For example, a helpful approach is to adjust a learner to detect anomalies: historical data can be used, where those events are identified, to find a pattern to classify new data as an anomaly (true event) or not (false event). In this example, the learner's objective is to act as a binary classifier, where a balance between false negatives (predict a typical operation, when in fact an anomaly exists) and false positives (predict an anomaly, when there is not). This balance is attained via an optimization (learning phase), where the learner representation is adjusted. Multi-objective optimization techniques have a natural way of dealing with such problems. They perform a simultaneous optimization of conflicting objectives. As a result, a set of Pareto-optimal solutions, the Pareto front, is calculated. This idea could be used in the training process of binary classifiers.

Nevertheless, this requires an integral methodology, merging multi-objective optimization and multi-criteria decision making. While it is true that this idea is not new, methodologies and guidelines are still missing to conduct this process. In this work, we move toward the definition of an integrated methodology of multi-objective learning for binary classifiers for anomaly detection. An anomaly detection database for water distribution systems is used for such a purpose. Preliminary results show to be competitive regarding the F1-score to similar approaches.

Keywords
Machine learning, Logistic regression, Multi-objective optimisation, Water distribution systems.

1 INTRODUCTION

Machine learning techniques have shown to be a powerful tool for different kinds of applications [1,2,3,4,5]. Their data-driven approach makes them suitable for finding complex patterns and relationships with enough (and well processed) data. After a supervised learning process, a given machine learning representation can identify anomaly events from regular events [6]. In such cases, they are referred to as virtual or soft sensors: instead of having a physical device, information from other sources is mixed in order to infer, in this case, an anomaly event [7].

Such a learning process is usually performed via optimization with a single objective cost function [8]. Nevertheless, the trade-off between false positives and false negatives is evident for classification purposes. That means it is worthwhile to analyze the trade-off of a given classifier for anomaly detection between triggering a false alarm or letting pass risk situations. Even if both instances are considered equivalent misclassifications for practical purposes, they are not. On the one hand, false positives trigger an alarm that requires attention; an excess of false alarms could overwhelm the technical staff. On the other hand, false negatives are undetected situations that could be hazardous to the system's operation. Therefore, there is a trade-off that a learner must achieve through the learning process.

In such instances, where there is a clear trade-off between conflicting objectives, multi-objective optimization could be an interesting tool [9]. Multi-objective optimization deals with conflicting objectives simultaneously. Consequently, a Pareto front is approximated: a set of Pareto optimal solutions. In such a set, the only difference between two solutions is the trade-off that they exhibit among conflicting objectives. Therefore, it is possible to ponder benefits and drawbacks actively when the decision-maker favors one objective over another. Such an idea could bring compelling solutions for machine learning [10,11].

Nevertheless, the multi-objective nature demands a multi-criteria analysis to select a solution from the set to be implemented. While it is true that this idea is not new, methodologies and guidelines are still missing to conduct this process. Therefore, it is necessary to move toward the definition of an integrated methodology of multi-objective learning for binary classifiers.

This work deals with the binary classification problem using logistic regression and explores the advantages of using a multi-objective optimization approach for its training. Additionally, we point out some interesting facts and guidelines for the multi-criteria analysis. The remainder of this paper is as follows: In Section 2, a brief background on multi-objective optimization and supervised machine learning are given. In section 3, the proposal of this exploratory work is presented, whilst in Section 4, the study design is explained. In Section 5, results are commented on and discussed, and finally, conclusions and future works are presented.

2 DESCRIPTION

Next, fundamental ideas on machine learning and multi-objective optimization are given.

2.1 Machine learning, supervised learning, and binary classification

Machine learning refers to computer algorithms that improve themselves automatically through experience in the form of data. Such a learning process could be supervised (requiring inputs and targets), unsupervised (inputs required), or by reinforcement (via interaction with the surroundings) [12]. Supervised learning uses as inputs M instances (observations) with N features (explanatory variables) to train a given learner representation using reliable information of the targets T for each one of the instances. The main goal is to construct a relationship to provide an output (target prediction) for any new instance (with its features). Such training could be oriented for classification or regression. In both cases, several representations such as artificial neural networks [13], support vector machines [14], or decision trees [15] exist, among others.

