Evaluation of the Privacy of Images Generated by ImageCLEFmedical GANs 2025 Based on Pre-trained Model Feature Extraction Methods

Notebook for the <Lab name> Lab at CLEF 2025

Dengtao Zhang[1,*,†], Xutao Yang[2,†]

[1]School of Information Science and Engineering, Yunnan University, Kunming 650504, Yunnan, China

Abstract

Our team's primary contributions to the ImageCLEFmedical GANs 2025 task are as follows. This task evaluates whether medical images generated by Generative Adversarial Networks (GANs) utilized specific real images while training the generative models. We developed a methodology that fine-tuned and evaluated multiple pre-trained models based on a contrastive learning framework, combined with a Mixture of Experts (MoE) strategy to fuse these models. Leveraging the similarity of feature extractions between generated and real images, we performed a binary classification task to identify real images that were potentially used during GAN training. Our best-performing model achieved a Cohen's Kappa score of 0.108 among the submitted results. Our experimental findings demonstrate that our approach can effectively distinguish between "used" and "unused" real images in the context of GAN training. Our code is public available at https://github.com/zhangdt123/image.

Keywords

GANs, pre-trained model, contrastive learning, MoE, Medical Imaging

1. Introduction

Deep learning has achieved remarkable progress in medical image analysis, demonstrating powerful capabilities in tasks such as classification, detection, and segmentation. However, the high performance of deep neural networks typically depends on large-scale, high-quality, and accurately annotated datasets. Due to the high cost of image acquisition, the need for expert annotation, and concerns about patient privacy, obtaining sufficient training data in medical imaging is often challenging. As a result, models are limited in terms of generalization and robustness [1].

In this context, deep generative models—such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Models—offer a promising solution. By learning the underlying distribution of real medical images, these models can synthesize structurally coherent and semantically consistent images, effectively mitigating the problem of data scarcity to a certain extent [2]. Furthermore, synthetic images can be leveraged for data augmentation, enhancing the stability of model training under limited data conditions and even providing additional "virtual samples" for clinically rare conditions, thereby broadening the applicability of medical AI models [3].

Despite the significant potential of deep generative models in medical imaging, their application is accompanied by a range of non-negligible concerns, particularly in the areas of privacy protection, ethical compliance, and clinical usability.

First, training generative models often requires access to large volumes of real patient imaging data. Without stringent data de-identification and access control measures, there is a risk of compromising patient privacy [4]. Moreover, although the generated images are synthetic, their underlying feature representations may still retain identifiable information from the original data, especially when employing high-fidelity models such as Generative Adversarial Networks (GANs). This risk of "re-identification"

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain

*Corresponding author.

1643734352@qq.com (D. Zhang); yangxutao@ynu.edu.cn (X. Yang)

0009-0004-7787-1629 (D. Zhang); 0000-0003-1983-0971 (X. Yang)

makes it difficult for synthetic data to fully comply with data protection regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) [5].

Second, generative models are susceptible to misuse. Utilizing synthetic medical images without thorough validation may cause models to rely on spurious features, ultimately compromising their diagnostic accuracy in real-world clinical settings [6]. For example, models may learn to recognize abnormal structures or lesion patterns that do not exist in authentic data, thereby undermining the reliability of clinical decisions. Additionally, if synthetic images are used in diagnostic tasks without expert annotation or review, this could lead to legal disputes concerning medical accountability and malpractice [7].

To advance research on the controllability, quality assessment, and clinical applicability of generative models in medical imaging, the ImageCLEF initiative has organized a series of medical challenge tasks.Our team's username is taozi. The 2025 ImageCLEFmedical competition includes a dedicated subtask focusing on generative models,[8] aiming to evaluate generative methods' feasibility and practical value in real-world medical scenarios. This study focuses on the competition's subtask titled "ImageCLEFmed GAN 2025: Training Data Analysis and Fingerprint Detection," [9] which centers on analyzing synthetic biomedical images to determine whether specific real images were used during the training of generative models. It's also named Subtask 1: "Detect Training Data Usage". Specifically, for each real image in the test set, the goal is to predict whether it was used in generating a given synthetic image (label 1) or not (label 0). The core challenge is to detect the presence of "fingerprints" of training data within the synthetic outputs.

