Selection of Augmented Data for Overcoming the Imbalance Problem in Facies Classification

Dowan Kim and Joongmoo Byun

***Abstract*— Facies classification refers to the classification of rock types and pore fluids using information obtained from well log data and core samples. A range of elastic properties provide the main input for classification models. The elastic properties are closely related to water saturation, porosity, and shale volume. In addition, if impedance inversion is performed, the same elastic properties can be obtained from the surface seismic area, thus linking well log and surface seismic data. Machine learning (ML)-based facies classification has the advantage of minimizing the subjectivity associated with human interpretations and maximizing the time efficiency. However, due to the insufficiency of well log data, class imbalance and absolute data shortages can easily arise. Therefore, in this study, we used a cycle-consistent generative adversarial network (CycleGAN) to augment the synthetic data simulating well log data. In addition, we determined which classes of data required augmentation when using CycleGAN and proposed criteria for selecting the augmented data to be used for class-balanced training. The developed algorithm was verified using the Vincent oil field data. The classification results were improved, and more physically valid predictions were achieved in the surface seismic survey area. The data augmentation scheme developed in this study will be useful for facies classification in environments where well log data are very limited.**

***Index Terms*— Class imbalance problem, data augmentation, facies classification, neural networks.**

I. Introduction

FACIES classification involves the identification of the facies of subsurface rock based on mineral composition and pore fluids. Since facies classification results assist in understanding the sedimentary environment, they play an important role in reservoir characterization.

The changes in the characteristics of rock types and pore fluids have often been estimated from the changes in amplitude or spectral components [1]–[3]. These approaches are convenient, but also have the limitation of being qualitative [4]. When well log data are available, a range of elastic properties can provide good input attributes for quantitative seismic interpretation [5]–[8]. Such elastic properties are directly related to facies classifications that vary according to water saturation, porosity, and shale volume. In addition, if an impedance inversion or amplitude variation with offset (AVO) inversion is performed using prestack seismic data, the same elastic properties can be obtained in the seismic survey area, thus serving as a bridge between well log data and surface seismic survey data.

Recently, due to advancements in machine learning (ML) methods, studies have applied ML methods to facies classification. Jin [9] compared the performance of various ML techniques for facies classification. Ashgar *et al.* [10] applied a semi-supervised learning technique that uses both labeled and unlabeled data around well locations. Liu *et al.* [11] suggested a method for facies classification and reservoir prediction using multikernel relevance vector machine. Zhang *et al.* [12] proposed an ensemble learning method with encoder and decoder networks.

However, using ML approaches to perform facies classification can cause a severe class imbalance problem. The class imbalance problem occurs when there is a large difference in the amount of data among different classes in the training dataset. In many cases, it is difficult to build ideally class-balanced training datasets in facies classification. In such cases, a heavily weighted model may be trained with majority classes.

Over-sampling techniques for improving training data have been studied with the aim of alleviating the negative effects of imbalanced training datasets. Over-sampling techniques balance classes by adding synthetic data to minority classes. Synthetic minority over-sampling technique (SMOTE) [13], a representative over-sampling approach, generates new data between neighboring minority class samples. Borderline-SMOTE [14] and adaptive synthetic sampling (ADASYN) [15] are improved SMOTE techniques in which over-sampling is performed only at the boundaries between classes, to avoid augmentation of noise or unhelpful data. However, these methods generate data showing only monotonous changes, which can easily lead to overfitting.

In recent years, data augmentation techniques using a generative adversarial network (GAN) [16] have attracted significant research attention [17], [18]. A GAN is suitable for application to data augmentation, because it trains a generative model that creates synthetic data similar to real data. In particular, the cycle-consistent GAN (CycleGAN) [18] can be used for image-to-image translation, which includes generative models that can produce specific classes of data using data from another class. Therefore, Kim and Byun [19] generated synthetic samples of minority classes using majority class samples by applying CycleGAN to solve the imbalance problem in facies classification. However, this work has a limitation that it does not provide a standard for which classes and how many synthetic data should be generated in a multiclass classification problem.