A binary classification must predict if arriving data belong to class 0 or class 1. Such classes could be identified in the anomaly detection case by answering the following question: Is there an anomaly? A class 0 event is a situation where no anomaly exists; on the opposite, a class 1 event is an anomaly situation. A given learner will enter a training phase using a dataset $[MxN|T]$ to adjust its parameters β via an optimization phase, using some evaluation criteria or cost function.

Usually, this learning process's cost function for optimization is an aggregation function of correct classification and misclassifications. Furthermore, given a parameter vector β for a learner representation, it is usual to evaluate its final performance in the same way, merging true positives ($TP(\beta)$), false negatives ($FN(\beta)$), false positives ($FP(\beta)$) and true negatives ($TN(\beta)$). For example:

$$Error\ rate = \frac{FP + FN}{TP + FP + TN + FN} \tag{1}$$

$$F1\ score = \frac{2 * TPR * PPV}{TPR + PPV} \tag{2}$$

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

Also, it is usual to depict such trade-offs with a confusion matrix (Figure 1). All performance index reveals that a trade-off between *FN* and *FP* is pondered. Both should be minimized, and normally, they are conflicting objectives. Therefore, multi-objective optimization could be an interesting tool to deal with such a situation.

Figure 1. Confusion matrix. It is possible to visualise the performance of a given learner, identifying TN and TP (green boxes) and FP and FN (red boxes).

2.2 Multi-objective optimization

As commented in [16], a multi-objective problem (MOP) with m objectives can be stated as follows:

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = [J_1(\boldsymbol{\theta}), \dots, J_m(\boldsymbol{\theta})] \tag{6}$$

Subject to:

$$\boldsymbol{K}(\boldsymbol{\theta}) \leq \mathbf{0} \tag{7}$$

$$\boldsymbol{L}(\boldsymbol{\theta}) \leq \mathbf{0} \tag{8}$$

$$\underline{\theta_i} \leq \theta_i \leq \overline{\theta_i}, i = [1, \dots, n] \tag{9}$$

Where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]$ is defined as the decision vector with dim($\boldsymbol{\theta}$) = n; $\boldsymbol{J(\theta)}$ as the objective vector and $\boldsymbol{K(\theta)}$, $\boldsymbol{L(\theta)}$ 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 has been noticed that there is not a single solution in MOPs because there is not generally a better solution for all the objectives. Therefore, a set of solutions, the Pareto set $\boldsymbol{\Theta_P}$, is defined. Each solution in the Pareto set defines an objective vector in the Pareto front $\boldsymbol{J_P}$ (See Figure 2). It is important to notice that most of the time, we rely only on the Pareto front and set approximations $\boldsymbol{J_P^*}$, $\boldsymbol{\Theta_P^*}$. P. All the solutions in the Pareto front are a set of Pareto optimal and non-dominated solutions, where:

- Pareto optimality [16]: An objective vector $\boldsymbol{J}(\boldsymbol{\theta^1})$ is Pareto optimal if there is not another objective vector $\boldsymbol{J}(\boldsymbol{\theta^2})$ such that $J_i(\boldsymbol{\theta^2}) \leq J_i(\boldsymbol{\theta^1})$ for all *i* ∈ *[1,2,...,m]* and $J_j(\boldsymbol{\theta^2}) < J_j(\boldsymbol{\theta^1})$ for at least one *j*, *j* ∈ *[1,2,...,m]*.
- Dominance [17]: An objective vector $\boldsymbol{J}(\boldsymbol{\theta^1})$ is dominated by another objective vector $\boldsymbol{J}(\boldsymbol{\theta^2})$ iff $J_i(\boldsymbol{\theta^2}) \leq J_i(\boldsymbol{\theta^1})$ for all *i* ∈ *[1,2,...,m]* and $J_j(\boldsymbol{\theta^1}) < J_j(\boldsymbol{\theta^1})$ for at least one *j, j* ∈ *[1,2,...,m]*. This is denoted as $\boldsymbol{J}(\boldsymbol{\theta^2}) \preccurlyeq \boldsymbol{J}(\boldsymbol{\theta^1})$.

