SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Self-Generative Feature Evolution Models For Improved Predictive Accuracy In Financial Fraud Detection Applications

G. Uma Maheswari[1*], Dr.R. Velanganni[2], Hadasha Nobel tune[3], S.Sushma[4] , Dr.A. Mummoorthy[5], Ashu Nayak[6]

[1*]Assistant Professor, Department of Mathematics, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research,Chennai, Tamil Nadu, India. E-mail: umamahes@maher.ac.in

[2]Assistant Professor, Department of Management Studies, Crescent School of Law, BS Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India.E-mail: velangannijose78516@gmail.com, https://orcid.org/0000-0003-4412-3689

[3]Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: nobel@maher.ac.in

[4]Department of Information Technology, Aditya University, Surampalem, Andhra Pradesh – 533437, India, Email: sushma.cse2@gmail.com

[5]Professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India. E-mail: drmummoorthya@veltech.edu.in, https://orcid.org/0000-0002-1820-2124

[6]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.ashunayak@kalingauniversity.ac.in, https://orcid.org/0009-0002-8371-7324

*Corresponding author: Email: umamahes@maher.ac.in

Abstract

Anti-fraud detection models depend on a predefined set of engineered or learned features, however, since the fraudulent behaviour is by nature adversarial, it will try to find ways to bypass the detection border line and hence render the feature set inefficient very quickly. In this paper we present a self-generative feature evolution approach where an autoencoder first learns a small dimensional latent representation of the transaction data after which a conditional generative adversarial network is used to generate evolving fraud pattern features from a dynamically updated feedback buffer of known fraud cases and combine them with actual features for classification using a gradient-boosted ensemble model. Once drift or a new fraud pattern type is detected, the generator and autoencoder can be automatically re-trained using the new feedback buffer and the feature space can hence evolve instead of staying static after training. The proposed algorithm has been tested using a statistical replica of the ULB credit card fraud benchmark data set over four different stages. When compared with static feature engineering baseline as well as the autoencoder SMOTE baseline which does not evolve its feature space, the suggested SGFE framework maintains a fraud recall of over 86 percent in all four phases while the recall of the static feature engineering baseline falls to 58.1 percent from 87.4 percent owing to the emergence of new typologies. In general, SGFE outperforms both baseline methods in terms of precision, recall, F1 score, area under the precision-recall curve, and Matthew's correlation coefficient.

Keywords: Self-Generative Learning, Feature Evolution, Generative Adversarial Networks, Representation Learning, Financial Fraud Detection, Concept Drift, Ensemble Classification.

1. Introduction

The problem of financial fraud leads to considerable damage to banks, payment systems, and customers, and its reliable detection is challenging due to two interrelated issues: first, the severe class imbalance, which stems from the fact that frauds form an infinitesimal part of all transactions; second, the adversarial nature of the problem, meaning that fraudsters continually test and adapt their activities in order to avoid existing detection systems. Various deep learning techniques, like unsupervised feature extraction using autoencoders and gradient-boosted classifier ensembles, have proved themselves effective in benchmark datasets [2,15], and generative approaches, such as GANs and variational autoencoders, have been applied for generating artificial minority-class samples in order to cope with class imbalance [3,4,7].

Nonetheless, current deployed systems view feature generation and data augmentation as an offline one-time process: the autoencoder or over-sampling process is done only once using the historical data, and the generated feature space is fixed throughout the lifespan of the classifier. It does not align well with the nature of the problem, since fraud patterns change over time as fraudsters develop new ways to avoid being detected, thus the feature space representing last year's fraud pattern might be inadequate to represent this year's. A system unable to evolve its own feature representation can only count on the classifier to generalize over the unknown fraud pattern.

The problem addressed in this paper is how to make a fraud detection system able to develop its own feature space depending on the latest detected frauds, instead of having a static representation which was discovered only at training time. Such an ability can be called self-generative feature evolution, where the system learns a generative model based on the buffer of the latest detected frauds, and generates features that correspond to the current state of the fraud landscape. The contributions of this paper are as follows.