Manuscript received May 7, 2021; revised July 1, 2021 and July 27, 2021; accepted August 8, 2021. Date of publication August 17, 2021; date of current version January 6, 2022. This work was supported by Korea Institute of Energy Technology Evaluation and Planning funded by Korea Government (MOTIE) under Grant 20182510102470 and Grant 20194010201920. *(Corresponding author: Joongmoo Byun.)*

The authors are with the Reservoir Imaging with Seismic and EM Technology. Machine Learning (RISE.ML) Laboratory, Hanyang University, Seoul 04763, South Korea (e-mail: kdw9492@hanyang.ac.kr; jbyun@hanyang.ac.kr).

Digital Object Identifier 10.1109/LGRS.2021.3103997

When applying over-sampling, there are two important considerations. First, the appropriate amount of synthetic data must be determined. If too little data are synthesized, it is difficult to solve the imbalance problem sufficiently, whereas if too much data are synthesized, overfitting to minority classes will occur [15]. Second, synthetic data suitable for training must be selected. If synthetic data that are insufficiently similar to the original data are added to the training data, the training process will be adversely affected, including through the introduction of noise. Conversely, synthetic data too similar to the original data can easily lead to overfitting.

Therefore, in this article, we proposed a technique to identify suitable synthetic data generated using CycleGAN. The contributions of this study include: 1) designing new evaluation metrics to quantitatively determine how much imbalance exists in the original dataset and 2) developing an iterative data augmentation and selecting algorithm to determine the appropriate synthetic data to be added to each class.

II. Methodology

A. Rock- and Pore-Fluid-Type Substitution Using CycleGAN

CycleGAN is a method for image-to-image translation using GANs. Image-to-image translation is a field of computer vision and graphics that can map input images to paired output images. CycleGAN consists of three loss terms; the first and second terms of these are as follows [18]:

$$L_{\text{total}} = L_{x\to y} + L_{y\to x} + L_{\text{cyc}} \tag{1}$$

$$L_{x\to y} = \boldsymbol{E}_x\left[\log D_x(x)\right] + \boldsymbol{E}_y\left[\log\left(1 - D_x\left(G_{yx}(y)\right)\right)\right] \tag{2}$$

$$L_{y\to x} = \boldsymbol{E}_y\left[\log D_y(y)\right] + \boldsymbol{E}_x\left[\log\left(1 - D_y\left(G_{xy}(x)\right)\right)\right] \tag{3}$$

$$L_{\text{cyc}} = \boldsymbol{E}_x\left\|G_{yx}\left(G_{xy}(x)\right) - x\right\|_1 + \boldsymbol{E}_y\left\|G_{xy}\left(G_{yx}(y)\right) - y\right\|_1 \tag{4}$$

where G_{ab} is the generator that converts the domain of class A into that of class B, and D_b is a discriminator that distinguishes between the original data belonging to class B and the synthetic data generated through G_{ab}. As a result, G and D have an adversarial relationship, so the first and second loss terms have two-way adversarial forms.

The third loss term is known as the cycle-consistency loss. This allows for full recovery of the synthetic data generated using G when returned to the original class through the "reverse direction of G." If cycle-consistency loss is not used, the generators may produce only the most plausible output. This form of failure is called mode collapse. However, cycle-consistency loss places restrictions on translation, which prevents mode collapse.

In this study, we applied the CycleGAN for fluid- or rock-type substitution in a data augmentation process for facies classification. The labeled data in this study are well log data, which are 1-D arrays along the depth (or time) direction; thus, our CycleGAN model was based on a 1-D CNN structure. The generators were designed to generate output patches through three encoding and decoding steps, and the discriminators were configured to determine the output patches.