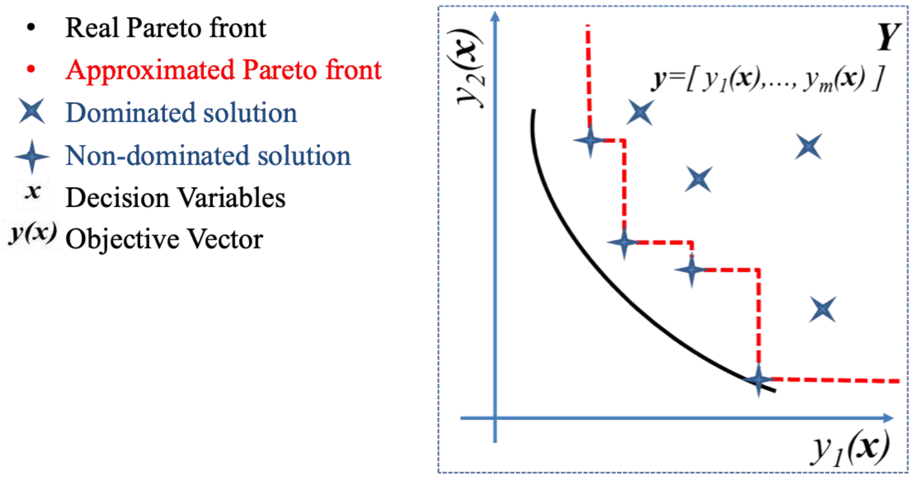


Figure 2. Pareto optimality and dominance concepts for a min-min MOP. Non-dominated solutions approximate (red dotted line) the unknown Pareto front (blsck solid line) in the objective space Y. Remainder solutions are dominated solutions.

The multi-objective optimization approach, from a practical point of view, requires three main steps:

Multi-objective optimization statement: this implies defining the design objectives to optimize, the decision variables, and the parametric model to establish an unequivocal correspondence. The multi-objective optimization (MOO) process corresponds to the optimization process itself. Requires to define an optimization algorithm (with its hyper-parameters), running platform, and hardware requirements/conditions. Multi-criteria Decision Making (MCDM) step: the final process where a solution from the Pareto front approximation should be selected, and its correspondent design alternative from the Pareto set implemented. As a multi-criteria analysis needs to be performed, visualization tools and multi-criteria methods are usually required. The procedure integrating those three steps is named the multi-objective optimization design (MOOD) procedure [18]. Next, it will be presented how to use such a procedure in the multi-objective learning process.

3 TOOLS AND METHODS

Here, the proposal is presented to adjust a binary classifier via a MOOD procedure. It covers the MOP statement, the MOO process, and the MCDM stage.

3.1 Multi-objective problem statement

The parametric model to be used is the logistic regression due to its simplicity and because it is the simplest binary classifier. Future work will focus on different representations or tribes [19]. The logistic regression uses the sigmoid function (equation 10) to compute the probability of a given observation to be 1.

$$h_\theta(x) = \frac{1}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\,\dots\,+\beta_n x_n}} \tag{10}$$