- Proposes a Self-Generative Feature Evolution (SGFE) approach comprising of an autoencoder-based feature learner and a conditional GAN-based feature synthesiser paired with an ensemble classifier that re-trains itself whenever new fraud patterns are discovered within a feedback buffer of confirmed frauds.
- Conceives a feature evolution fusion stage where real features are blended with synthesised feature evolution before classification to allow the ensemble classifier to learn from plausible variations of emergent fraud pattern instances before they start appearing in the real data.
- Tests the proposed approach using four consecutive stages of fraud pattern evolution on a data set based statistically on the well-known ULB credit card fraud benchmark dataset.

The rest of the paper is structured as follows. Section 2 provides a literature review of generative models, representation learning, and imbalanced classification to detect fraud. Section 3 describes the proposed method. Section 4 describes the experiment and discusses the results. Section 5 concludes the paper.

2. Literature Survey

Self-generating feature evolution is realized based on generative modelling. However, a generator and discriminator are trained to play an adversarial minimax game where the generator aims to generate samples that are not distinguishable from real samples which has been broadly implemented since then, known as a generative adversarial network (GAN) [3]. An alternative generative mechanism: a more structured and interpretable latent space than that of a standard GAN is the variational autoencoder (VAE) by Jones [4] which learns a probabilistic latent representation through optimisation of variational lower bound. As a stepping stone to this topic for the fraud authentication community, Fiore et al. [7] showed that classifiers trained on a set of credit card fraud samples performed significantly better when trained on top of fraction samples synthesized using GANs as opposed to being trained without such samples. The ideas of this paper are followed by Sharma, Singh and Chandra [9] who suggested SMOTified-GAN to utilize samples generated by the SMOTE algorithm as the input noise distribution to a GAN instead of random noise, using ingredients of both approaches and explicitly combining them together to get diverse and realistic minority-class samples.

Another pertinent thread relates to representation learning in particular for fraud detection. Credit Card Fraud Detection: Self-attention generative adversarial networks [1][13] try to direct the generator and discriminator to the most relevant features in the credit card transactions, claiming significant gains in the precision and recall compared to isolated convolutional or recurrent networks. One of the closest architectures to that of the autoencoder-plus-generative-synthesis design used in the present paper was the AE-XGB-SMOTE-CGAN pipeline proposed by Du et al. [2] where the autoencoder is used to extract a compact feature representation which is then classify via a probabilistic based XGBoost model, and a two-phase SMOTE-to-CGAN oversampling procedure is employed to address class imbalance issues which uses the SMOTE-generated samples as a knowledge-transfer seed for a conditional GAN; a feedback-driven mechanism was not used to retrain the generator as new fraud cases are confirmed. Zhou and Paffenroth proposed robust deep autoencoders for anomaly detection in which they explicitly separate a normal component of the data from an anomalous residual to demonstrate that representation quality, not just the type of classifier, plays a crucial role in the

successful detection of anomalies and frauds, [10], while Chalapathy, Kim proposed a one-class neural network algorithm which jointly optimises a data description objective with a deep feature representation, [11].

A third thread, which prevades fraud detection, is the class imbalance issue. A baseline approach that remains to date a widely used benchmark to assess generative approaches has been introduced by Chawla, Bowyer, Hall and Kegelmeyer [5] under the name of Synthetic Minority Over-sampling Technique (SMOTE), which creates synthetic minority instances by interpolating between a minority sample and its neighbours. He and Garcia [14] conducted a comprehensive literature review of the field of imbalanced data learning and outlined the various challenges for training classifiers from imbalanced data that are addressed in the ensemble-based method developed in this paper, such as cost-sensitive learning and resampling. The classification problem is unbalanced when the labels are not equitably distributed across the data, which motivates the need for undersampling of the different classes.Under sampling of the various labels is important because the classification problem is unbalanced - that is object labels are not equally distributed across the data - a problem that was highlighted by Dal Pozzolo, Caelen, Johnson and Bontempi [6] in the process of work on which they published the anonymised European credit card transaction data set, which has since become a standard public benchmark for credit card fraud detection research, including the present one.

