Analysis of one-vs-all versus one-vs-one approaches in lithofacies classification

Gallileu Genesis[1], Igor Fernandes Gomes[2], José Antonio Barbosa[3]
[12]LMCG, Department of Civil Engineering, UFPE.
gallileu.genesis@ufpe.br; igor.fernandes@ufpe.br
Av da Arquitetura, s/n, Cidade Universitária, Recife-PE, 50740540
[3]GEOQUANTT, Department of Geology, UFPE.
Av da Arquitetura, 954-995, Cidade Universitária, Recife-PE
jose.antoniob@ufpe.br

Abstract. The lithofacies classification provides valuable information about the geological history and the depositional environment. For multiclass classification tasks, as is usually the case with lithofacies classification, there are two basic approaches for building the models: the One-vs-One (OvO) and the One-vs-All (OvA) approaches. In general, when building models, default settings are typically employed, without conducting a detailed study to determine the most suitable approach for each scenario. In this study, we evaluated the performance of the OvA and OvO in the lithofacies classification task. The results indicate that the OvA model has lower computational cost and consistently outperforms the OvO model across almost cross-validation folds. Results on test data indicate that OvA can be a good alternative in scenarios with imbalanced classes or when there is a limited amount of data for training and testing. The OvO model can be an interesting alternative when the imbalance between classes is an important factor for the problem.

Keywords: Lithofacies classification, Multiclass classification, One-vs-One, One-vs-All, XGBoost.

1 Introduction

Lithofacies classification plays a crucial role in understanding the geology and depositional environment within sedimentary formations, providing valuable insights for constructing computational models in oil reservoirs [1][2][3]. In recent years, the application of machine learning models for this task has garnered significant attention [4][5][6]. However, this task presents several challenges due to limited labeled data, complex relationships between input variables and lithofacies, scale dependence, heterogeneity of the porous medium, and the number of labels to classify [7][8].

For multiclass problems, such as lithofacies classification, two fundamental approaches are commonly employed: the One-vs-One (OvO) [9] approach, where binary classifiers are trained for each pair of classes and directly compared, and the One-vs-All (OvA) [10] or One-vs-Rest (OvR) approach, where binary classifiers are trained for each class against all other classes. Despite the widespread use of open-source libraries for model construction, the selection of the classification strategy is often overlooked, with default configurations being applied without considering their impact on the model's performance for different types of applications.

To explore this question, this study was conducted with the aim of evaluating the performance of the OvA and OvO approaches in lithofacies classification. The Geophysical Tutorial Machine Learning Contest 2016 dataset [11][12] was employed for this purpose. We chose XGBoost as the base model, with two versions configured: one using the OvO approach and the other using the OvA approach. Hyperparameter optimization was performed using the Optuna library, with F1-score as the optimization metric.

The study aimed to provide insights into the performance of the OvA and OvO approaches in lithofacies classification using machine learning models and contribute to the optimization of classification strategies for building more accurate computational models in oil reservoir exploration.

2 Methodology

2.1 Data source and preprocessing

The dataset comprises measurements from 10 wells, including Gamma ray (GR), Resistivity logging (ILD_log10), Photoelectric effect (PE), Neutron-density porosity difference (DeltaPHI), Average neutron-density porosity (PHIND), Nonmarine/marine indicator (NM_M), and relative position (RELPOS). Prior to analysis, the dataset was preprocessed to handle missing values, normalize features, and ensure data consistency.

The nine lithofacies present in the database, along with their respective codes, are shown in Tab. 1.

Table 1: Lithofacies present in the database.

Lithofacie	Código
Nonmarine sandstone	SS
Nonmarine coarse siltstone	CSiS
Nonmarine fine siltstone	FSiS
Marine siltstone and shale	SiSh
Mudstone (limestone)	MS
Wackestone (limestone)	WS
Dolomite	D
Packstone-grainstone (limestone)	PS
Phylloid-algal bafflestone (limestone)	BS

2.2 Model selection and configuration

XGBoost [13], a popular machine learning algorithm, renowned for its effectiveness in handling complex classification tasks, was selected as the foundational model for this study. Two versions of the model were configured: one based in the OvO approach and the other using the OvA approach. The construction of these approaches can be summarized as follows:

- **OvA**: a binary classifier is trained for each class in the problem. In other words, the model trains one class against all other classes combined. To classify a new instance, the OvA model compares the outputs of all the classifiers and assigns the instance to the class that has the highest confidence score.

