SCFD: A Self-Calibrated Feature Denoising Framework for Robust Medical Image Classification Under Extreme Class Imbalance

Guanqi Cheng
School of Computer Science
University of Technology Sydney
Ultimo, NSW 2007, Australia
Email: Guanqi.Cheng@student.uts.edu.au

Mukesh Prasad
School of Computer Science
University of Technology Sydney
Ultimo, NSW 2007, Australia
Email: Mukesh.Prasad@uts.edu.au

Ali Braytee
School of Computer Science
University of Technology Sydney
Ultimo, NSW 2007, Australia
Email: Ali.Braytee@uts.edu.au

***Abstract*—Deep learning models for medical image classification often underperform when class imbalance is extreme, as supervised methods overfit the majority class, leading to biased feature representations. To address this, we propose Self-Calibrated Feature Denoising (SCFD), a framework that enhances the performance of self-supervised models in imbalanced settings. SCFD integrates a pretrained DINOv2 encoder for class-specific feature extraction, a lightweight autoencoder to denoise minority-class embeddings, and a latent-space regularization mechanism that leverages majority-class information for calibration. By combining these components, SCFD generates balanced, discriminative embeddings without modifying the backbone, enabling seamless integration into existing pipelines. Experiments on three disease diagnosis datasets, Lab-shanghai, Brain Tumor 4C, and Chest CT, demonstrate that SCFD consistently improves model robustness and generalization under extreme class imbalance, providing a reliable approach for effective medical image classification.**

***Index Terms*—Self-supervised learning, medical image, DINOv2 , class imbalance, image classification**

I. Introduction

Medical risks such as cerebral infarction, brain tumors, and lung cancer remain major causes of morbidity and mortality worldwide, severely threatening patient health and quality of life [1]–[3]. Accurate and timely diagnosis is essential for effective treatment; however, conventional diagnostic procedures are labor-intensive, time-consuming, and subject to inter-observer variability, motivating the development of automated imaging-based solutions [4]. Deep learning (DL) has significantly advanced medical image analysis by enabling lesion detection, segmentation, and classification, thereby improving diagnostic accuracy and workflow efficiency [5]. Despite this progress, several challenges hinder widespread clinical adoption. Annotated medical data remain scarce and expensive to obtain; labels are often noisy due to inter-observer disagreement; imaging protocols vary across scanners and institutions [6]; and current models still suffer from limited interpretability and robustness, reducing clinical trust [7], [8].

Among these issues, class imbalance is particularly critical. In real-world datasets, rare pathological findings are much less frequent than normal or common cases, leading to biased optimization and poor sensitivity for minority categories [9]. Traditional remedies, such as re-sampling and class weighting, offer limited benefits and may destabilize training in highly imbalanced scenarios [10]. Supervised learning (SL) with models such as VGG, ResNet, and ViT has been widely explored in medical imaging [11], [12]. However, their performance deteriorates significantly under severe imbalance. To reduce dependence on annotated data, self-supervised learning (SSL) has emerged as a promising paradigm [13], [14]. By leveraging large-scale unlabeled data for pretraining and fine-tuning with minimal supervision, SSL provides transferable visual representations that often rival or surpass SL in medical applications [15], [16]. Nevertheless, SSL alone does not fully overcome imbalance: minority-class features remain noisy, weakly structured, and prone to collapse during training [17], [18].

To address this, we propose a Self-Calibrated Feature Denoising (SCFD) framework. SCFD uses a self-supervised encoder (DINOv2) to extract unbiased features and a lightweight autoencoder trained on minority-class samples to suppress intra-class noise. The autoencoder latent space is regularized with a majority-class prior via KL divergence, producing denoised features that are passed to a lightweight classifier. SCFD requires no modifications to the pretrained backbone, enabling easy integration. Evaluation across three disease diagnosis datasets demonstrates its superiority over both supervised and self-supervised models, particularly under severe class imbalance.

The main contributions of this work are:

- A novel SCFD framework that enhances minority-class feature representations via denoising and latent-space regularization, without requiring backbone retraining.
- Extensive evaluation of both supervised and self-supervised methods under varying class imbalance ratios, across multiple backbones including ResNet, VGG, DenseNet, ShuffleNet, ViT, and DINOv2.

