

Detecting Pneumatic Failures on Temporary Immersion Bioreactors

Octavio Loyola-González^{1,2(✉)}, José Fco. Martínez-Trinidad¹,
Jesús A. Carrasco-Ochoa¹, Dayton Hernández-Tamayo²,
and Milton García-Borroto³

¹ Instituto Nacional de Astrofísica, Óptica y Electrónica,
Luis Enrique Erro No. 1, Sta. María Tonanzintla, 72840 Puebla, Mexico
{octavioloyola, fmartine, ariel}@inaoep.mx

² Centro de Bioplantitas, Universidad de Ciego de Ávila,
Carretera a Morón km 9, 69450 Ciego de Ávila, Cuba
{octavioloyola, dayton}@bioplantitas.cu

³ Instituto Superior Politécnico José Antonio Echeverría,
Calle 114 No. 11901, 19390 Marianao, La Habana, Cuba
mgarciab@ceis.cujae.edu.cu

Abstract. Temporary immersion bioreactors are an effective procedure to increase plant multiplication rates. The pneumatic system is an important part of a bioreactor, which should be controlled to guarantee both the efficiency and efficacy in the system. Therefore, bioreactors have been automated using a pneumatic drive to execute the immersion time. Sometimes, the pneumatic system presents failures which can affect the plant quality; therefore, pneumatic failure detection is an important task. Since failures are a few compared with the normal behavior, it is a class imbalance problem. In this paper, we study the use of contrast pattern-based classifiers, designed for class imbalance problems, for creating an understandable and accurate model for detecting pneumatic failures on temporary immersion bioreactors. Our experiments over eight real-world databases show that a decision tree ensemble obtains significantly better AUC results than other tested classifiers.

Keywords: Supervised classification · Contrast patterns · Imbalanced databases · Failure detection

1 Introduction

Temporary immersion systems have been widely used as an effective technology for increasing the plant multiplication rates and plant quality [2, 10, 12, 13, 24]. A temporary immersion bioreactor (TIB) is a type of temporary immersion system which allows immerse periodically the plants into nutrient medium. A TIB has several variables which should be controlled, as duration and frequency of nutrient medium through air pressure, for guaranteeing both the efficiency and

the efficacy¹ in TIBs. Therefore, temporary immersion bioreactors have been automated using a pneumatic drive to execute the immersion time in specific time intervals [10–13]. An important task for temporary immersion bioreactors is detecting failures on the pneumatic system, since these failures could affect the multiplication rate, the plant quality, or even discontinue the plant micro-propagation into the temporary immersion bioreactor [13]. Failure detection on the pneumatic system of temporary immersion bioreactors could be considered as a class imbalance problem, because failures usually are a few compared with the normal process of immersion.

In real-world imbalanced problems, the objects are not equally distributed among classes, which produces a bias of classification results to the majority class (the class with more objects). Frequently, the most interesting class, for an expert of the application domain, is the one that contains significantly less objects (minority class), because it is commonly associated to rare cases [28]. This type of problems is known as the class imbalance problem. Currently, there are several classifiers designed to deal with the class imbalance problem [20–22]. However, not all classifiers make a model that can be understood by experts in the application domain [5, 9, 16, 22].

Contrast pattern-based classifiers are an important family of both understandable and accurate classifiers. These classifiers have been used to solve real-world problems in fields like bioinformatics [9], intruder detection [5], anomaly detection in network connection data [29], rare event forecasting [27], and privacy preserving data mining [1], which are well-known class imbalance problems [20, 22].

In this paper we present a study of the use of contrast pattern-based classifiers for pneumatic failures detection in temporary immersion bioreactors, using eight real-world databases of pineapple plants. Through this study, we show that the Area Under the receiver operating characteristic Curve (AUC) [17] is significantly improved when the HeDex classifier [19] is used. The main contribution of this paper is the use of contrast pattern-based classifiers for pneumatic failure detection in temporary immersion bioreactors (a class imbalance problem), which allows creating an understandable and accurate model that forewarns the experts in practice about failures in the temporary immersion bioreactor.