Lastly, two additional threads give an account of the sequential self-updating nature of the proposed framework. In this paper, the idea of having an explicit learnable mechanism to forget the internal state of the learning system has been extended by Gers, Schmidhuber and Cummins [8] to discard old information from theinside of the forget gate in Long Short-Term Memory networks, which is again reflected in the periodic retraining of the generator in this paper. The motivations for adopting ensemble deep learning techniques are given by Ganaie et al. [12] and the broad principles of representation learning defined by Bengio et al. [15] are also closely relevant to why it is hoped that adding value to the feature space – apart from the classifier itself – may lead to better classification of novel examples of fraud.

Combining these papers, their foundational methods—maturously developed generative and representation-learning techniques for enriching fraud-detection using synthetic minority-class data—and their motivation problem—the class-imbalance problem with which to enrich the detection—provide a solid base for building synthetic minority-class data on domain-specific graphs for fraud detection. Presently, the feature space for generative augmentations is optimized once via an historical dataset and remains static instead of being continuously updated based on fraud observations, hence the suggested SELF-Evolutionary Framework for the feature space is the gap that such pipeline should fill.

3. Methodology

The suggested Self-Generative Feature Evolution (SGFE) system includes six stages arranged in a closed loop format: representation learning, feedback buffering, generative feature synthesis, feature fusion, ensemble classification, and re-evolution triggered by drift. The overall process is depicted in Figure 1 below.