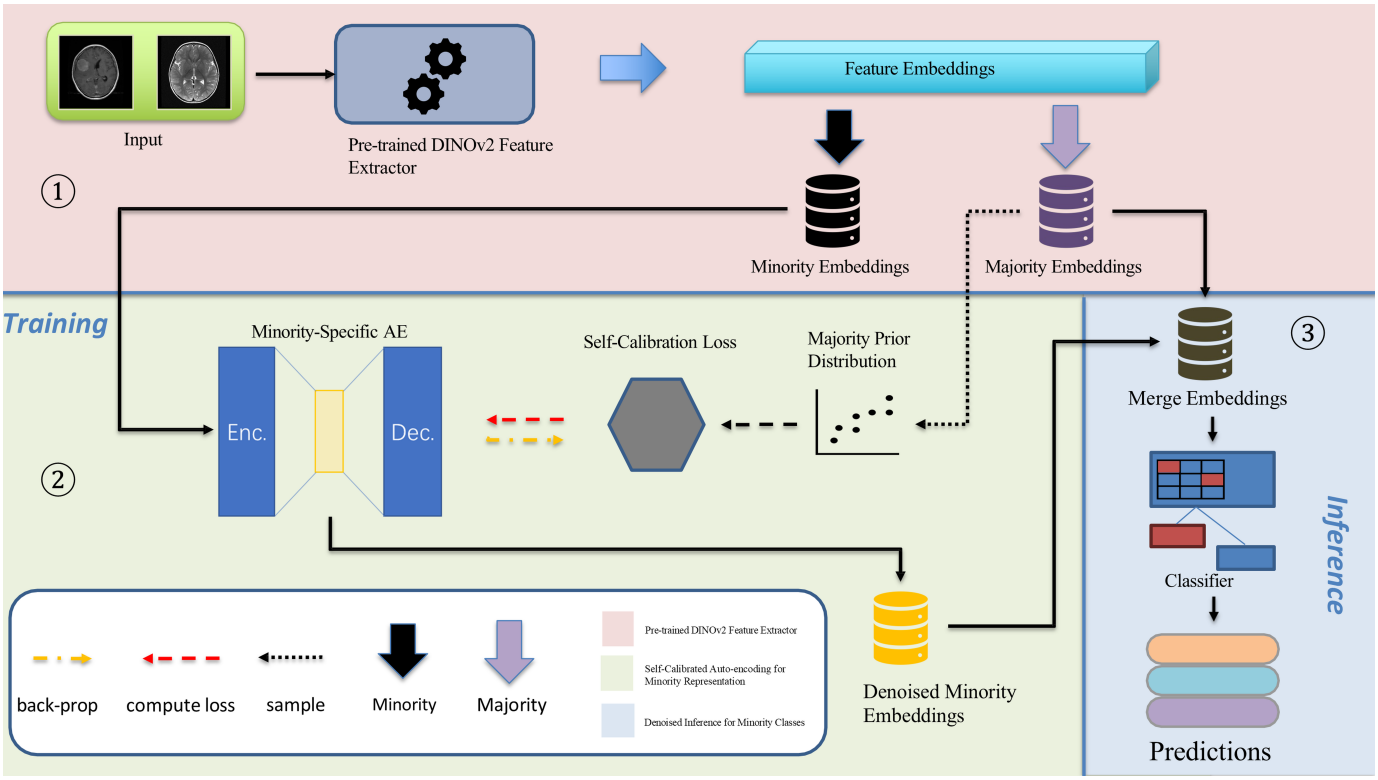


Fig. 1. The proposed Self-Calibrated Feature Denoising (SCFD) framework.

II. Related Work

This section reviews representative supervised and recent self-supervised learning approaches for addressing class imbalance in medical imaging.

A. Class Imbalance Approaches with Supervised Learning

Rajendran et al. [19] developed a breast cancer prediction pipeline that incorporates data selection, pre-processing, class-balancing techniques, and multiple supervised classification methods, such as Bayesian Networks, Random Forests, and Decision Trees. While the approach improved overall predictive performance, it still exhibited limitations in accurately identifying the minority class (i.e., cancer cases) under severe class imbalance, particularly lacking a precise mechanism for detecting clinically high-risk individuals. Liu et al. [20] proposed an ensemble learning framework that integrates SMOTE over-sampling with CVCF noise filtering and employs diverse SVM classifiers with multiple kernel functions, optimizing classifier weights via the SAGA algorithm to effectively improve classification performance on imbalanced data; however, existing methods still struggle to fully eliminate the impact of noisy samples and require enhanced generalization and robustness when handling highly imbalanced datasets. Hu et al. [21] proposed a deep supervised learning framework with a self-adaptive auxiliary loss (DSN-SAAL) that re-weights cross-entropy using the effective number of samples and incorporates reverse cross-entropy regularization to improve COVID-19 diagnosis from imbalanced and noisy CT images; however, existing methods struggle to simultaneously address severe class imbalance and label noise, leading to biased predictions and poor generalization on minority classes.