By training a two-way translation between wet sand and oil sand, we simulated substitution between fluid classes. Similarly, by training a translation between wet sand and shale, we simulated substitution between rock types. We wanted to perform over-sampling by adding the synthetic data transferred from the majority classes to the minority classes. Example CycleGAN results obtained by translating wet sand into the shale class are shown in Fig. 1. Fig. 1(a) shows a patch of wet sand in the original training dataset. Each column represents a scaled elastic property. In this study, the patch size was fixed at five samples, and seven elastic properties were used as inputs for training. The elastic properties correspond to the P-impedance (AI), S-impedance (SI), lambda-rho ($\lambda\rho$), mu-rho ($\mu\rho$), Poisson's ratio (ν), velocity ratio (V_P/V_S), and E-rho ($E\rho$) values, which were extracted from the well log data. Here, λ is the first Lamé's parameter, ρ is the density, μ is the shear modulus, V_P is the P-wave velocity, V_S is the S-wave velocity, and E is the Young's modulus. Fig. 1(b) presents the result obtained by converting the wet sand patch into the shale class using CycleGAN, and Fig. 1(c) shows the reconstructed wet sand patch (converted from the synthetic shale patch). The reconstructed data are almost identical to the original data because of the use of the cycle-consistency loss.

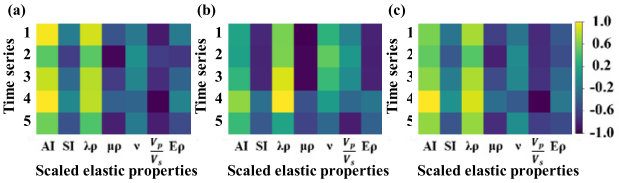

Fig. 1. (a) Patch belonging to the wet sand class in the original dataset. (b) Result obtained by converting the wet sand patch into the shale class using CycleGAN. (c) Reconstructed wet sand patch converted from the synthetic shale patch.

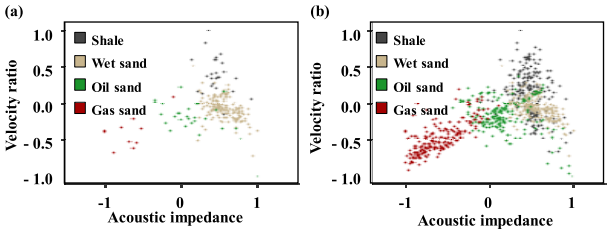

Fig. 2. Crossplots of P-impedance versus velocity ratio obtained using (a) original dataset and (b) augmented synthetic data.

We used the synthetic data [Fig. 1(b)] created using these trained generators as augmented data. Fig. 2 shows acoustic impedance versus velocity ratio crossplots of the original well log data and augmented synthetic data. Large amounts of synthetic data with sufficient diversity compared with the original data were obtained.

B. Evaluation Metrics Considering the Imbalance Problem

It is important to select appropriate evaluation metrics for a given ML problem. In particular, because we wanted to address the class imbalance problem, it was necessary to quantify the extent of the imbalance in our original dataset. We also needed to ascertain whether the class imbalance was resolved through data augmentation.

The metrics used most commonly to evaluate the performance of classification models are accuracy, precision, and recall. Accuracy refers to the percentage of correct predictions across the entire dataset. Precision refers to the percentage of correct answers predicted for a particular class of data, and

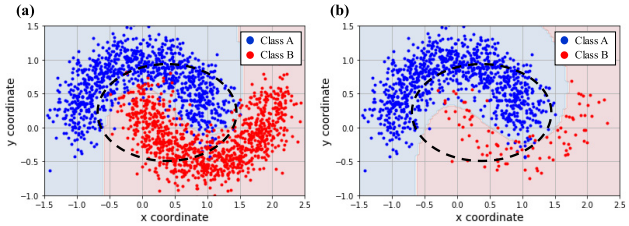

Fig. 3. Comparison of the training classification models when (a) there is no imbalance between classes and (b) there is an imbalance problem. There is a large difference in the decision boundary in the area outlined with the black dashed ellipse.

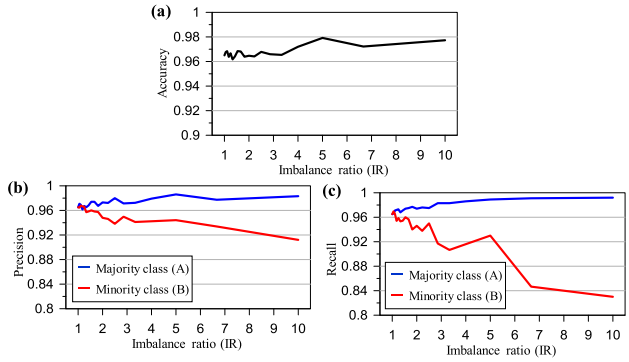

Fig. 4. Change in (a) accuracy, (b) precision, and (c) recall with increasing IR.

recall refers to the percentage of successful predictions for a particular class. Because the three metrics have different meanings, it is common to use accuracy, precision, and recall together to conduct a composite evaluation.