Where $x = [x_1,\dots,x_N]$ are the N explanatory variables or features of the learner; $\boldsymbol{\beta} = [\beta_1, \dots, \beta_N]$ are the regression coefficients adjusted given M observations or instances. Usually, given this set of M instances, the parameter β is adjusted using the loss function of Equation (11):

$$Loss(h_{\boldsymbol{\beta}}(\boldsymbol{x}), y) = \begin{cases} -\log(h_{\boldsymbol{\beta}}(\boldsymbol{x})) & \text{if } y = 1 \\ -\log(1 - h_{\boldsymbol{\beta}}(\boldsymbol{x})) & \text{if } y = 0 \end{cases} \tag{11}$$

and the cost function for optimisation of (12):

$$CE(\boldsymbol{\beta}) = -[y\log(h_{\boldsymbol{\beta}}(\boldsymbol{x})) + (1-y)\log(1-h_{\boldsymbol{\beta}}(\boldsymbol{x}))]/M \tag{12}$$

Instead of using an aggregation function for FP and FN, such performance will be evaluated simultaneously via multi-objective optimization. Therefore, a multi-objective problem is considered as shown in Equation (13):

$$\min_{\boldsymbol{\beta}} J(\boldsymbol{\beta}) = [FP + CE_m(\boldsymbol{\beta}), FN + CE_m(\boldsymbol{\beta})\ ] \tag{13}$$

With:

$$CE_m(\boldsymbol{\beta}) = -[y\log(h_{\boldsymbol{\beta}}(\boldsymbol{x})) + (1-y)\log(1-h_{\boldsymbol{\beta}}(\boldsymbol{x}))]/(M \cdot \log(\epsilon)) \tag{14}$$

Subject to:

$$\underline{\beta_i} \le \beta_i \le \overline{\beta_i}, i = [1, \dots, n] \tag{15}$$

3.2 Multi-objective optimization process

For the experiments presented here, the spMODEx algorithm will be used. It is a multi-objective evolutionary algorithm based on Differential Evolution [20, 21], using as diversity mechanism a spherical pruning [22]. The following hyperparameters are used [23]:

Mutation: binomial;

Scaling factor: 0.5;

Crossover rate: 0.9;

Population: 20; Function evaluations: 1e4; Spherical arcs: 20.

Next, a proposal for the decision-making step is placed. It includes an analysis of the Pareto front and the Pareto set approximations.

3.3 Multi-criteria decision making

Next, a proposal for the decision-making step is placed. It includes an analysis of the Pareto front and the Pareto set approximations.

- Evaluate the design alternatives from the Pareto set with the data set for training ($J_P{}^{t*}$) and depict them using a 2-dimensional plot.
- Evaluate the design alternatives from the Pareto set with the data set for testing (J^{v*}) and plot it in the same 2-dimensional plot as $J_P{}^{t*}$.
- Perform a work scenario comparison [24]. Different from a design concept comparison [25], where different concepts are used to perform the multi-objective process, the work scenario comparison evaluates the performance of a given Pareto set approximation under different conditions.
- Determine the deformation from the $J_P{}^{t*}$ towards J^{v*}. This evaluation will be helpful in measuring the internal coherence of the learners and their trade-offs. That is if the trade-off ordering among solutions is preserved.
- Define a region of interest (or pertinent region).
- Pick the most suitable solution and evaluate it with the test data.
- Plot Pareto set via parallel coordinates and boxplot. Perform a critical analysis of the most important features.

4 STUDY DESIGN

Providing clean and safe drinking water is crucial for any water supply company [26]. To guarantee such a supply, automatic anomaly detection plays a critical role in drinking water quality monitoring [27]. Recent anomaly detection techniques incorporate tools from the machine learning area [1]. This work uses a real-world data set generated in a research project on drinking water. The data set consists of data from *Thüringer Fernwasserversorgung*, a major German water supplier located in central Germany. This data set has been used for different competitions about anomaly detection for drinking water in major international conferences [26]. The data and additional documentation are available for download [28]. In Table 1, the features used are depicted. More details are provided in [6].

For this example, the following data science methods are implemented:

- Pre-processing: an imputing mechanism has been implemented for instances with missing or not interpretable values (repeat last value).
- Feature engineering: No additional features are included. This means that this problem is being treated as a static problem instead of a dynamic time series.
- Processing: A moving average of 1440 samples was used for detrending.