B. Class Imbalance Approaches with Self-supervised Learning

Elbatel et al. [22] proposed Fed-MAS, a federated learning framework that leverages self-supervised priors to compute class-aware divergence and rescue factors for weighted model aggregation, effectively addressing the challenge of learning from minority classes in highly imbalanced medical image datasets; however, existing federated approaches lack mechanisms to capture class-specific discrepancies and often fail to generalize under extreme class imbalance and non-IID data distributions. Su et al. [23] proposed STGAN, a two-stage GAN-based data augmentation framework that leverages self-supervised learning and Freeze-D to transfer universal and class-specific knowledge for generating high-quality skin lesion images, thereby improving classification performance on imbalanced multi-class datasets; however, existing methods struggle with limited labeled data and severe class imbalance, resulting in poor accuracy for minority classes and limited clinical applicability. Li et al. [24] proposed a self-supervised dual-track learning-to-rank framework that selects diverse and difficult negative CT samples to efficiently train COVID-19 classification networks in the presence of data imbalance, addressing the limitations of existing methods that are either time-consuming or lack effective negative-sample selection strategies.

III. Proposed Method

This section presents the Self-Calibrated Feature Denoising (SCFD) framework, designed to enhance minority-class representations in class-imbalanced medical image classification. SCFD leverages a minority-specific autoencoder, regularized by a majority-class prior, to denoise features and improve robustness under severe imbalance. We first briefly review DINOv2, the self-supervised vision transformer serving as the backbone.

A. A Brief on DINOv2

DINOv2 is a self-supervised model that learns high-level semantic features from large-scale unlabeled data [25], [26]. It combines image-level and block-level contrastive objectives with techniques like masked image modeling, high-resolution fine-tuning, and KoLeo regularization, enabling robust feature extraction. This makes it well-suited for class-imbalanced medical imaging with limited minority-class samples.

B. Self-Calibrated Feature Denoising (SCFD) Method

The SCFD framework (Fig. 1) consists of three components: (1) Pretrained DINOv2 feature extractor, (2) Self-Calibrated Autoencoding for minority features, and (3) Denoised inference for minority classes. The first component extracts feature embeddings from all samples and partitions them based on the class distribution. In the second component, a lightweight autoencoder is trained solely on minority-class embeddings to suppress noise. The reconstruction objective encourages the model to recover reliable features. To further structure the latent space, a self-calibration loss aligns minority latent embeddings with a reference distribution derived from the majority-class latent space using KL divergence. The third component applies the trained autoencoder to denoise minority embeddings. These refined embeddings are then merged with majority-class embeddings to form a unified feature set for training downstream classifiers. This procedure enhances classification performance and robustness, particularly for underrepresented classes.

1) Pretrained DINOv2 Feature Extractor: Feature embeddings are extracted via the fixed pretrained DINOv2 encoder $\phi(\cdot)$:

$$f_i = \phi(x_i), \quad f_i \in \mathbb{R}^d. \tag{1}$$

Embeddings are divided into majority-class $\{f_i^{(maj)}\}$ and minority-class $\{f_i^{(min)}\}$ sets, with minority embeddings generally exhibiting higher noise due to fewer samples.

2) Self-Calibrated Autoencoder for Minority Features: The autoencoder consists of encoder $E : \mathbb{R}^d \to \mathbb{R}^{d_z}$ and decoder $D : \mathbb{R}^{d_z} \to \mathbb{R}^d$:

$$z_i = E(f_i), \quad \hat{f}_i = D(z_i). \tag{2}$$

Reconstruction loss is defined as

$$\mathcal{L}_{\text{recon}} = \frac{1}{N} \sum_{i=1}^{N} \|f_i - \hat{f}_i\|_2^2. \tag{3}$$