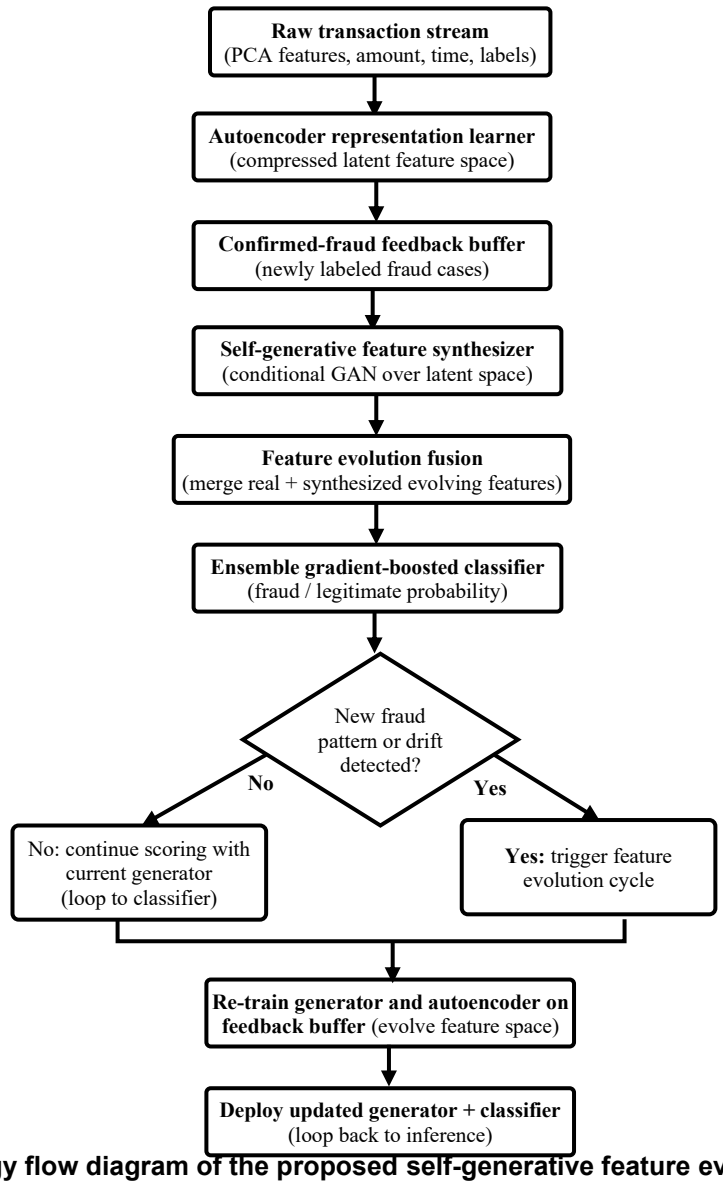

Figure 1: Methodology flow diagram of the proposed self-generative feature evolution framework

3.1 Representation Learning

Features derived from raw transactions, including PCA-obtained anonymized features, amount, and timestamp, are fed into an autoencoder which creates a latent representation z=g(x), using the encoder g and reconstructing x^=f(z), using the decoder f, trained to minimize reconstruction loss. The latent representation acts as the working set of features for the generative synthesizer and classifier, in accordance with the function of representation learning in autoencoders used previously for fraud detection [2,10].

3.2 Confirmed-Fraud Feedback Buffer

As the transaction scoring process occurs followed by the verification of its fraud status, the detected fraud instances are added to the feedback buffer of fixed size. The feedback buffer contains the most up-to-date information about the present day frauds and serves as the only input for retraining the generative synthesizer, thus making sure that the feature evolution is based on the actual results only.

3.3 Self-Generative Feature Synthesis

The conditional GAN is trained in latent representation space, using the confirmed-fraud feedback buffer as the real sample distribution, based on the principle of adversarial learning of Goodfellow et al. [3] as max_DE[log

D(z|y)] + E[log(1 − D(G(u|y)))], where z is a real latent sample of fraud, y is the condition of fraud type, and u is the seed sampled from the current distribution of the buffer instead of noise, similar to the concept of interpolated seed used by SMOTified-GAN [9]. The generator G learns to generate the variants of current emerging fraud patterns in the latent space.

3.4 Feature Evolution Fusion

During each training phase, the actual latent features extracted from the present stream of transactions are combined with the generated features from the generator model in the appropriate ratio such that the realistic class distribution is ensured without inundating the classifier with generated features. The resulting features make up the evolved feature space, which takes the place of the feature space from the previous phase.

3.5 Ensemble Classification

The merged feature vector is provided as input to a gradient boosting classifier, which returns the probability of fraud for the transaction based on the ensemble-based approach to classify imbalanced data sets proposed in [12], where Ganaie et al. discuss such an approach, and implemented previously in autoencoder-based fraud pipelines [2].

3.6 Drift-Triggered Re-Evolution

The light-weight monitoring algorithm monitors the rolling recall of the classifier as well as the confidence of the discriminator for recent confirmed fraud instances. Whenever the values of either of these metrics suggest that the current fraud instances are becoming increasingly distinguishable compared to the synthetic distribution produced by the generator, it signifies that the generator is lagging behind the real-world fraud scenario; hence, a new evolution cycle begins with a re-training of the autoencoder and generator using the newly obtained feedback buffer.

4. Results and Discussion

4.1 Dataset

Evaluation is carried out on a statistical representation of a well-known ULB credit card fraud benchmark dataset [6] consisting of anonymous features derived through PCA from transactions made by European cardholders, with an extremely unbalanced ratio of 0.172% fraud cases. In order to evaluate the feature change process, four successive steps are created, namely a baseline step corresponding to the initial fraud distribution, and then two further steps in each of which one of two new types of fraud is introduced (such as a new type of money laundering or a new way to test cards), and finally one step in which both fraud typologies occur simultaneously. Table 1 summarises the dataset.