Since synthetic images are generated by modeling the data distribution of real images, they often exhibit strong statistical similarity to authentic samples. The closer a generated image's distribution is to that of real images, the higher its perceived quality and visual realism. In this study, we formulate the task as a binary classification problem: determining whether a given real image was used during the generation process. To achieve this, we compute image similarity scores—higher similarity indicating likely usage and lower similarity suggesting non-usage.

Our approach primarily leverages a contrastive learning strategy combined with three pre-trained models and a Mixture-of-Experts (MoE) framework for training and evaluation. The pre-trained models include ResNet50, ViT-B/16, and EfficientNet-B0. First, we utilize these models to precompute similarity matrices between synthetic images and real training images, forming a candidate pool of positive samples for the contrastive learning framework. Next, the pre-trained models act as feature extractors, with a dynamic projection head applied for dimensionality reduction, enabling multi-level feature decoupling and adaptive parameter tuning. Finally, the trained deep learning models extract features from input images, and inter-feature similarities are computed. Within the contrastive learning framework, these features provide a more accurate representation of image similarity.

By integrating contrastive learning with diverse pre-trained models, we can comprehensively evaluate the similarity between generated and real images, thereby effectively solving the binary classification task. Based on the individual performance of the pre-trained models, we apply a Mixture-of-Experts (MoE) mechanism to fuse their outputs, offering richer feature representations and more robust guarantees for image analysis and interpretation.

2. Related Work

With the widespread application of deep learning in medical image analysis, models' reliance on large-scale, high-quality annotated datasets has become increasingly prominent. However, acquiring medical images is often constrained by high collection costs, stringent ethical approvals, specialized manual annotations, and concerns over patient privacy. These practical limitations hinder the generalization and robustness of AI models in tasks such as lesion detection, organ segmentation, and modality transformation. As a result, synthesizing medical images using deep generative models—as a means of data augmentation, supplementation, or even substitution for real samples—has emerged as a prominent

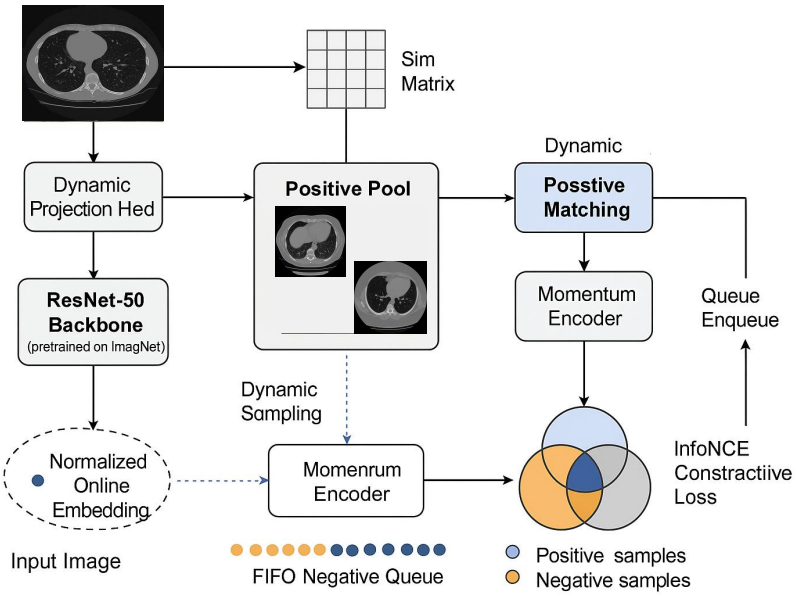

Figure 1: Overview of the Proposed Dynamic Contrastive Learning Framework

research focus in recent years [2, 3].

Typical applications of synthetic images in the medical domain include data augmentation, where additional training images are generated in few-shot scenarios to enhance model performance [10]; modality completion and transformation, such as generating one modality from another (e.g., CT to MRI), thereby facilitating multimodal learning and registration [11]; and privacy-preserving data sharing, where synthetic images are used to construct open-source medical datasets, alleviating restrictions on real data distribution.