Minority and majority latent distributions $q(z)$ and $p(z)$ are approximated as diagonal Gaussians:

$$\mu_p^{(k)} = \frac{1}{N_{\text{maj}}} \sum_{j=1}^{N_{\text{maj}}} z_j^{(\text{maj},k)}, \quad \sigma_p^{2(k)} = \frac{1}{N_{\text{maj}}} \sum_{j=1}^{N_{\text{maj}}} \left(z_j^{(\text{maj},k)} - \mu_p^{(k)}\right)^2 + \epsilon,$$
$$\mu_q^{(k)} = \frac{1}{N_{\text{min}}} \sum_{i=1}^{N_{\text{min}}} z_i^{(\text{min},k)}, \quad \sigma_q^{2(k)} = \frac{1}{N_{\text{min}}} \sum_{i=1}^{N_{\text{min}}} \left(z_i^{(\text{min},k)} - \mu_q^{(k)}\right)^2 + \epsilon. \tag{4}$$

KL divergence is used as the self-calibration loss:

$$\mathcal{L}_{\text{SCFD}} = \frac{1}{2} \sum_{k=1}^{d_z} \left[\log \frac{\sigma_p^{2(k)}}{\sigma_q^{2(k)}} - 1 + \frac{\sigma_q^{2(k)}}{\sigma_p^{2(k)}} + \frac{(\mu_p^{(k)} - \mu_q^{(k)})^2}{\sigma_p^{2(k)}} \right]. \tag{5}$$

The total loss is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda \mathcal{L}_{\text{SCFD}}, \quad \lambda = \frac{1}{B} \sum_{i=1}^{B} \text{std}(z_i). \tag{6}$$

3) Denoised Inference for Minority Classes.: After training, the autoencoder refines minority embeddings:

$$\tilde{f}_i = D(E(f_i^{(min)})). \tag{7}$$

The refined feature set is

$$\mathcal{F}_{ref} = \mathcal{F}_{maj} \cup \{\tilde{f}_i\}. \tag{8}$$

Downstream classifiers are trained on $\mathcal{F}_{ref}$, thereby improving the robustness of the minority class without requiring retraining of the large, pre-trained backbone.

IV. Experiment

A. Datasets and Experimental Settings

Our method is evaluated on three datasets: Lab-shanghai (MRI, 23,148 cerebral infarction images), Brain Tumor 4C [27] (MRI, 7,023 images, 4 classes), and Chest CT [28] (CT, 1,000 images, 4 classes). The Lab-shanghai dataset was labeled by two experienced radiologists, with informed consent obtained from all patients and ethical approval for secondary use granted by the Human Research Ethics Committee of the University of Technology Sydney (UTS, ETH24-10033). Minority-class imbalance was simulated by reducing the lesion samples to 100%, 50%, 20%, and 5% of their original size, while keeping the majority-class samples unchanged. Performance was measured using macro-averaged Recall, Precision, and F1 score.

B. Implementation Details

All experiments were conducted on a Windows 11 workstation with an NVIDIA RTX 4090 GPU, Intel Core i9-13900KF CPU, and 64 GB DDR5 RAM (5200 MHz). The self-supervised model (DINOv2) was fine-tuned for 50 epochs with Adam (learning rate 1×10^{-6}), batch size 16, and Cross-Entropy Loss. Our SCFD was trained for 50 epochs with Adam (initial learning rate 1×10^{-3}, decayed by 1% per epoch), batch size 16, using a combined objective of mean squared error reconstruction loss and self-calibration loss. Supervised models were trained for 50 epochs with a batch size of 16, using Cross-Entropy Loss and SGD (momentum 0.9, initial learning rate 0.001), with a cosine annealing scheduler.

TABLE I
CLASSIFICATION PERFORMANCE OF SELECTED MODELS ACROSS DATASETS (MACRO-AVERAGED PRECISION (PREC.), RECALL (REC.), AND F1-SCORE (F1)) AT DIFFERENT IMBALANCE RATIOS

