RadiomicsFill-Mammo: Synthetic Mammogram Mass Manipulation with Radiomics Features

Inye Na[1], Jonghun Kim[1], Eun Sook Ko[2], and Hyunjin Park[1,3]*

[1] Department of Electrical and Computer Engineering,
Sungkyunkwan University, Suwon 16419, South Korea
[2] Department of Radiology, Samsung Medical Center,
Sungkyunkwan University School of Medicine, Seoul 06351, South Korea
[3] Center for Neuroscience Imaging Research,
Institute for Basic Science, Suwon 16419, South Korea
hyunjinp@skku.edu

Abstract. Motivated by the question, "Can we generate tumors with desired attributes?" this study leverages radiomics features to explore the feasibility of generating synthetic tumor images. Characterized by its low-dimensional yet biologically meaningful markers, radiomics bridges the gap between complex medical imaging data and actionable clinical insights. We present **RadiomicsFill-Mammo**, the first of the *Radiomics-Fill* series, an innovative technique that generates realistic mammogram mass images mirroring specific radiomics attributes using masked images and opposite breast images, leveraging a recent stable diffusion model. This approach also allows for the incorporation of essential clinical variables, such as BI-RADS and breast density, alongside radiomics features as conditions for mass generation. Results indicate that *RadiomicsFill-Mammo* effectively generates diverse and realistic tumor images based on various radiomics conditions. Results also demonstrate a significant improvement in mass detection capabilities, leveraging *RadiomicsFill-Mammo* as a strategy to generate simulated samples. Furthermore, *RadiomicsFill-Mammo* not only advances medical imaging research but also opens new avenues for enhancing treatment planning and tumor simulation. Our code is available at https://github.com/nainye/RadiomicsFill.

Keywords: Synthetic Tumor Generation · Radiomics Features · Mammography.

1 Introduction

In medical imaging, acquiring high-quality data is challenging due to privacy concerns and high labeling costs. Additionally, the inherent imbalance in available datasets complicates the training of effective neural networks, presenting a significant barrier to advancements in medical artificial intelligence (AI) [22,2].

---

* Corresponding author

To overcome these challenges, researchers have explored new methods, including the generation of synthetic data using advanced techniques like Generative Adversarial Networks and diffusion models [21,1,9,3,15,12]. Rouzrokh et al. [18] developed a denoising diffusion model for brain MRI slice inpainting, enabling customization with five types of masks, including tumor-free and various tumor states. Sagers et al. [19] utilized DALL·E 2 to generate photorealistic skin disease images for diverse skin types, improving classification accuracy, notably for underrepresented groups. These methods aim to bridge the gaps in data availability and balance but often fall short in customizing synthetic data with specific features.

Radiomics features, despite being lower in dimension than the full images, carry important biological information that has proven useful in diagnosis and prognosis [23,13,6]. Radiomics is less likely to overfit compared to models trained directly on images, especially when the available data is limited [16,25].

Building on this foundation, we introduce *RadiomicsFill-Mammo*, a novel approach that leverages radiomics features for generating synthetic mammogram images. This method can help balance out data and provide more specific control over the generation of realistic tumors. Our approach utilizes stable diffusion [17] as the backbone generator for *RadiomicsFill-Mammo*, selected for its denoising-based operation and flexibility in handling various input formats, which is crucial for mammography. The generator's pretraining on large-scale datasets ensures a robust foundation, enhancing the quality and realism of synthetic images. As the first in the *RadiomicsFill* series, *RadiomicsFill-Mammo* demonstrates the practicality and transformative potential of employing radiomics features in medical imaging research and clinical practices. Our main contributions are:

1. Development of a prompt encoder capable of learning and reflecting specific conditions from tabular radiomics features for accurate tumor generation.
2. Incorporation of clinical conditions and optimization for mammography environments, including images of the opposite breast, breast density, and Breast Imaging Reporting & Data System (BI-RADS) ratings, enhancing medical applicability.
3. Evidence of *RadiomicsFill-Mammo*'s effectiveness as an augmentor in downstream tasks, particularly in mass detection, underscoring its role in enhancing model performance across medical AI applications.
4. Demonstrated adaptability to small external datasets, showcasing the potential for broad application and research in medical imaging.