Nevertheless, medical image synthesis faces multiple challenges. First, unlike natural images, medical images exhibit highly specialized anatomical and pathological structures. If these are not accurately expressed in the synthetic output, the result may appear visually plausible yet lack clinical relevance [6]. Second, there is currently no standardized, objective, and reproducible framework for evaluating the quality of generated medical images. This not only hampers the assessment of diagnostic utility but also impedes fair benchmarking across models [8]. Moreover, if model training does not adequately mitigate data leakage risks, there remains the possibility that real patient images are "implicitly memorized" and reproduced in the generated output [12].

In terms of technical approaches, the dominant generative models currently include:

Generative Adversarial Networks (GANs), such as pix2pix, CycleGAN, and StyleGAN, are widely used for generating and translating CT, MRI, and X-ray images due to their ability to produce highly detailed and visually realistic results [13, 14]. Variational Autoencoders (VAEs) are better suited for modeling the latent distribution of images, generating stable but less detailed outputs, and are often employed in scenarios requiring control over anatomical shape or organ structure [15]. Diffusion Models have recently gained traction in medical imaging for their iterative generation process, offering improved stability and higher-quality synthesis compared to GANs, particularly in high-resolution image generation tasks [16]. Conditional Generative Models (e.g., cGANs, VAE-GANs) incorporate structural information such as labels, semantic maps, or medical text, enabling the production of synthetic images with higher clinical fidelity.

3. Methods

Figure 1 illustrates the overall architecture of our proposed dynamic contrastive learning model. The system begins with an input image, which is augmented and passed through a pre-trained ResNet-50

backbone (initialized with ImageNet weights) to extract high-level visual features. These features are subsequently projected into an embedding space using a dynamic projection head incorporating multi-layer perceptrons and generating normalized representations.

To facilitate contrastive learning, a similarity matrix precomputed from image features is utilized to dynamically construct a positive pool for each generated image. This enables adaptive positive matching by selecting real samples with similarity scores exceeding a predefined threshold or selecting top-k similar examples when insufficient matches are available. Simultaneously, negative samples are obtained through a FIFO memory queue, which stores normalized embeddings from previous mini-batches, ensuring stable and diverse contrastive pairs.

A momentum encoder—synchronized with the online encoder using exponential moving average updates—is employed to encode the positive samples, reducing noise and enhancing representation consistency. Both online and momentum embeddings are input into the InfoNCE contrastive loss function, where the model is trained to minimize the distance between positive pairs while maximizing separation from negative instances.

This dynamic sampling strategy, in combination with momentum encoding and a structured memory queue, significantly enhances the model's capacity for robust representation learning, especially under distributional shifts between generated and real samples.

3.1. MoCo Framework

MoCo (Momentum Contrast), proposed by Kaiming He's team, is a self-supervised learning framework designed to extract effective visual representations from unlabeled data using contrastive learning. The core idea of MoCo involves two key innovations:

First, it introduces a dynamic queue of negative samples, which stores feature representations of previous batches extracted by a momentum encoder. This allows for a large and consistent set of negative examples, which is crucial for effective contrastive learning. Second, MoCo employs a momentum encoder whose parameters are updated using an exponential moving average (EMA) from the online encoder. This design ensures stability in feature representation across different training iterations.

$$\theta_{\text{momentum}} \leftarrow m \cdot \theta_{\text{momentum}} + (1 - m) \cdot \theta_{\text{online}} \tag{1}$$

On the one hand, this mechanism generates stable feature representations, preventing the rapid updates of the online encoder from destabilizing the contrastive learning objective. On the other hand, since the momentum encoder extracts all features in the negative sample queue, it ensures **consistency** within the queue.

Regarding the loss function, MoCo employs the InfoNCE loss (a contrastive loss), which encourages the model to bring positive sample pairs closer in the feature space while pushing apart negative pairs.

$$\mathcal{L} = -\log \frac{\exp(q \cdot k^+/\tau)}{\exp(q \cdot k^+/\tau) + \sum_{k^-} \exp(q \cdot k^-/\tau)} \tag{2}$$

Here, q represents the query feature, k^+ is the positive key feature, k^- is the negative key feature, and τ is the temperature coefficient. The query feature is obtained by feeding an augmented version of the current input image through the online encoder (e.g., a ResNet). The momentum encoder processes a different augmentation of the same image (or another positive sample image) to generate the corresponding key feature.