Dataset	Model	100%			50%			20%			5%		
		Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
Lab-shanghai	VGG19	0.994	0.984	0.989	0.956	0.956	0.956	0.952	0.816	0.870	0.993	0.714	0.796
	ResNet50	0.963	0.985	0.973	0.915	0.920	0.917	0.982	0.773	0.844	0.488	0.500	0.494
	DenseNet169	0.970	0.988	0.979	0.947	0.929	0.938	0.939	0.748	0.813	0.993	0.714	0.796
	ShuffleNetV2×1.0	0.977	0.991	0.984	0.950	0.936	0.942	0.948	0.794	0.852	0.895	0.784	0.830
	ViT-B/16	0.985	0.986	0.985	0.747	0.747	0.747	0.637	0.614	0.624	0.488	0.500	0.494
	DINOv2	**0.993**	0.992	0.992	**0.992**	0.990	0.991	**0.994**	0.973	0.983	**0.992**	0.959	0.975
	SCFD (Ours)	0.991	**0.995**	**0.993**	**0.992**	**0.992**	**0.992**	0.985	**0.986**	**0.986**	0.983	**0.976**	**0.979**
Brain Tumor 4C	VGG19	0.960	0.959	0.960	0.907	0.901	0.902	0.800	0.796	0.794	0.689	0.681	0.672
	ResNet50	0.940	0.939	0.940	0.844	0.842	0.841	0.782	0.774	0.773	0.676	0.674	0.661
	DenseNet169	0.953	0.952	0.952	0.904	0.900	0.901	0.828	0.819	0.820	0.711	0.706	0.695
	ShuffleNetV2×1.0	0.924	0.922	0.922	0.844	0.835	0.837	0.688	0.656	0.660	0.409	0.520	0.451
	ViT-B/16	0.859	0.859	0.859	0.810	0.808	0.808	0.735	0.723	0.723	0.635	0.576	0.562
	DINOv2	**0.996**	**0.996**	**0.996**	0.978	0.976	0.977	0.961	0.957	0.958	0.857	0.837	0.842
	SCFD (Ours)	0.993	0.993	0.993	**0.990**	**0.989**	**0.989**	**0.965**	**0.963**	**0.963**	**0.893**	**0.887**	**0.888**
Chest CT	VGG19	0.627	0.626	0.534	0.499	0.540	0.478	0.144	0.306	0.161	0.117	0.263	0.099
	ResNet50	0.639	0.585	0.526	0.518	0.552	0.477	0.455	0.501	0.429	0.322	0.488	0.381
	DenseNet169	0.542	0.572	0.524	0.496	0.523	0.487	0.524	0.507	0.481	0.389	0.488	0.409
	ShuffleNetV2×1.0	0.095	0.250	0.138	0.043	0.250	0.073	0.043	0.250	0.073	0.043	0.250	0.073
	ViT-B/16	0.490	0.519	0.473	0.476	0.501	0.463	0.456	0.497	0.445	0.443	0.493	0.437
	DINOv2	**0.915**	**0.933**	**0.919**	0.830	0.832	0.830	**0.748**	0.730	0.731	0.665	0.627	0.583
	SCFD (Ours)	0.887	0.867	0.874	**0.853**	**0.853**	**0.851**	0.744	**0.786**	**0.745**	**0.685**	**0.728**	**0.669**

V. RESULTS

Table I reports classification performance on Lab-Shanghai, Brain Tumor 4C, and Chest CT datasets under varying lesion class ratios (100%, 50%, 20%, 5%). Conventional supervised models (VGG19, ResNet50, DenseNet169, ShuffleNetV2, ViT-B/16) exhibit substantial performance degradation as the number of lesion samples decreases. For example, ResNet50 F1-scores drop from 0.973 to 0.494 on Lab-Shanghai, 0.940 to 0.661 on Brain Tumor 4C, and 0.526 to 0.381 on Chest CT at 5% lesion data. Lightweight networks such as ShuffleNetV2 and ViT-B/16 often fail under extreme imbalance, especially on Chest CT (F1 $\sim$0.073–0.437).

DINOv2 maintains strong and relatively stable performance across datasets, achieving F1-scores of 0.975, 0.842, and 0.583 at 5% lesion data for Lab-Shanghai, Brain Tumor 4C, and Chest CT, respectively. The proposed SCFD framework further enhances minority class representation and robustness, achieving F1 scores of 0.979, 0.888, and 0.669 at 5%, consistently outperforms DINOv2 in Recall and F1 while maintaining comparable Precision. These results demonstrate SCFD's effectiveness in highly imbalanced medical image classification scenarios.

A. Ablation Experiment

To evaluate the effectiveness of the proposed SCFD framework, we focus on the feature extractor. An ablation experiment is conducted to analyze its impact on the overall performance.

1) Impact of Feature Extractor : As shown in Fig. 2, we conducted an ablation study to evaluate the impact of different feature extractors on SCFD across class imbalance ratios of 100%, 70%, 50%, 20%, and 5%. Conventional CNNs (VGG19, ResNet50, DenseNet169, ShuffleNetV2 ×1.0) and a transformer-based model (ViT-B/16) were compared with the self-supervised DINOv2 backbone. The results show that SCFD benefits from high-quality feature representations. DINOv2 consistently achieves the highest Macro-F1 across datasets and imbalance levels, maintaining robust performance even under extreme scarcity of the minority class. In contrast, conventional CNNs and lightweight networks exhibit notable performance drops as the minority ratio decreases, indicating that effective feature embeddings are essential for SCFD to handle severe class imbalance. When the feature extractor produces weak or noisy embeddings, SCFD's self-calibration and denoising mechanisms are less effective, which can lead to lower Macro-F1 scores. These findings highlight that the choice of feature extractor is a key factor determining the framework's overall performance.