However, if there is a class imbalance problem, accuracy and precision, unlike recall, are difficult to be used as evaluation metrics. Because accuracy and precision are expressed as a ratio between different classes, the results show a bias toward the majority class. Fig. 3 shows how classification outcomes differ depending on whether there is an imbalance problem. The dataset used in Fig. 3 is the two-moon dataset, provided by Scikit-learn. The decision boundaries were determined using a support vector machine (SVM). In Fig. 3(a), there are 1000 samples in both Classes A and B, while in Fig. 3(b), Class A has 1000 samples, but Class B only has 100 samples. Fig. 3(b) shows that the decision boundary has shifted and predicts more Class A samples than in the case shown in Fig. 3(a). It shows that the decision boundaries may differ depending on whether there is a class imbalance even if the distribution characteristics are the same.

We observed how the accuracy, precision, and recall of the trained model changed as the number of samples in Class B of the two-moon dataset gradually decreased from 1000 to 100 (Fig. 4). The accuracy, precision, and recall are as follows [20]:

$$\text{Accuracy} = \frac{\sum_i^n M_{ii}}{\sum_i^n \sum_j^n M_{ij}} \tag{5}$$

$$\text{Precision}_i = \frac{M_{ii}}{\sum_j^n M_{ji}} \tag{6}$$

$$\text{Recall}_i = \frac{M_{ii}}{\sum_j^n M_{ij}} \tag{7}$$

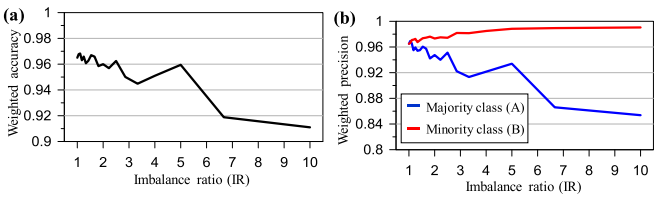

Fig. 5. Change in (a) weighted accuracy and (b) weighted precision with increasing IR.

where M_{ab} is the number of samples predicted as Class B that actually belong to Class A, and n is the number of classes.

The horizontal axis represents the imbalance ratio (IR), defined as the ratio of the number of samples in the majority class to the number of samples in the minority class. Fig. 4(c) shows that as the imbalance becomes more severe, the recall of the majority class rises slightly and the recall of the minority class decreases significantly. This means that as IR increases, the decision boundary is shifted and predicts more samples in the majority class, resulting in larger areas of misinterpretation. However, it is difficult to observe meaningful changes in precision graphs because the proportion of minority samples also decreases as the area predicted by minority classes gradually decreases in size. Thus, although the decision boundary gradually deteriorates, the precision of the majority class predictions tends to increase slightly. The same problem occurs with accuracy, which increases along with IR.

In such cases, accuracy and precision are not suitable evaluation metrics because they do not take the class imbalance problem into account. Nevertheless, many studies have used precision, or harmonic mean of precision and recall, as an evaluation index even if there is an imbalance between classes, because precision is important when evaluating performance [21]–[23].

This study seeks to develop evaluation metrics that can be used meaningfully even if there is an imbalance between classes. The greater the imbalance, the more the accuracy and precision are affected by the majority samples rather than by minority samples. Therefore, we designed weighted accuracy and weighted precision metrics that assign an IR weight to the minority classes as

$$\text{IR}_i = N/\big(\text{The number of samples in the } i^{\text{th}}\text{class}\big), \quad i \in [1, \ldots, n] \tag{8}$$

$$\text{Weighted accuracy} = \frac{\sum_i^n (M_{ii} \cdot \text{IR}_i)}{\sum_i^n \sum_j^n \left(M_{ij} \cdot \text{IR}_i\right)} \tag{9}$$

$$\text{Weighted precision}_i = \frac{M_{ii} \cdot \text{IR}_i}{\sum_j^n \left(M_{ji} \cdot \text{IR}_j\right)} \tag{10}$$

where N is the number of samples in the most majority class.