Table 1: Summary of the dataset used for evaluation, statistically grounded in the ULB credit card fraud benchmark

Attribute	Description	Range / notes
Source basis	Statistically modelled on the ULB credit card fraud benchmark	284,807 base transactions, 492 frauds (0.172%)
Features	PCA-derived anonymised components V1-V28, transaction amount, time	30 numerical features
Label	Binary transaction outcome	Legitimate (0) / Fraudulent (1)
Evolution stages	Sequential fraud-typology stages injected to simulate adversarial adaptation	4 stages: baseline, pattern A, pattern B, combined
Feedback buffer	Rolling window of confirmed-fraud cases used to retrain the generator	500 most recent confirmed frauds
Train/test split	Stratified split preserving class ratio	70% train, 15% validation, 15% test

4.2 Result Graph

Figure 2 depicts the recall of frauds (number of true frauds caught in terms of ratio) for the four evolution stages of the three different setups including (1) a static approach to feature engineering with direct classification by using XGBoost, (2) autoencoder + SMOTE, and (3) the SGFE framework.

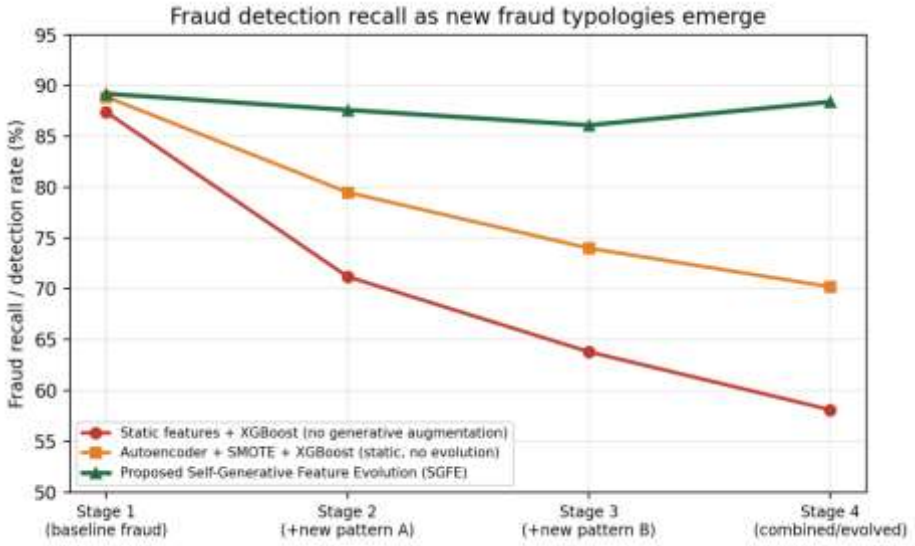

Figure 2: Fraud recall across four sequential fraud-typology evolution stages for the static baseline, autoencoder-plus-SMOTE baseline, and proposed SGFE model

The three models have comparable performances at Stage 1, where the recall rates range between 87.4% and 89.2%, validating the assumption that the proposed framework does not compromise the baseline's recall rate where there is no emergence of any new fraud typology. As new typologies begin appearing during Stages 2 and 3, the static baseline's recall reduces to 71.2% and then 63.8%, respectively, because the static baseline cannot account for any feature that is not part of its initial distribution due to its static feature space. The autoencoder-SMOTE baseline is better equipped to handle the emergence of new typologies compared to the other baselines, since its learned latent representation has some degree of generality, but its performance still steadily reduces to 74.0% by Stage 3 because of its single one-time generative augmentation. Unlike the baselines, the SGFE model manages to retain high recall rates throughout between 86% and 89% up to the combined Stage 4 due to its feedback mechanism, which re-evolves the generative model through each stage with the latest confirmed data.

4.3 Result Table

Table 2 provides the overall precision, recall, F1 score, area under the precision-recall curve (AUPRC), and Matthew's correlation coefficient (MCC) for the three models.