Table 1. Some Letters and Numbers [Caption, Cambria, 10pt, Italic, centred]

Parameter	Unit	Description
Time	*datetime*	Time Stamp
WT	°C	Water Temperature
$ClO2_1$, $ClO2_2$	*mg/l*	Chlorine Dioxide (2 values)
pH	*pH*	pH Value
Redox	*mV*	Redox Potential
EC	*μS/cm*	Conductivity
TURB	*NTU*	Turbidity
FR_1, FR_2	m^3/h	Water Flow Rate (2 values)
EVENT	*binary*	Anomaly Label

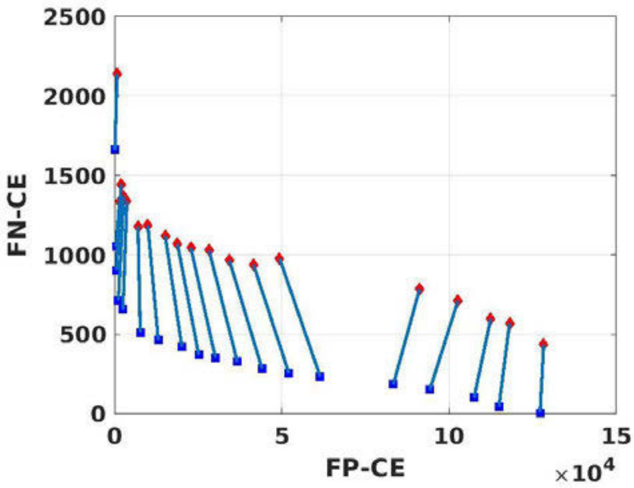
5 MULTI-OBJECTICE LEARNING RESULTS AND DISCUSSION

In Figure 3a, the Pareto front approximation is depicted. With blue squares, the Pareto front approximation J_P^{t*} while the red diamonds represent the deformation of the approximation J^{v*}. When evaluated with the testing data, the translation of a given design alternative β from J_P^{t*} is represented with blue lines. Such an analysis could reveal when an over-fitting occurs: it is expected to keep the trade-off coherence, as well as a similar trade-off or a reasonable mismatch between performance with the training set and the testing set (see [29]). In this case, as practically no blue lines are crossing, the learners in J_P^{t*} keep their internal coherence for decision making. Multiple crosses or gathering to a single region could reveal an over-fitting.

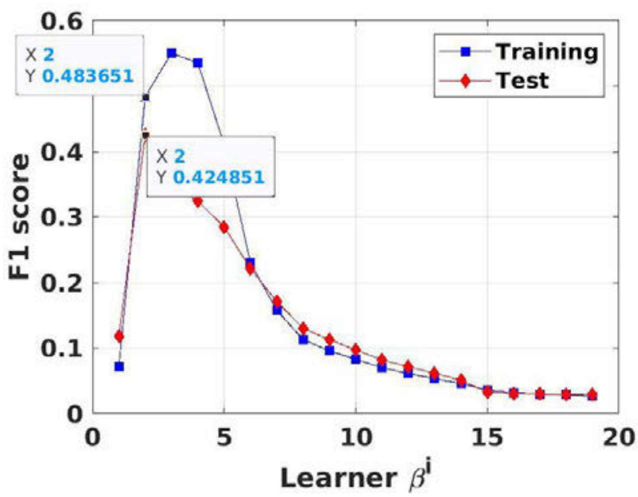
In Figure 3b F1-score is calculated for each of the trained learners. This index is a usual choice to evaluate overall performance whit imbalanced data. The closest to 1, the better the classifier according to this score. As it can be noticed, the learner with the highest F1 score in the training phase is not the better in the validation test. This reveals how important could be the decision-making stage, given that a good performance on the training set does not assure good generalization abilities. The best learner in the validation set is the second one, which F1-score value is relatively close to that achieved in the training set. Given that, this is the learner that is recommended for further implementation. Sixth to nineteenth learners exhibit practically the same performance.