2   Method

2.1   Dataset

For our study, we employed the publicly available VinDr-Mammo dataset [14] from Vietnam, containing 5,000 full-field digital mammography cases with breast density and annotation details. The data were pre-split into training and test sets
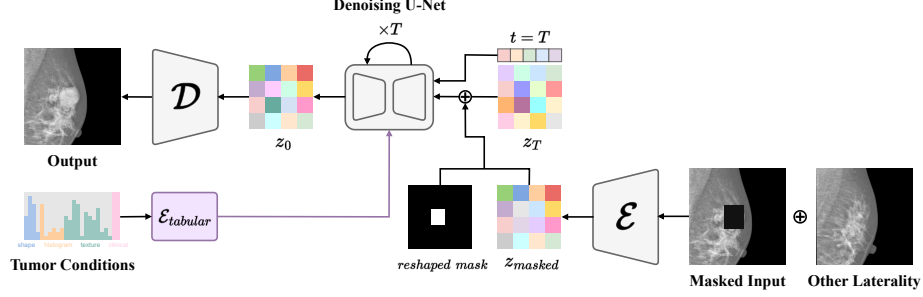
Fig. 1. Overview of the *RadiomicsFill-Mammo* model, a stable diffusion-based architecture. This model performs iterative denoising for tumor inpainting within masked regions on a noisy latent vector, utilizing information from unmasked regions, the opposite breast, and specific tumor conditions.

at a 4:1 ratio at the patient level. Our focus was on samples with either a Mass finding or No findings. Due to the provision of only bounding boxes for the mass regions, we worked with an experienced breast radiologist to semi-automatically (i.e., vanilla U-Net followed by manual correction) delineate mass contours for extracting radiomics features, employing these in our experiments.

The data utilized in our research was distributed as follows: for mass cases, we allocated 886 masses for training (423 subjects; 26 low density-benign [L-B], 96 low density-malignant [L-M], 376 high density-benign [H-B], and 388 high density-malignant [H-M]), 96 for validation (46 subjects; 4 L-B, 4 L-M, 38 H-B, and 50 H-M), and 235 for testing (115 subjects; 12 L-B, 14 L-M, 110 H-B, and 99 H-M). For normal cases, 12,680 images for training (3,170 subjects), 1,408 for validation (352 subjects), and 3,540 for testing (885 subjects) were used. We trained our model without distinguishing between left/right and cranial-caudal/mediolateral oblique (CC/MLO) mages, incorporating all variations into a single model to leverage the comprehensive dataset for enhanced learning outcomes.

2.2   RadiomicsFill-Mammo

RadiomicsFill-Mammo, as illustrated in Fig. 1, is a framework designed for inpainting masses in masked regions of mammograms. It takes as input the masked image, an image of the opposite breast (other laterality image), and tumor conditions. The opposite side image is included to leverage breast symmetry, aiding in inpainting by providing a reference [20]. In 2D mammograms, normal and abnormal tissues overlap due to projection, and a rectangular mask may contain both normal and abnormal tissue depending on the radiomics condition. Using the opposite side helps fill the masked region with normal tissue texture while generating masses that reflect the given conditions.

The generator within this framework is adapted from the stable-diffusion-2-inpainting [17] model, with modifications made to suit our specific requirements.
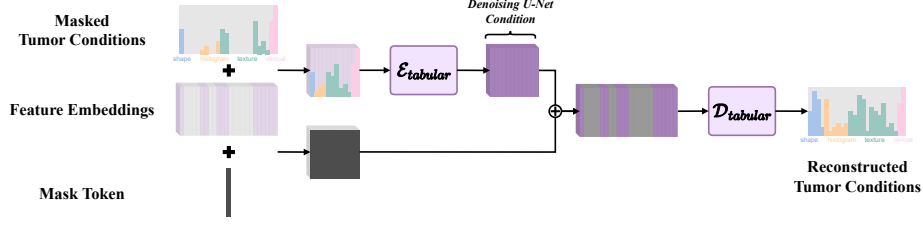
Fig. 2. Pretraining Process of the Tabular Encoder. Unmasked feature values (i.e., radiomics features) and their embeddings are concatenated for the encoder input. For masked features, a mask token and the masked feature's embeddings are concatenated, forming the input for the decoder, which focuses on reconstructing tumor conditions. Following pretraining, the encoder is frozen, and its output is utilized for tumor inpainting.

To improve tumor generation, we replace the standard text encoder with a pretrained tabular encoder based on radiomics features, as detailed in Section 2.3. This choice tailors the generator's capabilities more closely to medical imaging requirements. We specifically fine-tune the denoising U-Net for mass inpainting, ensuring precise integration of tumor masses into mammograms. This focused adjustment enhances the generated images' realism and clinical relevance.