The weighted accuracy and weighted precision are the same as the original accuracy and precision if there is no class imbalance problem. However, if there is a class imbalance, IR is used as a weighting for the minority class. Fig. 5 shows the graphs of the variations in weighted accuracy and weighted precision according to IR in the two-moon dataset. Even though the IR increases and the decision boundary moves away from the ideal decision boundary, the original precision

of the majority class and accuracy have rather increased. On the other hand, the weighted precision demonstrates a trade-off between precision and recall [Figs. 4(c) and 5(b)]. Similarly, the weighted accuracy also decreases as IR increases, which provides a good representation of the fact that the greater the data imbalance, the further away from the ideal decision boundary.

C. Data Augmentation and Selection Strategy

When a class imbalance problem exists, the recalls of the minority classes will be low and the weighted precisions will be high. In contrast, the recalls of the majority classes will be high and the weighted precisions will be low. Based on this, the difference between the recall and weighted precision for each class was calculated, and the class was categorized as the majority or minority classes according to the difference. In this study, we categorized the classes with the difference greater than 0.03 as the majority classes and the classes with the difference less than -0.03 as the minority classes.

After categorizing all classes as majority or minority classes, we created synthetic data for the minority classes to solve the class imbalance problem. In this stage, generators of CycleGAN, which translate from majority to minority classes, were used. If the number of majority classes is determined as N and that of minority classes is K, the total number of generators used at this step is equal to $N \times K$. We generated as much synthetic data as possible using these generators, and the data to be added to the training dataset were selected from among these data with the following criteria.

First, the generated minority class data were applied to the previously trained classification model as inputs. In this step, the probability of the classification model correctly predicting minority class data was computed using the softmax function. A probability of less than or equal to 0.5 indicates that the characteristics of the generated data are insufficiently similar to the original data. Thus, the data added to the training dataset could act as noise, which may have an adverse effect on training. Conversely, if the probability is greater than 0.95, the data may be too similar to the original data. When these data are added to the training dataset, they may act as duplicates, which leads to overfitting. Therefore, in this study, only synthetic data with computed probabilities ranging from 0.5 to 0.95 were used. Then, some of these data were randomly extracted and added to the original training dataset. Finally, the classification model was retrained with the newly built training dataset, and these processes were repeated until the class imbalance problem was solved. Ultimately, the differences between recall and weighted precision of all classes can be minimized. As the amount of synthetic data added to the training dataset increases in a single iteration, the loss function converges more quickly. However, the risk of overfitting to the minor classes also increases. Thus, we added only five synthetic data samples to each minority class in a single iteration during training.

III. Field Data Example

To verify the proposed algorithm, we used seismic survey data obtained from the Vincent oil field located in Western Australia. Log data obtained from three wells (Vincent 1, Vincent 2, and Vincent 3) were used as training data. The input elastic properties used for training the classification model were the same as those used for CycleGAN training, including acoustic impedance, elastic impedance, lambda-rho, mu-rho, Poisson's ratio, velocity ratio, and E-rho. In addition, the surface seismic data in a 2-D arbitrary line passing through Vincent 1, 2, and 3 were also converted into the same elastic properties, using prestack impedance inversion [24], to predict the facies in the surface seismic survey area.

TABLE I
Evaluation Metrics When Training the Facies Classification Model Using the Imbalanced Original Dataset

	Class			
	Shale	Wet sand	Oil sand	Gas sand
Weighted precision	1.000	0.842	0.817	1.000
Recall	0.833	0.993	0.897	0.889
Weighted accuracy	0.903			

The facies type, which is classified based on rock types and pore fluids, of the rocks in the survey area was determined by interpreting the well log data. First, we classified the rock type as shale or sand based on a shale volume of 50%. Next, the sand was divided into wet sand and hydrocarbon sand based on a water saturation of 70%. Finally, based on the density logs, the hydrocarbon sand was divided into oil sand and gas sand. The well log data were divided into patches comprising five samples each, for use in the classification ML model formed of the CNN structures. A total of five hidden layers were included, and batch normalization and dropout were carried out for stable training. The output is divided into four categories, each of which corresponds to the facies. The shale, wet sand, oil sand, and gas sand classes contained 12, 153, 29, and 9 patches, respectively, indicating an imbalance problem between the classes. We used 60% of the patches as the training data (121 patches) and designated the remaining 40% as validation data (82 patches).