The rest of the paper has the following structure. Section 2 provides a brief introduction to temporary immersion bioreactors and their pneumatic system. Section 3 reviews some of the most popular contrast pattern-based classifiers designed to deal with class imbalance problems. Section 4 presents our study about pneumatic system failure detection using the classifiers presented in Sect. 3 over eight real-world databases, including the experimental setup and a brief interpretation, issued by the pneumatic system expert, of the most frequent patterns belonging to the minority class (failure). Finally, Sect. 5 provides conclusions and future work.

¹ By efficiency and efficacy we mean high plant growing rate, high quality plants, and low production cost.

2 Pneumatic Systems in Temporary Immersion Bioreactors

Temporary immersion bioreactors are a type of temporary immersion system, which are commonly categorized into mechanically agitated and pneumatically agitated bioreactors [24].

The temporary immersion bioreactor at the *Centro de Bioplantas*² is classified as a pneumatically agitated bioreactor, which is constituted by two transparent glass (or plastic) containers, autoclavable silicone tubes, hydrophobic air filters, electric valves, and an air compressor (see Fig. 1) [10]. One container is for growing plants and the other container is for liquid medium. These are connected by silicone tubes. In each case, the air flow is sterilized by passage through hydrophobic filters. Then, air pressure from an air compressor pushes the medium from one container to the other to immerse the plants completely. The air flow is reversed to withdraw the medium from the culture container. Also, three-way solenoid valves provided on/off operation; where the frequency and length of the immersion period is controlled by using a programmable logic controller (PLC) [3] connected with a supervisory control and data acquisition (SCADA) system through the *modus* protocol [4].

The main failures that arise in the pneumatic system of this temporary immersion bioreactor are:

- (i) Bad air pressure into the central distribution line.
- (ii) Bad air pressure into the plant container.

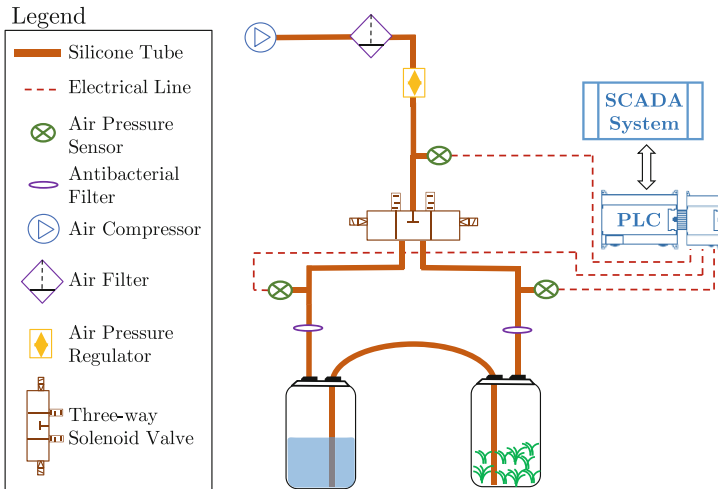


Fig. 1. Temporary immersion bioreactor diagram.

² www.bioplantas.cu.

- (iii) Bad air pressure into the liquid medium container.
- (iv) There is an obstruction³ into the silicone tube connecting the containers.

If these failures are not detected as soon as they take place in the pneumatic system then the plant micropropagation into the temporary immersion bioreactor is discontinued.

In this paper, we propose to use contrast pattern-based classifiers, which allow creating understandable and accurate models. In this way, the pneumatic system experts could introduce this model into the PLC to forewarn failures in the temporary immersion bioreactor.

Despite there are several contrast pattern-based classifiers reported in the literature, not all can deal with class imbalance problems.