- **OvO**: a binary classifier is trained for each pair of classes. If there are N classes, the number of required binary classifiers is N*(N-1)/2. Once again, to classify a new instance, the OvO classifiers vote, and the class with the most votes is assigned to the instance.

2.3 Hyperparameter tuning

The hyperparameters tuning was performed using Optuna library [14]. Optuna uses a Sequential Model-Based Optimization (SMBO) strategy to efficiently search the hyperparameter space. One hundred trials were conducted, and the F1-score was selected as the optimization metric to strike a balance between Precision and Recall.

2.4 Performance evaluation

Following the competition guidelines, nine wells (Alexander D, Shankle, Luke G U, Kimzey A, Cross H Cattle, Nolan, Recruit F9, Newby, Churchman Bible) were used for training, while one well (Shrimplin) served as the test set. Performance evaluation of the models involved two distinct phases. First, a 5-fold cross-validation was conducted using the training set to assess the models' performance on different subsets of data. Metrics such as accuracy, precision, recall, and F1-score were calculated for each fold. Second, a blind test on the complete well was carried out to evaluate the models' performance on new data. The confusion matrix was utilized to assess the classification results of the blind test.

3 Results and discussion

3.1 Cross-validation results

Table 2 presents the hyperparameter tuning times in hours and cross-validation times in minutes.

Table 2. Hyperparameter tuning and cross-validation times for the OvA and OvO models.

Model	Tuning (h)	Cross-validation (min)
OvA	2,1	3.3
OvO	3,8	4.0

The OvO model has a computational cost that is approximately 81% higher for hyperparameter tuning and around 21% higher for cross-validation, when compared to the OvA model. These differences stem from the distinct computational complexities of the models. In this problem, we have 9 classes, which means the OvA model requires only 9 classifiers (one for each class). On the other hand, OvO necessitates a total of 36 binary classifiers. This disparity makes both training and inference for the OvO model more resource-intensive compared to the OvA model.

Figure 1 illustrates the comparison of accuracy, precision, recall, and f1_score scores for the OvA and OvO models across each fold of cross-validation.

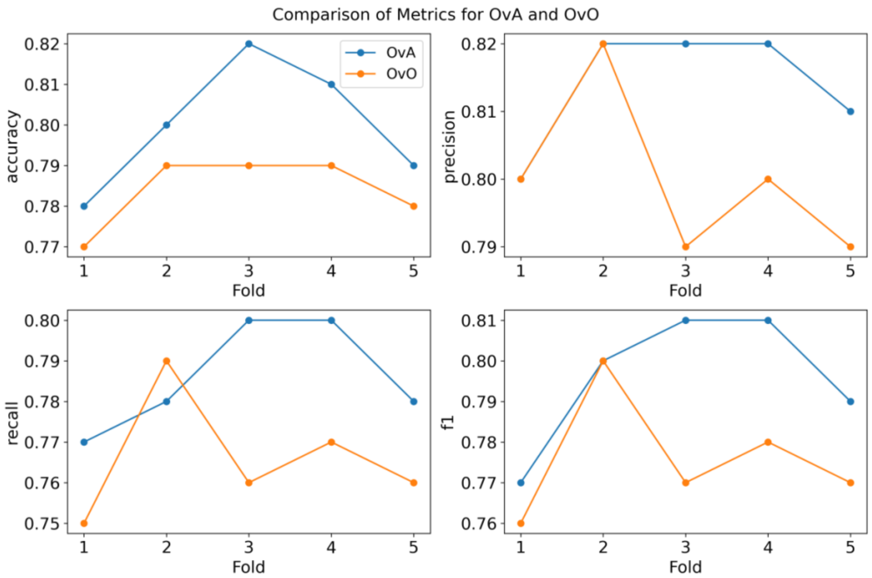

Figure 1: Accuracy, precision, recall, and F1-score results for each of the 5 cross-validation folds.

It can be noted that the OvA model consistently outperforms the OvO model in most of the cross-validation folds. This performance difference can be attributed to the impact of class imbalance during model construction: due to its construction characteristics, the OvO model tends to face greater challenges with problems where there is an imbalance between classes, as is the case in the problem addressed in this study. This is because the number of samples in each class determines the amount of training data for the model, leading to inadequate learning for rare classes. This is not the case in the OvA model construction, where each binary classifier focuses solely on one class versus all others. The presence of significant overlap between classes could be another explanatory factor for the results shown in Fig. 1. In such cases, the OvA model may be more effective, as each classifier is trained to distinguish a specific class from the rest.