In Figure 3c, the values of the parameter vector β of the approximated Pareto set are depicted using boxplots and parallel coordinates. With such an analysis, it is possible to appreciate the impact of a given feature on the prediction capabilities of the classifier. For example, zero values on β_8 and β_9 (features 8 and 9, FR1 and FR2) indicate that this measure could be potentially omitted. That could mean saving a couple of measures and sensors in a practical sense. On the opposite, β_1 and β_6 (features 1 and 6, WT and EC) seems to be the one that the most impact has in the classification; β_2 and β_7 (features 2 and 7, ClO21 and TURB) seems to be responsible on the trade-off exchange in the set.

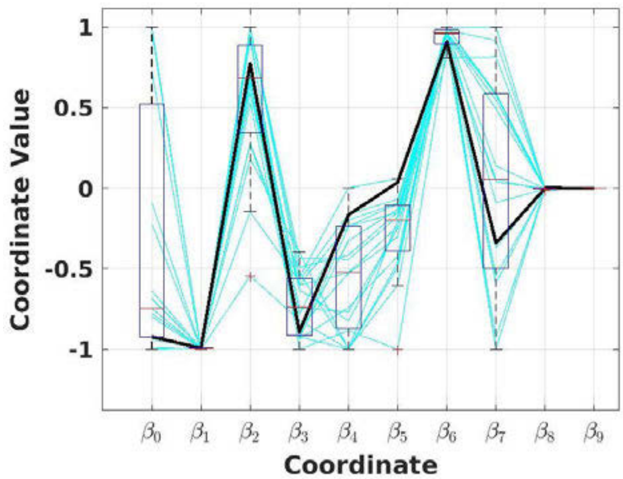
Finally, Figure 4 depicts the confusion matrix of the design alternative (learner) number two, the one selected in figure 3. The F1 score is competitive compared to other classifiers, with similar data science methods (pre-processing, feature engineering, processing, and splitting). For example, ensemble methods using support vector machines and decision trees (third and fourth place) had an F1 score of 0.39 and 0.45, respectively. Interestingly, no additional treatment for the imbalanced data was required by simultaneously considering both classes with the multi-objective approach.