3 Contrast Pattern-Based Classifiers in Class Imbalance Problems

There are three main approaches to deal with the class imbalance problem [21, 22]; data level, algorithm level, and cost-sensitive. The cost-sensitive approach has as drawback that the cost matrix is commonly unknown. On the other hand, the data level approach has well-known drawbacks like representative object exclusion or promoting of classifier overfitting. However, ensemble methods (algorithm level) have reported good classification results in class imbalance problems [18, 26]. Therefore, we will be focusing on the application of ensembles of contrast pattern-based classifiers to detect pneumatic failures on temporary immersion bioreactors.

The following classifiers are among the most popular contrast pattern-based classifiers designed to deal with the class imbalance problem.

Coverage creates balanced subsamples without oversampling the minority class in order to create a decision tree ensemble [18]. The main goal is to establish robustness to the *consolidated tree construction* algorithm (CTC) [25]. CTC is an algorithm based on decision trees that requires the use of several subsamples of the training sample (as a lot of multiple classifier systems), but it returns a single decision tree. Coverage uses internally the CTC algorithm.

HeDex is a decision trees ensemble [19] using the *Hellinger Distance* [6] as a decision tree splitting criterion. It builds extremely randomized ensemble trees through the randomization on both attribute selection and split-point selection, which allows to achieve high level of variety of decision trees. HeDex has shown good results over several imbalanced databases [19].

RUSBoost is a decision tree ensemble that uses a boosting algorithm [26]. RUSBoost applies a resampling method (RUS) that randomly removes examples from the majority class. RUSBoost, also is based on the SMOTEBoost algorithm (which use the AdaBoost.M2 algorithm [14]) but RUSBoost uses RUS

³ The most frequent obstruction is produced by waste of plant material.

rather than SMOTE. The RUSBoost classifier presents a simpler, faster, and less complex alternative to SMOTEBoost for learning from imbalanced databases [26].

iCAEP performs information-based Classification by Aggregating Emerging Patterns [30]. It uses the minimum encoding inference approach to classify an object, instead of the aggregation of support. iCAEP selects a smaller but more representative subset of contrast patterns from the object to be classified.

4 Detecting Pneumatic Failures on Temporary Immersion Bioreactors Through Contrast Pattern-Based Classifiers

This section presents an empirical study on using contrast pattern-based classifiers for detecting pneumatic failures on temporary immersion bioreactors, which is a class imbalance problem. The experimental setup is presented in Sect. 4.1 and the experimental results are presented in Sect. 4.2.

All dataset partitions used in this paper as well as the experimental results are available for downloading from our supplementary material website⁴.

4.1 Experimental Setup

For our experiments, we use eight pineapple databases, which were collected from the temporary immersion bioreactor at *Centro de Bioplantas* [11]. All databases contain 210 numerical attributes from the sensors of the pneumatic system. Each object represent an immersion time (of 70 seconds) in the temporary immersion bioreactor. Each database contains 70 attributes corresponding to measures of air pressure into the central distribution line, 70 attributes corresponding to measures of air pressure into the liquid medium container, and the last 70 attributes corresponding to measures of air pressure into the plant container (the air pressure sensors can be visualized in Fig. 1). Each object labeled as *failure* represents a problem during the immersion time, more specifically it means that the liquid medium was not transferred from a container to another.

Table 1 shows for each database: the name, the number of objects belonging to the minority (or failure) class (#Objects_Min), the number of objects belonging to the majority class (#Objects_Maj), and the class imbalance ratio (IR).

All databases were partitioned using 5-fold and distribution optimally balanced stratified cross validation (DOB-SCV) [23] with the goal of avoiding problems into data distribution on highly imbalanced databases [21].

The iCAEP classifier takes advantage of the patterns extracted through a contrast pattern miner; therefore, we used the *bagging miner* algorithm [15] to extract the contrast patterns to be used by the iCAEP classifier. The main reasons is that this miner has reported good classification results over databases

⁴ <http://sites.google.com/site/octavioloyola/papers/DPFonTIBs>.