As cross-validation assesses model performance across multiple subsets of data, ensuring that evaluation is unbiased due to a specific train-test split. The consistent superior performance of the OvA model across most folds indicates a more robust and stable performance compared to the OvO model. This suggests that the OvA model is better suited for this specific dataset and classification task, making its selection as the applied model for problem-solving reasonable.

3.2 Results on test data

The performance results of the models in classifying the Shrimplin well are presented in Tab. 3 and Fig. 2. The summary of evaluation metrics for both models for each lithofacies is provided in Tab. 3, while the distribution of actual and predicted classes in the confusion matrix is depicted in Fig. 2.

Table 3: Results of Precision, Recall, and F1-score for the classification of all lithofacies by both models in the Shrimplin well.

Lithofacie	Precision		Recall		F1-score		Support
	OvA	OvO	OvA	OvO	OvA	OvO	
SS	0.00	0.00	1.00	1.00	0.00	0.00	0
CSiS	0.63	0.67	0.70	0.75	0.66	0.70	118
FSiS	0.70	0.74	0.59	0.61	0.64	0.67	123
SiSh	0.57	0.42	0.72	0.56	0.63	0.48	18
MS	0.58	0.64	0.11	0.22	0.19	0.33	63
WS	0.50	0.53	0.81	0.83	0.62	0.65	63
D	0.50	0.00	0.40	0.00	0.44	0.00	5
PS	0.67	0.59	0.58	0.49	0.62	0.54	69
BS	0.40	0.39	1.00	1.00	0.57	0.56	12

The first point to note is that both models classified the SS class, even though it is not present in the Shrimplin well. When cross-referencing this information with Fig. 2, it can be observed that both models classify the CSiS class as SS, with 4 errors made by the OvA model and 3 by the OvO model. This suggests overfitting during the training phase for this class, most likely due to overlap between them.

For the CSiS lithofacies, the OvA model achieved a precision of 0.63 and a recall of 0.70, resulting in an F1-score of 0.66. Similar F1-score values are observed in the FSiS, SiSh, WS, and PS classes. The lowest F1-score (0.19) was obtained in the MS class. The overall accuracy of the model was 0.60. The macro averages for precision, recall, and F1-score were 0.51, 0.66, and 0.49, respectively. The weighted averages for the same metrics were 0.62, 0.60, and 0.58.

The OvO model achieved a precision of 0.67 and a recall of 0.75 for the CSiS lithofacies, resulting in an F1-score of 0.70, which is a superior performance compared to OvA. The same pattern holds for the FSiS, MS, and WS classes, with F1-scores of 0.67, 0.33, and 0.65, respectively. However, the OvO model's performance was inferior to OvA in the SiSh, D, PS, and BS classes. The overall accuracy of the OvO model was 0.61. The macro averages for precision, recall, and F1-score were 0.44, 0.61, and 0.44, respectively. The weighted averages were 0.63, 0.61, and 0.59.

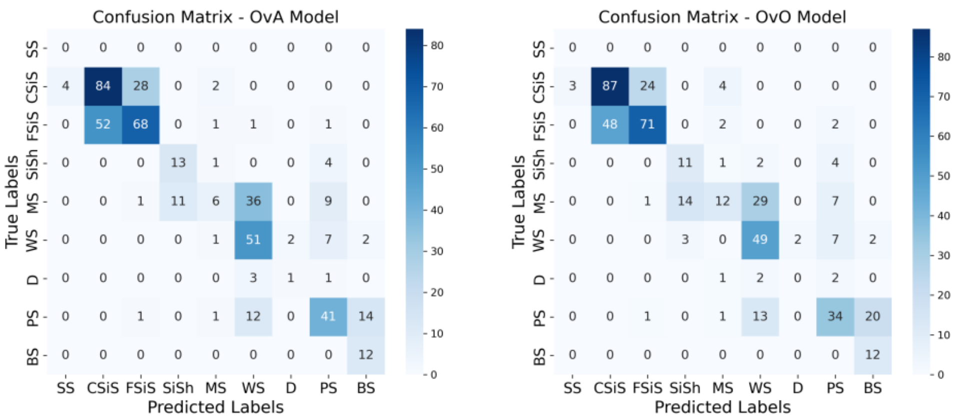

Figure 2: Confusion matrix of both models in Shrimplin well.