(a) *Pareto front approximation*



(b) *F1-score*



(c) *β distribution in the Pareto set approximation*

Figure 3. Performance visualisation of the approximated Pareto front and set

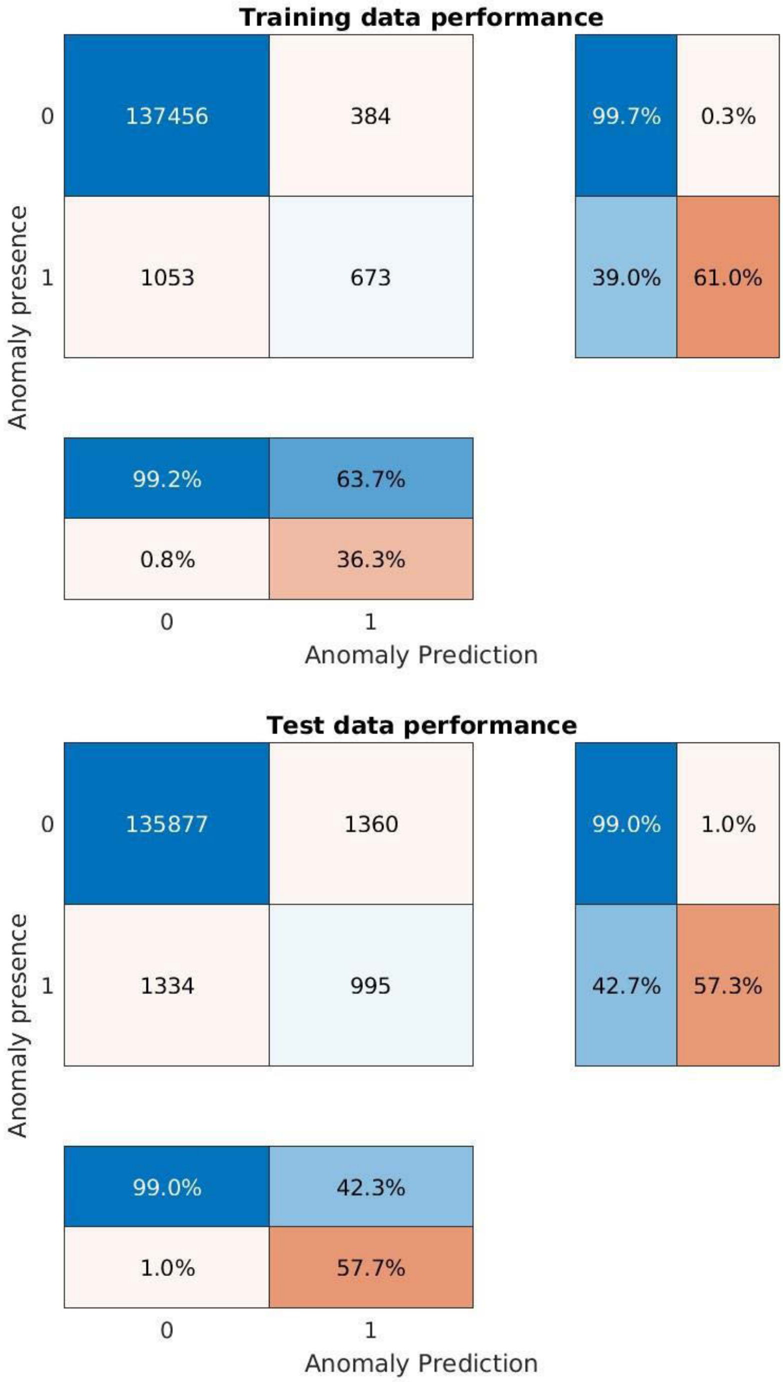


Figure 4. Confusion matrix of the learning vector selected using the F1 criteria.

6 CONCLUSIONS

In this work, it has been proposed a MOOD procedure for multi-objective training in binary classifiers. The multi-objective approach allows considering the *FP* and FN ratio simultaneously. That is, it is possible to train a set of learners with a different trade-off between false positives and false negatives. This could be interesting in the decision-making stage, given that it is possible to select a learner with an affordable cost regarding not-detected threats and/or triggering false alarms. Furthermore, a proposal for the decision-making stage has been presented to help in the selection of a suitable learner for implementation. An example using logistic regression was presented for anomaly detection in water distribution systems. It includes a Pareto front comparison to validate trade-off coherence with the data set for testing; this is important given that it also allows for verifying the generalization capabilities of the learner. An example of anomaly detection in a water distribution system was presented. Obtained results are competitive with other approaches under the same mechanisms for pre-processing, feature engineering, and processing of data. Future work will concentrate on the design concept comparison of different learners' representations and verifying different options for the MOP statement and indexes for the MCDM step.

7 REFERENCES

[1] Dogo, E.M., Nwulu, N.I., Twala, B., Aigbavboa, C.: A survey of machine learning methods applied to anomaly detection on drinking-water quality data. Urban Water Journal 16(3), 235–248 (2019).

[2] Weichert, D., Link, P., Stoll, A., Ru¨ping, S., Ihlenfeldt, S., Wrobel, S.: A review ofmachine learning for the optimization of production processes. The International Journal of Advanced Manufacturing Technology 104(5), 1889–1902 (2019)

[3] Ghassemi, M., Naumann, T., Schulam, P., Beam, A.L., Chen, I.Y., Ranganath,R.: A review of challenges and opportunities in machine learning for health. AMIA Summits on Translational Science Proceedings 2020, 191 (2020).