Table 1. Summary of the imbalanced databases used in our study

Name	#Objects_Min	#Objects_Maj	IR
Pineapple1	17	236	13.88
Pineapple2	11	236	21.45
Pineapple3	25	237	9.48
Pineapple4	14	237	16.93
Pineapple5	19	236	12.42
Pineapple6	16	237	14.81
Pineapple7	15	237	15.80
Pineapple8	12	234	19.50

with numerical attributes [15], and given that our databases contain only numerical attributes we chose bagging miner. Additionally, in [6] the authors suggested to use the bagging algorithm with the Hellinger distance to obtain good classification accuracy; therefore, we used the Hellinger distance as a decision tree splitting criterion into the bagging miner.

For the contrast pattern methods presented in Sect. 3, we used the parameter values recommended by their authors. For Coverage, we used the WEKA⁵ implementation provided by its authors. For RUSBoost, we used the implementation into the KEEL⁶ Data-Mining tool. For HeDex, iCAEP, and Bagging Miner we implemented our own versions.

We used the Friedman test [7] and the Bergmann-Hommel dynamic post-hoc procedure [8] to statistically compare our results. Moreover, post-hoc results are presented using *critical distance* (CD) diagrams [7]. Usually, in a CD diagram, the position of the classifier within the segment represents its rank value, where the rightmost classifier is the best one. If two or more classifiers share a thick line it means they have statistically similar behavior.

We used the AUC measure [17] to evaluate the classification performance because it is the most used measure for class imbalance problems [18, 19, 21, 22, 25].

4.2 Experimental Results

Figure 2 shows a CD diagram with the classification ranking for each contrast pattern-based classifier using all imbalanced databases described in Table 1.

From our results, the best contrast pattern-based classifier for detecting pneumatic failures on a temporary immersion bioreactor is HeDex, which statistically outperforms the remainder other classifiers used in our experiments. A possible explanation for this behavior is that HeDex uses randomization on both attribute selection and split-point selection, which can achieve high level of variety of decision trees. Furthermore, HeDex selects more than one split-point to

⁵ <http://www.sc.ehu.es/aldapa/2014/coverage-ctc>.

⁶ <http://www.keel.es>.

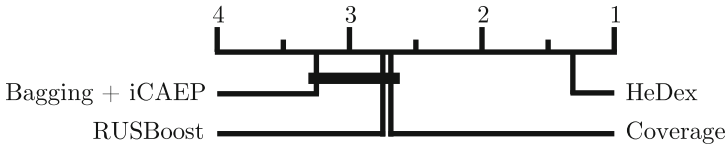


Fig. 2. CD diagram with a statistical comparison (using $\alpha = 0.10$) of the AUC results for contrast pattern-based classifiers over all the tested databases.

find a sub-optimal split-point that attains better results [19]. Also, HeDex uses the Hellinger distance as a decision tree splitting criterion, which has been widely used to deal with the class imbalance problem [6].

On the other hand, Coverage is based on a resampling strategy using multiple subsamples, which could exclude some representative objects to train the classifier. Additionally, Coverage uses a decision tree induction algorithm (CTC) which does not use a skew-insensitive splitting criteria. In the case of RUSBoost, it uses an undersampling method (RUS), which could exclude some important objects. Finally, bagging miner jointly with iCAEP obtained the worst AUC results, which could be attributed to the lower support of the patterns belonging to the minority class (failure).

We analyzed the patterns extracted from all databases using the HeDex algorithm where the AUC results were equal to 1.0. Among them, we can see, several times, the following patterns:

- (i) Air pressure into the central distribution line is lower than or equal to 0.104
- (ii) Air pressure into the plant container is lower than or equal to 0.136 and air pressure into the central distribution line is greater than 0.127

These patterns indicate to the expert that the temporary immersion bioreactor has a failure in the pneumatic system. These contrast patterns were analyzed and confirmed, by the pneumatic expert in the temporary immersion bioreactor, as useful patterns to classify failures in this temporary immersion system. The explanation issued by the expert for the first pattern is that the air compressor does not operate correctly or the central distribution line is broken. For the second pattern the explanation is that the plant container has an air escape. Accordingly, the PLC was reprogrammed including these patterns, which will forewarn failures in the temporary immersion bioreactor.