To train the classification ML model with the field data, five additional data augmentation and ML model retraining iterations were conducted after the initial model training using only the original dataset. Each iteration consisted of 200 epochs, and the best model was selected when the validation loss was minimal. The cross-entropy loss was used for training. At the end of each iteration, data augmentation and selection were conducted using CycleGAN.

Table I shows the evaluation metrics of the facies classification results of the well log data when the ML model was trained using only the original dataset. In this case, the weighted precision of the wet sand class, which has the largest amount of data, was significantly lower than the recall value. On the other hand, both weighted precision and recall of the oil sand class were measured to be low, as the oil sand class contained more complex imbalances. For example, the oil sand class represents a majority class compared with shale or gas sand classes, but a minority class compared with wet sand class.

The classification performance when the model was trained after data augmentation is shown in Table II. Compared with those in Table I, the recall values of the shale and gas sand classes, which were the minority classes, and the weighted

TABLE II

EVALUATION METRICS WHEN TRAINING THE FACIES CLASSIFICATION MODEL AFTER DATA AUGMENTATION

	Class			
	Shale	Wet sand	Oil sand	Gas sand
Weighted precision	0.935	0.967	1.000	1.000
Recall	1.000	1.000	0.897	1.000
Weighted accuracy	0.974			

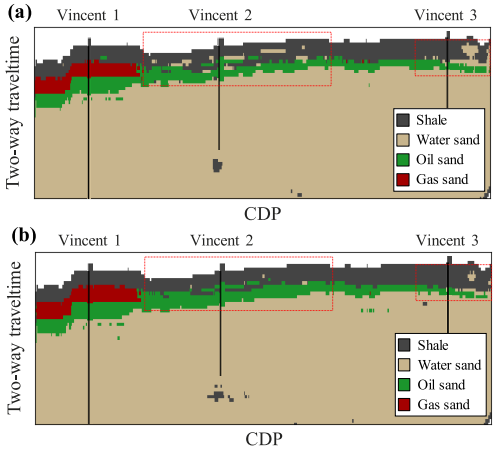

Fig. 6. Results of the facies classification in a seismic survey area. The facies are predicted using the trained model (a) before and (b) after data augmentation. The results in the areas outlined by red dotted boxes were noticeably improved.

precision of the wet sand class, which was the majority class, increased significantly. The weighted accuracy also improved.

Fig. 6 shows a comparison of the facies prediction results obtained in the surface seismic survey area using the ML model trained before and after data augmentation. As shown in the areas marked with red dotted boxes, the wet sand area is overestimated before applying the data augmentation; as a result, the boundaries between classes are blurred. However, data augmentation significantly alleviated the imbalance between classes, and the areas incorrectly predicted as wet sand were correctly predicted as shale or oil sand. As a result, the boundary between shale and oil sand shown in Fig. 6(b) is clearer than in Fig. 6(a).

IV. CONCLUSION

In this study, we developed an algorithm using CycleGAN for data augmentation and selection processes, to overcome the class imbalance problem. First, to measure the degree of class imbalance, evaluation metrics that consider class imbalance were developed. Because the proposed weighted accuracy and weighting precision metrics take the imbalance between classes into account, the more severe the imbalance between classes, the more effective the proposed evaluation metrics are compared with conventional accuracy and precision metrics. Next, we designed an algorithm classifying the classes as majority or minority classes. Data augmentation and selection (of data to be added to the training dataset) processes were then conducted using CycleGAN. The developed algorithm was verified using the Vincent oil field. From the results, we confirmed that degradation in the classification results due to the imbalance between classes was alleviated. The developed evaluation metrics and classification scheme can be extended to other classification schemes that use ML when there is an imbalance between classes in the training dataset.