[4] Kadlec, P., Gabrys, B.: Soft sensors: where are we and what are the current andfuture challenges? IFAC Proceedings Volumes 42(19), 572–577 (2009).

[5] Alsrehin, N.O., Klaib, A.F., Magableh, A.: Intelligent transportation and controlsystems using data mining and machine learning techniques: A comprehensive study. IEEE Access 7, 49830–49857 (2019).

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

[7] Souza, F.A., Arau´jo, R., Mendes, J.: Review of soft sensor methods for regressionapplications. Chemometrics and Intelligent Laboratory Systems 152, 69–79 (2016).

[8] Domingos, P.: A few useful things to know about machine learning. Communications of the ACM 55(10), 78–87 (2012).

[9] Meza, G.R., Ferragud, X.B., Saez, J.S., Dura´, J.M.H.: Controller tuning with evolutionary multiobjective optimization (2017).

[10] Ribeiro, V.H.A., Reynoso-Meza, G.: Ensemble learning by means of a multiobjective optimization design approach for dealing with imbalanced data sets. Expert Systems with Applications 147, 113232 (2020).

[12] Mitchell, T.M., et al.: Machine learning. McGraw-hill New York (1997).

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

[13] Najah, A., El-Shafie, A., Karim, O.A., El-Shafie, A.H.: Application of artificialneural networks for water quality prediction. Neural Computing and Applications 22(1), 187–201 (2013).

[14] Deka, P.C., et al.: Support vector machine applications in the field of hydrology: a review. Applied soft computing 19, 372–386 (2014).

[15] Kotsiantis, S.B.: Decision trees: a recent overview. Artificial Intelligence Review39(4), 261–283 (2013).

[16] Miettinen, K.: Concepts. In: Nonlinear multiobjective optimization, pp. 5–36. Springer (1998).

[17] Coello, C.A.C., Lamont, G.B.: Applications of multi-objective evolutionary algorithms, vol. 1. World Scientific (2004).

[18] Reynoso Meza, G.: Controller tuning by means of evolutionary multiobjective optimization: a holistic multiobjective optimization design procedure. Ph.D. thesis, Editorial Universitat Polit`ecnica de Val`encia (2014).

[19] Domingos, P.: The master algorithm: How the quest for the ultimate learningmachine will remake our world. Basic Books (2015).

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

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

[22] Reynoso-Meza, G., Sanchis, J., Blasco, X., Mart´ınez, M.: Design of continuouscontrollers using a multiobjective differential evolution algorithm with spherical pruning. In: European Conference on the Applications of Evolutionary Computation. pp. 532–541. Springer (2010).

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

[24] Reynoso-Meza, G., Carrillo-Ahumada, J., Torralba, L.M.: Pareto front analysis ofcontrol structures in multimodel systems: The work scenario concept introduction.

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

[26] Rehbach, F., Chandrasekaran, S., Rebolledo, M., Moritz, S., Bartz-Beielstein, T.: GECCO Challenge 2018: Online Anomaly Detection for Drinking Water Quality (2018).

[27] Banna, M.H., Imran, S., Francisque, A., Najjaran, H., Sadiq, R., Rodriguez, M.,Hoorfar, M.: Online drinking water quality monitoring: review on available and emerging technologies. Critical Reviews in Environmental Science and Technology 44(12), 1370–1421 (2014).

[28] Moritz, S., Rehbach, F., Chandrasekaran, S., Rebolledo, M., Bartz-Beielstein, T.: GECCO Industrial Challenge 2018 Dataset: A water quality dataset for the 'Internet of Things: Online Anomaly Detection for Drinking Water Quality' competition at the Genetic and Evolutionary Computation Conference 2018, Kyoto, Japan. (Feb 2018). https://doi.org/10.5281/zenodo.3884398.

[29] Reynoso-Meza, G., Vignoni, A., Boada, Y., Picu¨, J., Picu¨, E.: Model mismatch in multi-objective optimisation and preservation of trade-off order. IFACPapersOnLine 52(26), 249–254 (2019)