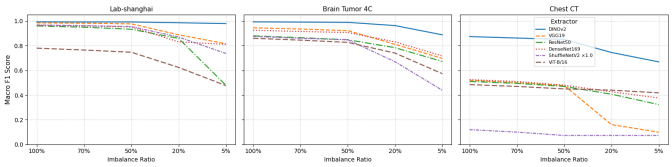


Fig. 2. Macro-F1 across three datasets with different feature extractors under varying imbalance ratios.

2) Impact of Self-calibration Loss Term : As shown in Fig. 3, we evaluated the impact of including the KL-based self-calibration loss under different class imbalance levels

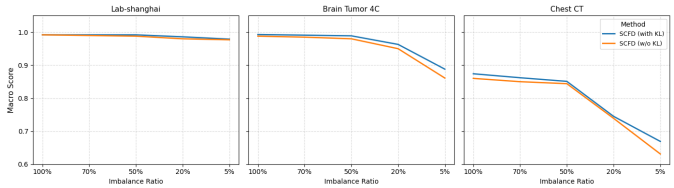
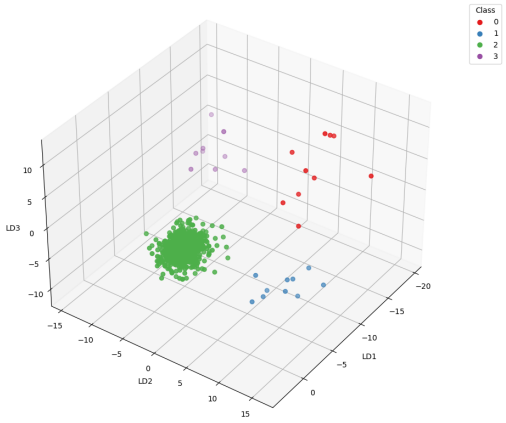
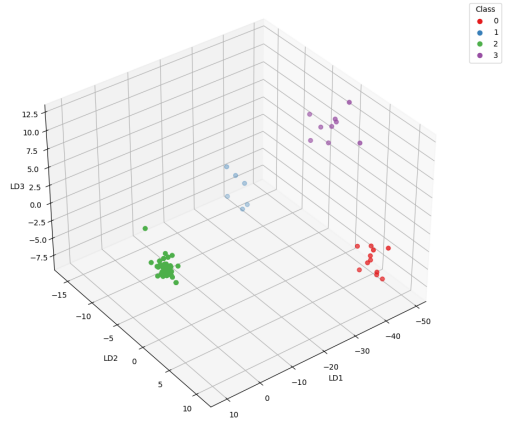


Fig. 3. Impact of Self-calibration Loss Term on SCFD Performance Measured by Macro-F1.

across three datasets. When the number of minority samples is sufficient, incorporating the KL term yields only marginal improvements in classification performance. However, under conditions of extreme scarcity, the KL loss plays a critical role in improving the model's robustness and accuracy.



(a) 3D visualization of Brain Tumor 4C dataset



(b) 3D visualization of Chest CT dataset

Fig. 4. 3D LDA-Based Visualization of Feature Embeddings under 5% Lesion Sample Proportion and Constant Normal Samples

B. Qualitative Visualization Results

To evaluate the model's ability to learn discriminative representations under severe class imbalance, only 5% of lesion (minority) samples were retained while the number of normal (majority) samples remained unchanged. Features were grouped according to ground-truth labels. Dimensionality reduction was performed in two steps: Principal Component Analysis (PCA) retained at least 80% of variance, followed by Linear Discriminant Analysis (LDA) to project data into a 3D space that maximizes class separability. Fig. 4 presents the resulting embeddings, with classes indicated by color, demonstrating SCFD's effectiveness in preserving minority class separability.

For the Brain Tumor 4C dataset (Fig. 4a), embeddings form distinct, compact clusters with minimal overlap, indicating strong inter-class separation. For the Chest CT dataset (Fig. 4b), despite the limited number of samples and severe class imbalance, the projected features still form well-separated clusters. These visualizations confirm SCFD's robustness in maintaining class separability and generating coherent embeddings under extreme imbalance, which is crucial for improving medical image classification.