ACKNOWLEDGMENT

The authors would like to thank the CGG for providing HampsonRussell (HRS) license of academic version.

REFERENCES

[1] M. T. Taner, F. Koehler, and R. E. Sheriff, "Complex seismic trace analysis," *Geophysics*, vol. 44, no. 6, pp. 1041–1063, 1979.
[2] D. Gao, "Volume texture extraction for 3D seismic visualization and interpretation," *Geophysics*, vol. 68, no. 4, pp. 1294–1302, Jul. 2003.
[3] S. Chopra and K. J. Marfurt, "Seismic attributes—A historical perspective," *Geophysics*, vol. 70, no. 5, pp. 1–28, Sep. 2005.
[4] D. Gao, "Latest developments in seismic texture analysis for subsurface structure, facies, and reservoir characterization: A review," *Geophysics*, vol. 76, no. 2, pp. W1–W13, Mar. 2011.
[5] B. Goodway, T. Chen, and J. Downton, "Improved AVO fluid detection and lithology discrimination using Lamé petrophysical parameters: '$\lambda\rho$', '$\mu\rho$', & 'λ/μ fluid stack,' from P and S inversions," in *Proc. SEG Tech. Program Expanded Abstr.*, 1997, pp. 183–186.
[6] X.-G. Chi and D.-H. Han, "Lithology and fluid differentiation using a rock physics template," *Lead. Edge*, vol. 28, no. 1, pp. 60–65, Jan. 2009.
[7] J. Nieto, B. Batlai, and F. Delbecq, "Seismic lithology prediction a Montney shale gas case study," *CSEG Rec.*, vol. 38, no. 2, pp. 34–43, 2013.
[8] J. Choi, S. Kim, B. Kim, and J. Byun, "Probabilistic reservoir characterisation using 3D PDF of stochastic forward modelling results in Vincent oil field," *Explor. Geophys.*, vol. 51, no. 3, pp. 341–354, May 2020.
[9] L. Jin, "Machine learning approaches for seismic-facies prediction and reservoir-property inversion," in *Proc. SEG Tech. Program Expanded Abstr.*, 2018, pp. 2147–2151.
[10] S. Ashgar, J. Choi, D. Yoon, and J. Byun, "Facies classification using semi-supervised deep learning with pseudo-labeling strategy," in *Proc. SEG Tech. Program Expanded Abstr.*, 2019, pp. 3171–3175.
[11] X. Liu, X. Chen, J. Li, X. Zhou, and Y. Chen, "Facies identification based on multikernel relevance vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7269–7282, Oct. 2020.
[12] Y. Zhang, Y. Liu, H. Zhang, and H. Xue, "Seismic facies analysis based on deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1119–1123, Jul. 2020.
[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
[14] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput.*, Aug. 2005, pp. 878–887.
[15] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[16] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 1–9.
[17] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.
[18] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. ICCV*, 2017, pp. 2223–2232.
[19] D. Kim and J. Byun, "Data augmentation using CycleGAN for overcoming the imbalance problem in petrophysical facies classification," in *Proc. SEG Tech. Program Expanded Abstr.*, 2020, pp. 2310–2313.
[20] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
[21] A. Amin, F. Rahim, I. Ali, C. Khan, and S. Anwar, "A comparison of two oversampling techniques (SMOTE vs MTDF) for handling class imbalance problem: A case study of customer churn prediction," in *New Contributions in Information Systems and Technologies*. Cham, Switzerland: Springer, 2015, pp. 215–225.
[22] A. More, "Survey of resampling techniques for improving classification performance in unbalanced datasets," 2016, *arXiv:1608.06048*. [Online]. Available: http://arxiv.org/abs/1608.06048
[23] Y. Suh, C. Kim, L. Song, J. Yu, and J. Mo, "A comparison of oversampling methods on imbalanced topic classification of Korean news articles," *J. Cognit. Sci.*, vol. 18, no. 4, pp. 391–437, Dec. 2017.
[24] D. P. Hampson, B. H. Russell, and B. Bankhead, "Simultaneous inversion of pre-stack seismic data," in *Proc. SEG Tech. Program Expanded Abstr.*, 2005, pp. 1633–1637.