We adopted the widely used MoCo v2 as the baseline contrastive learning framework, enhancing it with a nonlinear projection head (MLP) and richer data augmentation strategies. To better address the characteristics of the generated image detection task, we made the following extensions to the MoCo v2 framework:

Positive sample selection: Rather than relying solely on data augmentations to create positive pairs, we selected positives based on precomputed semantic similarity between the generated image and real images used during training. This improves the quality and relevance of positive samples.

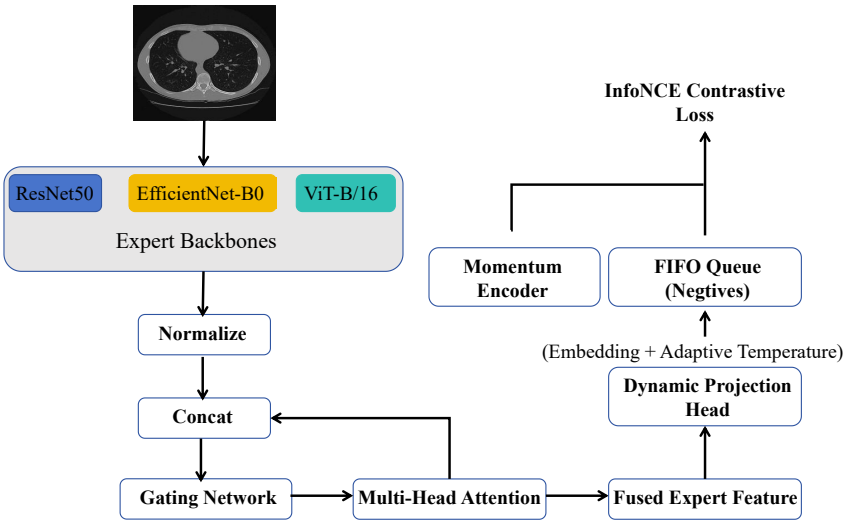

Figure 2: Overview of the MoE Framework

Adaptive temperature coefficient: Instead of using a fixed global temperature, we introduced sample-specific temperature values, allowing the model to adjust its learning focus dynamically for each sample. This is particularly beneficial in tasks with complex feature distributions and varying sample difficulty—such as in generated image detection.

Loss function: We adopted a Hybrid Contrastive Loss, which calculates similarity using negative samples from the queue and incorporates additional supervision from the momentum encoder's perspective. This hybrid formulation combines standard contrastive loss with momentum-aware guidance to enhance representation learning.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_1 + \mathcal{L}_2 \tag{3}$$

$$\mathcal{L}_1 = \alpha \cdot \frac{1}{B} \sum_{i=1}^{B} \left[-\frac{\text{pos\_sim}_i}{\tau_i} + \log \left(\exp\left(\frac{\text{pos\_sim}_i}{\tau_i} \right) + \sum_{j=1}^{K} \exp\left(\frac{q_i \cdot k_j^-}{\tau_i} \right) \right) \right] \tag{4}$$

$$\mathcal{L}_2 = (1-\alpha) \cdot \frac{1}{B} \sum_{i=1}^{B} \left[-\frac{\text{pos\_sim}_i}{\tau_{\text{fixed}}} + \log \left(\exp\left(\frac{\text{pos\_sim}_i}{\tau_{\text{fixed}}} \right) + \sum_{j=1}^{K} \exp\left(\frac{q_i \cdot k_j^-}{\tau_{\text{fixed}}} \right) \right) \right] \tag{5}$$

For each positive sample i in the batch, the cosine similarity between the online feature q_i (denoted as $online_emb$) and the momentum feature k_i^+ (denoted as mom_emb) is computed as follows:

$$\text{pos\_sim}_i = q_i \cdot k_j^+ \tag{6}$$

For each negative sample k_j^- in the queue, the dot product between the online feature q_i and the queue vector is calculated as follows:

$$\text{neg\_sim}_{i,j} = q_i \cdot k_j^- \quad (\text{j} \in \{1, 2, 3, ..., K\}) \tag{7}$$

Here, K denotes the capacity of the queue. The resulting negative sample similarity matrix has the shape $[B, K]$, where B is the batch size. Among them, the formula $\tau_{\text{fixed}} = \tau_i.detach()$ represents that the gradient of temperature values is blocked. The weight parameter α (denoted as $loss_alpha$ in the code) regulates the balance between online loss and momentum loss, with a default value of 0.7.