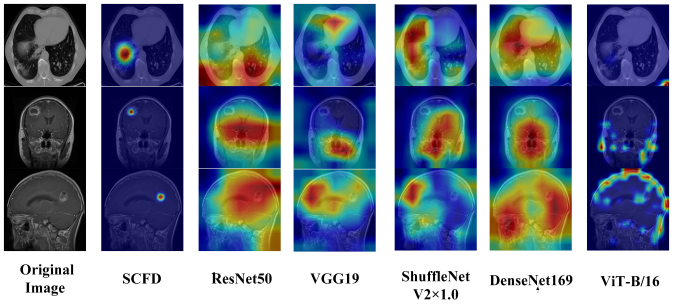


Fig. 5. Qualitative comparison of Grad-CAM results trained with 5% lesion annotations. From left to right: Input image, DINOv2-based classifier trained with SCFD, ResNet50, VGG19, ShuffleNetV2×1.0, DenseNet169, and ViT-B/16.

C. Grad-CAM Results

To evaluate the effectiveness of our method under an extremely low-data regime, Fig. 5 presents Grad-CAM visualizations on three representative samples from public datasets, where only 5% of the training data are used. The figure compares the original image with the Grad-CAM results produced by our SCFD method and several representative baselines, including ResNet50, VGG19, ShuffleNetV2×1.0, DenseNet169, and ViT-B/16.

As observed, the proposed method yields more localized and semantically meaningful activation regions that are well aligned with lesion-related structures, despite the severe data scarcity. In contrast, models trained without feature-level denoising tend to produce diffuse or fragmented activations, frequently responding to irrelevant background regions. This phenomenon suggests that conventional classifiers are more prone to overfitting noisy or spurious features when only a small fraction of training data is available.

The improved localization can be attributed to the self-calibrated feature denoising mechanism, which suppresses unreliable minority-class features while preserving discriminative semantic representations learned by the pretrained DINOv2 backbone. Consequently, the classifier trained with SCFD

attends to more consistent and informative regions during inference. These qualitative results provide visual evidence that SCFD enhances representation robustness under extreme class imbalance and limited supervision, complementing the quantitative improvements reported earlier.

VI. Conclusion

This study systematically evaluated DINOv2-based frameworks for disease detection across diverse medical imaging datasets and class imbalance scenarios. The results demonstrate that DINOv2, as a self-supervised vision encoder, offers strong generalization and robustness, outperforming several supervised models under balanced and imbalanced conditions. We proposed the Self-Calibrated Feature Denoising (SCFD) framework to improve performance in the presence of class imbalance by selectively enhancing minority-class representations without modifying the pretrained backbone. The integration of SCFD yields notable gains in classification accuracy and feature separability. These findings highlight the potential of combining self-supervised pretraining with targeted feature enhancement to advance diagnostic performance and facilitate the deployment of robust models in clinical environments.

References

[1] L.-S. Cheng, W.-J. Tu, Y. Shen, L.-J. Zhang, and K. Ji, "Combination of high-sensitivity c-reactive protein and homocysteine predicts the post-stroke depression in patients with ischemic stroke," *Molecular Neurobiology*, vol. 55, pp. 2952–2958, 2018.

[2] American Cancer Society, "Lung cancer. information and resources about cancer: Breast, colon, lung, prostate, skin," 2023, [Online]. Available: https://www.cancer.org/cancer/types/lung-cancer.html.

[3] Cancer.Net, "Brain tumor: Statistics," Mar. 2023, [Online]. Available: https://www.cancer.net/cancer-types/brain-tumor/statistics.

[4] A. T. Mohammadi, S. Bolhasani, F. Ghodrati, A. Mir, F. Sabouri, M. Amin, and E. Ghanbarzadeh, "Artificial intelligence in medical practice and research: Applications in nursing, hematology, radiology, and neuroscience," *Nobel Sciences*, 2025.

[5] X. Li, L. Zhang, J. Yang, and F. Teng, "Role of artificial intelligence in medical image analysis: A review of current trends and future directions," *Journal of Medical and Biological Engineering*, vol. 44, no. 2, pp. 231–243, 2024.

[6] A. Braytee, S. He, S. Tang, Y. Sun, X. Jiang, X. Yu, I. Khatri, K. Chaturvedi, M. Prasad, and A. Anaissi, "Identification of cancer risk groups through multi-omics integration using autoencoder and tensor analysis," *Scientific Reports*, vol. 14, no. 1, p. 11263, 2024.