The summary of model performance can be conducted through a comparison of macro and weighted averages. The results of the macro averages show that the OvA model performs better overall across all classes, without considering the class imbalance. In other words, when each class is considered equally important, regardless of its instance count. Thus, if the goal is to ensure that the model performs well for each class independently, the OvA model is indeed the more suitable choice.

The models exhibit similar performance when analyzing the weighted averages. In this case, the class imbalance is taken into account, weighing the performance of each class by the number of instances it has (the model gives more weight to classes with higher support).

4 Conclusions

The OvA model demonstrated better computational efficiency for hyperparameter tuning and training. This makes it suitable for problems where computational resources or available time are critical factors. Additionally, OvA can be a good alternative in scenarios with imbalanced classes or when there is a limited amount of training and testing data. On the other hand, the OvO model can be an interesting choice when class imbalance is a crucial factor for the problem.

These findings underscore the critical importance of carefully analyzing the choice of classification strategy when building models to optimize their performance.

Acknowledgements. This study was supported by CNPq/Brazil, Energi Simulation, and Petrobras that funded the projects “Geomecarb” and “Pseudo-poços” through research agreements approved by the National Agency of Petroleum, Natural Gas and Biofuels (ANP).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] X. Yu, S. Li, S. Zhao, J. Chen, G. Hou. Constraining Method of Stochastic Modeling for Fluvial Petroleum Reservoir Controlled by Depositional Facies Using Wells and Seismic Data. *Earth Science Frontiers*, 15(07), pp. 33–41, 2008. https://10.1016/S1872-5791(08)60037-8.

[2] D. Grana, X. Lang, W. Wu. Statistical facies classification from multiple seismic attributes: comparison between Bayesian classification and expectation-maximization method and application in petrophysical inversion: Statistical facies classification. *Geophysical Prospecting*, 65(08), 2016. https://doi.org/10.1111/1365-2478.12428.

[3] L. Zhao, J. Geng, J. Cheng, H. Han, G. Tonglou. Probabilistic lithofacies prediction from prestack seismic data in a heterogeneous carbonate reservoir. *Geophysics*, 79(08), M25–M34, 2014. https://doi.org/10.1190/geo2013-0406.1.

[4] G. Li, Y. Qiao, Y. Zheng, Y. Li, W. Wu. Semi-Supervised Learning Based on Generative Adversarial Network and Its Applied to Lithology Recognition. *IEEE Access*, 7, 67428 – 67437, 2019. https://doi.org/10.1109/ACCESS.2019.2918366.

[5] S. Asghar, J. Choi, D. Yoon, J. Byun. Facies classification using semi-supervised deep learning with pseudo-labeling strategy. *SEGTechnical Program Expanded Abstracts*, 08, 3171–3175, 2019. https://doi.org/10.1190/segam2019-3216086.1.

[6] M. Dunham, A. Malcolm, K. Welford. Improved Well-log classification using semisupervised label propagation and self-training, with comparisons to popular supervised algorithms. *Geophysics*, 85(01), 1–15, 2020. https://doi.org/10.1190/geo2019-0238.1.

[7] M. Liu, M. Jervis, W. Li, P. Nivlet; Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks. *Geophysics,* 85 (4), O47–O58, 2020. doi: https://doi.org/10.1190/geo2019-0627.1.

[8] A. Saleem, J.Choi, D Yoon, and B. Joongmoo. Facies classification using semi-supervised deep learning with pseudo-labeling strategy. *Paper presented at the SEG International Exposition and Annual Meeting*, San Antonio, Texas, USA, September 2019. doi: https://doi.org/10.1190/segam2019-3216086.1.

[9] J. Fürnkranz. Round robin rule learning, *In: Brodley CE, Danyluk AP, editors. ICML-01. Proceedings of the 18th International Conference on Machine Learning*. Williamstown, MA: Morgan Kauffman; 2001. P.146–153.

[10] R. Rifkin, A. Klautau. In defense of one-vs-all classification. *J Mach Learn Res*;5:101–141, 2004

[11] B. Hall, Facies classification using machine learning. *The Leading Edge,* 35, 906–909, 2016. doi: https://10.1190/tle35100906.

[12] Hall, M., and B. Hall, 2017, Distributed collaborative prediction: Results of the machine learning contest. *The Leading Edge*, 36, 267–269, doi: https://10.1190/tle36030267.1.

[13] T. Chen and C. Guestrin. Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, KDD’16. ACM, New York, pp 785–794, 2016.

[14] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A Next-Generation Hyperparameter Optimization Framework. *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631, 2019.