Finally, it is important to highlight that the use of contrast pattern-based classifiers for detecting pneumatic failures on temporary immersion bioreactors was qualified by the experts as meaningful and very useful, since it is very difficult for them to find these patterns manually.

5 Conclusions and Future Work

The main contribution of this paper is an empirical study of the use of contrast pattern-based classifiers for detecting pneumatic failures on temporary immersion bioreactors.

From our study, we can conclude that HeDex obtains the best AUC results for detecting pneumatic failures on temporary immersion bioreactors. Statistical tests prove that the differences among HeDex and the other tested contrast pattern-based classifiers are statistically significant. Additionally, through our study, we find useful patterns which were analyzed by a pneumatic system expert and these patterns were introduced into the PLC to forewarn failures in the temporary immersion bioreactor.

Finally, as future work, following the same approach presented in this paper, we will extend our study to other types of failures that can occur in temporary immersion bioreactors, e.g. if the liquid medium container is empty or if the plant grow rate is inappropriate. These studies would help to improve the plant quality in temporary immersion bioreactors.

Acknowledgment. This work was partly supported by National Council of Science and Technology of Mexico under the scholarship grant 370272. Also, the authors want to thank the laboratory members for cell culture and tissues at *Centro de Bioplasmas* for their valuable suggestions related to the plant micropropagation and the pneumatic system in the temporary immersion bioreactor, which significantly improved the quality of our work.

References

1. Andruszkiewicz, P.: Lazy approach to privacy preserving classification with emerging patterns. In: Ryzko, D., Rybiński, H., Gawrysiak, P., Kryszkiewicz, M. (eds.) *Emerging Intelligent Technologies in Industry*. SCI, vol. 369, pp. 253–268. Springer, Heidelberg (2011)
2. Barretto, S., Michoux, F., Nixon, P.J.: Temporary immersion bioreactors for the contained production of recombinant proteins in transplastomic plants. In: MacDonald, J., Kolotilin, I., Menassa, R. (eds.) *Recombinant Proteins from Plants*. MMB, vol. 1385, pp. 149–160. Springer, Heidelberg (2016)
3. Bolton, W.: *Programmable logic controllers*, 6th edn. Newnes, Bolton (2015)
4. Buchanan, W.J.: Modbus. In: *The Handbook of Data Communications and Networks*, vol. 1, pp. 677–687. Springer, Heidelberg (2004)
5. Chen, L., Dong, G.: Using emerging patterns in outlier and rare-class prediction. In: *Contrast Data Mining: Concepts, Algorithms, and Applications*, chap. 12, pp. 171–186. *Data Mining and Knowledge Discovery Series*, Chapman & Hall/CRC (2012)
6. Cieslak, D., Hoens, T., Chawla, N., Kegelmeyer, W.: Hellinger distance decision trees are robust and skew-insensitive. *Data Min. Knowl. Disc.* **24**(1), 136–158 (2012)
7. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**, 1–30 (2006)
8. Derrac, J., García, S., Molina, D., Herrera, F.: A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm Evol. Comput.* **1**(1), 3–18 (2011)