3.2. Dynamic Projection Head

Traditional contrastive learning frameworks (e.g., MoCo, SimCLR) typically employ fixed linear or multilayer perceptron (MLP) projection heads to merely map high-dimensional features—output by the backbone network—into a lower-dimensional space. As illustrated in the diagram, we extend this architecture in our work, primarily comprising two core enhancements:

Feature Dimension Adaptation: We project the backbone network's 2048-dimensional features (output from ResNet50) into a 256-dimensional contrastive learning space. Dynamic Temperature Generation: We propose generating sample-wise adaptive temperature parameters $\tau \in (0.05, 0.2)$ dynamically based on input features. The temperature paramete τ_i is adapted for each sample as follows:

$$\tau_i = 0.05 + 0.15 \cdot \sigma(W \cdot z_i + b)$$

where σ denotes the Sigmoid function, and z_i is the feature vector after projection.

3.3. Mixture of Experts (MoE)

Mixture of Experts (MoE) is a machine learning paradigm that integrates multiple sub-models (experts) and dynamically weights their outputs. The core idea is to allow different experts to specialize in learning distinct subspaces of the input data, enabling adaptive feature fusion through a gating network that intelligently assigns weights to each expert.

In this work, we introduce a novel design of the MoE architecture tailored to better meet the requirements of our task. Specifically, we incorporate three heterogeneous networks as experts: ResNet50 (for local texture), EfficientNet-B0 (for fine-grained features), and Vision Transformer (for global semantics). After unifying their outputs to the same dimensional space, we apply L2 normalization to eliminate discrepancies in magnitude. A gating network is then constructed to dynamically fuse expert features using a two-stage cascade structure:

$$G(X) = Softmax(W_2 \cdot GELU(W_1 \cdot Concat(f_{res}, f_{eff}, f_{vit}))) \tag{8}$$

$f_{res}, f_{eff}, f_{vit}$ denote the image feature vectors extracted by the three pre-trained models, respectively, and are processed with L2 normalization. The first layer of the gating network is parameterized by a weight matrix $W_1 \in R^{512\times6144}$, which projects the concatenated 6144-dimensional input into a 512-dimensional hidden space. The second layer is parameterized by $W_2 \in R^{3\times512}$, responsible for generating the weight scores corresponding to each expert.

Secondly, a multi-head cross-attention mechanism (with 8 heads) is introduced to enhance inter-expert feature interaction:

$$Attn_{out} = MultiHead(Q = K = V = ExpertFeats) \tag{9}$$

The query, key, and value matrices are all derived from the stacked feature representations of the three experts $ExpertFeats \in R^{B\times3\times2048}$. The MultiHead module employs an 8-head cross-attention mechanism, where each head has a dimensionality of $d_k = 2048/8 = 256$. The overall process is shown in Figure 2.

4. Experiments

4.1. Evaluation Metrics

To simplify our experimental analysis, we did not partition a validation set from the dataset but instead utilized the entire training set for model training. Subsequently, we submitted our results to the IMAGECLEFMED Gans 2025: Recognition of Training Data Fingerprints Challenge. This challenge is formulated as a binary classification task, with evaluation criteria comprising several key performance metrics: the Kappa value, accuracy, precision, recall, and F1-score. Notably, the Kappa value has been

designated as the primary evaluation metric for this year's competition. The definitions for these metrics are as follows:

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

P_o represents the probability of observed agreement, i.e., the proportion of instances where two evaluators (raters) assign the same classification in practice. P_e denotes the probability of expected chance agreement, assuming that the two evaluators classify instances independently and randomly. True Positives (TP) refer to the number of samples where the model correctly predicts the positive class, and the actual class is also positive. False Positives (FP) represent the number of samples where the model incorrectly predicts the positive class for instances that are actually negative. True Negatives (TN) indicate the number of samples where the model correctly predicts the negative class for true negative instances. False Negatives (FN) correspond to the number of samples where the model erroneously predicts the negative class for instances that are actually positive.