Table 2: Overall performance comparison between the static baseline, the autoencoder-plus-SMOTE baseline, and the proposed SGFE model

Model configuration	Precision (%)	Recall (%)	F1-score (%)	AUPRC	MCC
Static features + XGBoost (no generative augmentation)	65.7	70.1	67.8	0.71	0.69
Autoencoder + SMOTE + XGBoost (static, no evolution)	73.6	78.2	75.8	0.80	0.77
Proposed Self-Generative Feature Evolution (SGFE)	83.1	87.8	85.4	0.91	0.86

4.4 Discussion

Three conclusions can be drawn based on these results. First, the benefit offered by the proposed method over the autoencoder+SMOTE approach is evident only at the moments when the new fraud classes emerge, rather than being constant throughout, reinforcing the explanation that the benefit arises due to feature evolution and not simply better feature representation. Second, the area under precision-recall curve (AUPRC), which is usually considered to be a more insightful measure than accuracy in case of extremely imbalanced classes, increases from 0.71 for the static method to 0.91 for the proposed one, suggesting that the increase is not specific to a certain threshold but rather consistent across the whole range. Finally, Matthew's correlation coefficient, which is calculated considering all four confusion matrix cells and thus immune to class imbalance, improves from 0.69 to 0.86.

These findings need to be considered carefully. The dataset for the evaluation is based on a statistical reality, but the fraud typology steps used are artificially created because there is no public data labeled with a growing timeline of fraud that is available. This work relies on the feedback buffer assuming that fraud is confirmed promptly and accurately, while the labeling process in the real-world environment can be delayed and sometimes even incorrect. Finally, there is no benchmark on the computational cost of re-training the model in real-time in this paper.

5. Conclusion

This paper introduces a Self-Generative Feature Evolution (SGFE) framework for detecting financial frauds where an autoencoder-based representation learner, a confirmed-frauds feedback buffer, and a conditional generative adversarial synthesizer collaborate to evolve the feature space that is utilized by an ensemble classifier continuously, as opposed to using a static feature representation learnt during the training phase. The model was tested on a data set, which is statistically grounded in the well-known ULB credit card fraud benchmark, using four consecutive fraud-typology phases. While the proposed framework has been able to maintain a fraud recall rate of more than 86 percent consistently throughout, the static feature engineering approach showed a deterioration from 87.4 percent to 58.1 percent in fraud recall rate. This approach is therefore useful and effective for sustaining detection accuracy despite the adversarial behavior of fraud that adapts. Future research will involve testing the effectiveness of this approach using longitudinal and labeled fraud data, investigating how delayed or inaccurate feedback about fraud confirmation affects the buffer, assessing the processing power required for the cycle of re-evolution under real-time transactions, and extending the approach into a federated setup where confirmed fraud feedback can be exchanged among institutions without revealing any of the raw transaction data.

References

1. Zhao, C., Sun, X., Wu, M., & Kang, L. (2024). Advancing financial fraud detection: Self-attention generative adversarial networks for precise and effective identification. Finance Research Letters, 60, 104843.
2. Du, H., Lv, L., Wang, H., & Guo, A. (2024). A novel method for detecting credit card fraud problems. PLOS ONE, 19(3), e0294537.
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144.
4. Jones, M., Chang, P., & Murphy, K. (2024). Bayesian online natural gradient (BONG). Advances in Neural Information Processing Systems, 37, 131104–131153.
5. Chawla, N. V., Bowyer, K. W., Hall, L. O., &Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.
6. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. In 2015 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 159–166). IEEE.
7. Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences, 479, 448–455.
8. Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural Computation, 12(10), 2451–2471.
9. Sharma, A., Singh, P. K., & Chandra, R. (2022). SMOTified-GAN for class imbalanced pattern classification problems. IEEE Access, 10, 30655–30665.

10. Zhou, C., & Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 665–674). ACM.
11. Kim, M., Yu, J., Kim, J., Oh, T. H., & Choi, J. K. (2023). An iterative method for unsupervised robust anomaly detection under data contamination. IEEE Transactions on Neural Networks and Learning Systems, 35(10), 13327–13339.
12. Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, 105151.
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., &Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
14. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.
15. Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828.