9. Dong, G.: Overview of results on contrast mining and applications. In: Contrast Data Mining: Concepts, Algorithms, and Applications, chap. 25, pp. 353–362. Data Mining and Knowledge Discovery Series, Chapman & Hall/CRC, United States of America (2012)
10. Escalona, M., Lorenzo, C.J., González, B., Daquinta, M., González, J., Desjardins, Y., Borroto, G.C.: Pineapple (ananas comosus l. merr) micropropagation in temporary immersion systems. *Plant Cell Rep.* **18**(9), 743–748 (1999)
11. Escalona, M., Lorenzo, J., González, B., Daquinta, M., Fundora, Z., Borroto, C., Espinosa, P., Espinosa, D., Arias, E., Aspiolea, M.: New system for in-vitro propagation of pineapple (ananas comosus (l.) merr). *Trop. Fruits Newsl.* **29**, 3–5 (1998)
12. Escalona, M., Samson, G., Borroto, C., Desjardins, Y.: Physiology of effects of temporary immersion bioreactors on micropropagated pineapple plantlets. *Vitro Cell. Dev. Biol. Plant* **39**(6), 651–656 (2003)
13. Etienne, H., Berthouly, M.: Temporary immersion systems in plant micropropagation. *Plant Cell Tissue Organ Culture* **69**(3), 215–231 (2002)
14. Freund, Y., Schapire, R.E., et al.: Experiments with a new boosting algorithm. In: 13th International Conference on Machine Learning (ICML 1996). vol. 96, pp. 148–156 (1996)
15. García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A.: Finding the best diversity generation procedures for mining contrast patterns. *Expert Syst. Appl.* **42**(11), 4859–4866 (2015)
16. García-Borroto, M., Martínez-Trinidad, J., Carrasco-Ochoa, J.: A survey of emerging patterns for supervised classification. *Artif. Intell. Rev.* **42**(4), 705–721 (2014)
17. Huang, J., Ling, C.X.: Using AUC and accuracy in evaluating learning algorithms. *Knowl. Data Eng. IEEE Trans.* **17**(3), 299–310 (2005)
18. Iburguren, I., Pérez, J.M., Muguerza, J., Gurrutxaga, I., Arbelaitz, O.: Coverage-based resampling: Building robust consolidated decision trees. *Knowl. Based Syst.* **79**, 51–67 (2015)
19. Kang, S., Ramamohanarao, K.: A robust classifier for imbalanced datasets. In: Tseng, V.S., Ho, T.B., Zhou, Z.-H., Chen, A.L.P., Kao, H.-Y. (eds.) PAKDD 2014, Part I. LNCS, vol. 8443, pp. 212–223. Springer, Heidelberg (2014)
20. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* **250**, 113–141 (2013)
21. López, V., Fernández, A., Herrera, F.: On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed. *Inf. Sci.* **257**, 1–13 (2014)
22. Loyola-González, O., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., García-Borroto, M.: Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases. *Neurocomputing* **175**, 935–947 (2016). Part
23. Moreno-Torres, J.G., Saez, J.A., Herrera, F.: Study on the impact of partition-induced dataset shift on k-fold cross-validation. *Neural Netw. Learn. Syst. IEEE Trans.* **23**(8), 1304–1312 (2012)
24. Paek, K.Y., Chakrabarty, D., Hahn, E.J.: Application of bioreactor systems for large scale production of horticultural and medicinal plants. In: Hvoslef-Eide, A.K., Preil, W. (eds.) Liquid Culture Systems for in vitro Plant Propagation, pp. 95–116. Springer, Heidelberg (2005)
25. Pérez, J.M., Muguerza, J., Arbelaitz, O., Gurrutxaga, I., Martín, J.I.: Combining multiple class distribution modified subsamples in a single tree. *Pattern Recogn. Lett.* **28**(4), 414–422 (2007)

26. Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A.: RUSBoost: A hybrid approach to alleviating class imbalance. *Syst. Man Cybern. Part A: Syst. Hum. IEEE Trans.* **40**(1), 185–197 (2010)
27. Tsai, C.H., Chang, L.C., Chiang, H.C.: Forecasting of ozone episode days by cost-sensitive neural network methods. *Sci Total Environ.* **407**(6), 2124–2135 (2009)
28. Weiss, G.M., Tian, Y.: Maximizing classifier utility when there are data acquisition and modeling costs. *Data Min. Knowl. Discovery* **17**(2), 253–282 (2008)
29. Xue, J., Hu, C., Wang, K., Ma, R., Zou, J.: Metamorphic malware detection technology based on aggregating emerging patterns. In: *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, pp. 1293–1296 (2009)
30. Zhang, X., Dong, G.: Information-based classification by aggregating emerging patterns. In: Leung, K.-S., Chan, L., Meng, H. (eds.) *IDEAL 2000. LNCS*, vol. 1983, pp. 48–53. Springer, Heidelberg (2000)