[7] A. Anaya-Isaza, L. Mera-Jiménez, and M. Zequera-Diaz, "An overview of deep learning in medical imaging," *Informatics in Medicine Unlocked*, vol. 26, p. 100723, 2021.

[8] Y. Zhao, M. Prasad, and A. Braytee, "Dualprompt-medcap: a dual-prompt enhanced approach for medical image captioning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2025, pp. 205–215.

[9] M. Temraz and M. T. Keane, "Solving the class imbalance problem using a counterfactual method for data augmentation," *Machine Learning with Applications*, vol. 9, p. 100375, 2022.

[10] M. S. Shelke, P. R. Deshmukh, and V. K. Shandilya, "A review on imbalanced data handling using undersampling and oversampling technique," *Int. J. Recent Trends Eng. Res*, vol. 3, no. 4, pp. 444–449, 2017.

[11] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: a survey," *Evolutionary Intelligence*, vol. 15, no. 1, pp. 1–22, 2022.

[12] S. Takahashi, Y. Sakaguchi, N. Kouno, K. Takasawa, K. Ishizu, Y. Akagi, R. Aoyama, N. Teraya, A. Bolatkan, N. Shinkai, H. Machino, K. Kobayashi, K. Asada, M. Komatsu, S. Kaneko, M. Sugiyama, and R. Hamamoto, "Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review," *Journal of Medical Systems*, vol. 48, no. 1, p. 84, 2024.

[13] S. Shurrab and R. Duwairi, "Self-supervised learning methods and applications in medical imaging analysis: A survey," *PeerJ Computer Science*, vol. 8, p. e1045, 2022.

[14] H. Feng, Y. Jia, R. Xu, M. Prasad, A. Anaissi, and A. Braytee, "Integration of self-supervised byol in semi-supervised medical image recognition," in *International Conference on Computational Science*. Springer, 2024, pp. 163–170.

[15] P. Goyal, D. Mahajan, A. Gupta, and I. Misra, "Scaling and benchmarking self-supervised visual representation learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6391–6400.

[16] H. Liu, J. Z. HaoChen, A. Gaidon, and T. Ma, "Self-supervised learning is more robust to dataset imbalance," *arXiv preprint arXiv:2110.05025*, 2021.

[17] B. Kang, Y. Li, S. Xie, Z. Yuan, and J. Feng, "Exploring balanced feature spaces for representation learning," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020.

[18] Y. Yang and Z. Xu, "Rethinking the value of labels for improving class-imbalanced learning," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 19 290–19 301, 2020.

[19] K. Rajendran, M. Jayabalan, and V. Thiruchelvam, "Predicting breast cancer via supervised machine learning methods on class imbalanced data," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 8, pp. 57–61, 2020.

[20] N. Liu, X. Li, E. Qi, M. Xu, L. Li, and B. Gao, "A novel ensemble learning paradigm for medical diagnosis with imbalanced data," *IEEE Access*, vol. 8, pp. 171 263–171 280, 2020.

[21] K. Hu, Y. Huang, W. Huang, H. Tan, Z. Chen, Z. Zhong, X. Li, Y. Zhang, and X. Gao, "Deep supervised learning using self-adaptive auxiliary loss for covid-19 diagnosis from imbalanced ct images," *Neurocomputing*, vol. 458, pp. 232–245, 2021.

[22] M. Elbatel, H. Wang, R. Mart, H. Fu, and X. Li, "Federated model aggregation via self-supervised priors for highly imbalanced medical image classification," *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 334–346, October 2023.

[23] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, "A gan-based data augmentation method for imbalanced multi-class skin lesion classification," *IEEE Access*, vol. 12, pp. 16 498–16 513, 2024.

[24] Y. Li, D. Wei, J. Chen, S. Cao, H. Zhou, Y. Zhu, and Y. Zheng, "Efficient and effective training of covid-19 classification networks with self-supervised dual-track learning to rank," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2787–2797, 2020.

[25] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[26] Y. Huang, J. Zou, L. Meng, X. Yue, Q. Zhao, J. Li, C. Song, G. Jimenez, S. Li, and G. Fu, "Comparative analysis of imagenet pre-trained deep learning models and dinov2 in medical imaging classification," *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 297–305, Jul. 2024.

[27] M. Nickparvar, "Brain Tumor MRI Dataset," 2021. [Online]. Available: https://www.kaggle.com/dsv/2645886

[28] M. Hany, "Chest CT-scan images dataset," 2020. [Online]. Available: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images