4.2. Experimental Results

In this experiment, we evaluated the task across three pre-trained models. During the testing/inference phase, inheriting the objective design from contrastive learning pre-training (a characteristic of the MoCo framework), we primarily employed cosine similarity to assess the relationship between generated images and real images. Based on similarity scores, classifications were performed to derive the Kappa value, accuracy, precision, recall, and F1 score.

To enhance the diversity of the training data and improve the model's generalization capability, we applied data augmentation techniques to the training images, including random cropping, flipping, color jittering, and input normalization. These operations ensured that contrastive learning could effectively discriminate semantic features. Additionally, within the contrastive learning framework, for each generated image, we selected at least two positive samples and five negative samples for training. This approach allowed us to evaluate the performance of different pre-trained models during testing.

By systematically incorporating diverse pre-trained models, we aimed to assess their individual contributions to the final task outcomes. This enabled the integration of these models to construct a suitable Mixture of Experts (MoE) hybrid model. Through this comprehensive experimental design, we could more accurately evaluate the feature extraction capabilities of each pre-trained model for image characterization, providing valuable insights for future improvements in image processing and analysis. Key parameter categories and the specific values used in the experiments are summarized in the table.

For ResNet50, we systematically tested various combinations of the following critical parameters: $Num\_neg\_samples$: The number of negative samples sampled. $Min\_pos\_samples$: The minimum number of positive samples selected. $Loss\_alpha$: The weighting coefficient balancing the online loss and momentum loss. $Batch\_size$: The size of training batches. $Sim\_threshold$: The similarity threshold for classification decisions (based on cosine similarity). These parameters represent a critical hyperparameter set that directly influences model training and evaluation outcomes.

We performed predictions on all 500 generated images and submitted these results. To evaluate model performance, we adopted Cohen's Kappa as the primary evaluation metric due to its significant

Table 1
Some important parameter lists.

Parameter Category	Configuration 1	Configuration 2	Configuration 3
Num_neg_samples	5	1	10
Min_pos_samples	2	1	5
Loss_alpha	0.7	0.7	0.7
Batch_size	128	128	128
Sim_threshold	0.6	0.65	0.6
Learning_rate	3e-4	3e-4	3e-4

Table 2
Presentation of experimental results.

ID	Model type	Kappa	accuracy	precision	recall	F1
1875	ResNet50	0.108	0.554	0.554	0.552	0.553
1874	MoE	-0.032	0.484	0.487	0.604	0.539
1169	EfficientNet-B0	-0.084	0.458	0.465	0.56	0.51
1140	Vision Transformer	-0.02	0.49	0.49	0.70	0.58
1179	ResNet50	0.004	0.502	0.504	0.228	0.314
1107	ResNet50	0.06	0.53	0.539	0.42	0.47

advantages in computer vision tasks involving imbalanced class distributions or scenarios requiring consistency assessment, making it particularly suitable for this task. Additionally, the F1-score was used as a secondary metric, as it holistically integrates precision and recall, providing a more comprehensive performance evaluation. This evaluation protocol ensured a thorough understanding of model performance under varying conditions. In total, we submitted seven distinct sets of results. The table summarizes partial detailed scores, displaying the specific conditions and corresponding evaluation metrics for each submission. These outcomes facilitate further analysis and model refinement to enhance its practical applicability and effectiveness.

As shown in the table, we submitted a total of ten results and selected six representative results for presentation. Through experimentation, we observed that different parameter combinations significantly influence the outcomes. Due to hardware constraints, our parameter combination optimizations were primarily focused on ResNet50. When ID number is 1107, ResNet50 uses Configuration 3 from Table 1. ID number 1179 uses Configuration 2. ID number 1875 uses Configuration 1. Notably, the ResNet50 of configuration 3 demonstrated relatively superior performance on the test set compared to other models.

5. Conclusions

In this study, we employed multiple pre-trained models integrated with contrastive learning frameworks for classification. By defining a similarity threshold, images were classified based on their similarity scores between real and generated images. Features of the generated images were extracted using different pre-trained models and then matched against real images. Depending on the performance of each pre-trained model, their outputs were fused via a Mixture of Experts (MoE) strategy to leverage their complementary strengths. Moving forward, we plan to investigate methods to further optimize the MoE framework, focusing on refining dynamic weighting mechanisms and adaptive model selection. This aims to enhance cross-model collaboration for more precise feature alignment and maximize the collective performance of the integrated models.

Declaration on Generative AI

During the preparation of this work Chat-GPT-4o and Grammarly were used to check grammar and spelling. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, 2017. doi:10.1016/j.media.2017.07.005.

[2] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, Medical Image Analysis 58 (2019). doi:10.1016/j.media.2019.101552.

[3] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification, Neurocomputing 321 (2018). doi:10.1016/j.neucom.2018.09.013.

[4] G. A. Kaissis, M. R. Makowski, D. Rückert, R. F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nature Machine Intelligence 2 (2020). doi:10.1038/s42256-020-0186-1.

[5] R. Shokri, M. Stronati, C. Song, V. Shmatikov, Membership inference attacks against machine learning models, in: Proceedings - IEEE Symposium on Security and Privacy, 2017. doi:10.1109/SP.2017.41.

[6] J. P. Cohen, M. Luck, S. Honari, Distribution matching losses can hallucinate features in medical image translation, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 11070 LNCS, 2018. doi:10.1007/978-3-030-00928-1_60.

[7] S. Tonekaboni, S. Joshi, M. D. McCradden, A. Goldenberg, What clinicians want: Contextualizing explainable machine learning for clinical end use, in: Proceedings of Machine Learning Research, volume 106, 2019.

[8] B. Ionescu, H. Müller, D.-C. Stanciu, A.-G. Andrei, A. Radzhabov, Y. Prokopchuk, Ştefan, Liviu-Daniel, M.-G. Constantin, M. Dogariu, V. Kovalev, H. Damm, J. Rückert, A. Ben Abacha, A. García Seco de Herrera, C. M. Friedrich, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, T. M. G. Pakull, B. Bracke, O. Pelka, B. Eryilmaz, H. Becker, W.-W. Yim, N. Codella, R. A. Novoa, J. Malvehy, D. Dimitrov, R. J. Das, Z. Xie, H. M. Shan, P. Nakov, I. Koychev, S. A. Hicks, S. Gautam, M. A. Riegler, V. Thambawita, P. Halvorsen, D. Fabre, C. Macaire, B. Lecouteux, D. Schwab, M. Potthast, M. Heinrich, J. Kiesel, M. Wolter, B. Stein, Overview of imageclef 2025: Multimedia retrieval in medical, social media and content recommendation applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 16th International Conference of the CLEF Association (CLEF 2025), Springer Lecture Notes in Computer Science LNCS, Madrid, Spain, 2025.

[9] A.-G. Andrei, M. G. Constantin, M. Dogariu, A. Radzhabov, L.-D. Ştefan, Y. Prokopchuk, V. Kovalev, H. Müller, B. Ionescu, Overview of imageclefmedical 2025 GANs task: Training data analysis and fingerprint detection, in: CLEF2025 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Madrid, Spain, 2025.

[10] H. C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, M. Michalski, Medical image synthesis for data augmentation and anonymization using generative adversarial networks, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 11037 LNCS, 2018. doi:10.1007/978-3-030-00536-8_1.

[11] A. Chartsias, T. Joyce, M. V. Giuffrida, S. A. Tsaftaris, Multimodal mr synthesis via modality-

invariant latent representation, IEEE Transactions on Medical Imaging 37 (2018). doi:10.1109/TMI.2017.2764326.
[12] J. Hayes, L. Melis, G. Danezis, E. D. Cristofaro, Logan: Membership inference attacks against generative models, Proceedings on Privacy Enhancing Technologies 2019 (2019). doi:10.2478/popets-2019-0008.
[13] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, I. Išgum, Deep mr to ct synthesis using unpaired data, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10557 LNCS, 2017. doi:10.1007/978-3-319-68127-6_2.
[14] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (2021). doi:10.1109/TPAMI.2020.2970919.
[15] D. P. Kingma, M. Welling, Auto-encoding variational bayes, in: 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, 2014. doi:10.61603/ceas.v2i1.33.
[16] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: Advances in Neural Information Processing Systems, volume 2020-December, 2020.

A. Online Resources

The sources for the ceur-art style are available via

- GitHub,
- Overleaf template.