2.3 Tabular Encoder

We aim to produce prompts detailing tumor characteristics for integration into tumor generation within the stable diffusion framework. A key component of our *RadiomicsFill* model is the tabular encoder, trained as part of the Masked Encoding for Tabular data (MET) [10] process, incorporating its strategy for feature-specific embeddings. This encoder effectively identifies complex feature relationships by reconstructing masked values, acknowledging the stable but variable nature of these characteristics. While some features may be redundant, their relationships are effectively captured by the feature-specific embeddings. The embeddings are structured as *feature number × embedding dimension*, allowing for precise representation of each feature's unique properties. After pretraining on a masked feature reconstruction task with mean squared error loss, this encoder becomes the prompt encoder in our tumor generator. Fig. 2 demonstrates the MET network's architecture and its pretraining approach, enabling the generation of clinically relevant and detailed images by leveraging the structured complexity of tabular data.

Evaluation of the Tabular Encoder We adopted a Masked Tumor Condition, incorporating 9 shape, 18 histogram, and 40 texture features, alongside clinical variables (i.e., breast density as low/high and BI-RADS assessments as normal, benign, and malignant). Radiomics features were extracted using PyRadiomics [24] and normalized to a range of 0 to 1. In the case of normal samples, which lack
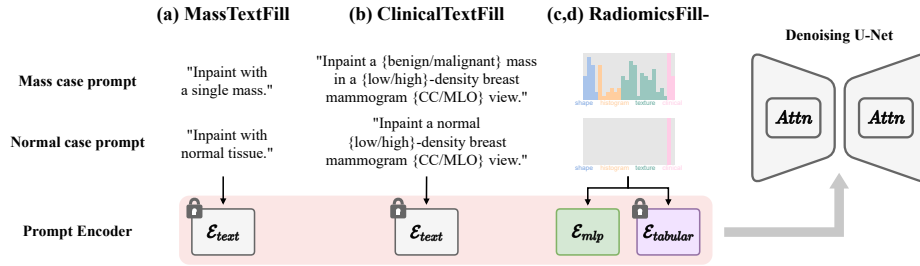
(a) MassTextFill   **(b) ClinicalTextFill**   **(c,d) RadiomicsFill-**

Denoising U-Net

Mass case prompt   "Inpaint with a single mass."   "Inpaint a {benign/malignant} mass in a {low/high}-density breast mammogram {CC/MLO} view."

Normal case prompt   "Inpaint with normal tissue."   "Inpaint a normal {low/high}-density breast mammogram {CC/MLO} view."

Prompt Encoder $\mathcal{E}_{text}$ $\mathcal{E}_{text}$ $\mathcal{E}_{mlp}$ $\mathcal{E}_{tabular}$ *Attn*   *Attn*

Fig. 3. Different Configurations of Prompt Encoders. Each configuration employs a prompt encoder specialized for its prompt type. Configurations (a) Mass-TextFill, (b) ClinicaltextFill, and (d) RadiomicsFill-MET utilize pretrained and frozen prompt encoders, whereas the encoder for (c) RadiomicsFill-MLP consists of three MLP layers and is trained alongside the generator.

mass, all radiomics features were set to 0, without any masking in the training process.

To evaluate the pretrained tabular encoder's effectiveness, we employed logistic regression for BI-RADS prediction based on radiomics features, with clinical variables masked during evaluation. For BI-RADS prediction, the baseline model used only the 67 radiomics features, whereas the MET model also included trained feature embeddings, with the encoder output used for logistic regression. This approach allowed for hyperparameter optimization, including the embedding dimension and encoder-decoder architecture, with outcomes detailed in the supplementary materials.

Furthermore, we enhanced learning efficiency by adopting a curriculum learning strategy that progressively increased the feature masking rate.

2.4   Experimental Setting

To assess the performance of *RadiomicsFill-Mammo*, we experimented with four configurations for tumor condition prompts and encoders, as detailed below and illustrated in Fig. 3:

- **(a) MassTextFill:** Simplifies the approach by categorizing inputs into two broad categories: mass and normal cases.
- **(b) ClinicalTextFill:** Utilizes 12 distinct text prompts based on breast density, BI-RADS assessment, and view for more nuanced clinical inputs.
- **(c) RadiomicsFill-MLP** and **(d) RadiomicsFill-MET:** Both configurations use 67 radiomics features and 2 clinical variables (breast density and BI-RADS) as prompts but differ in their prompt encoder architectures, exploring various encoding strategies.

During the denoising U-Net training phase, normal and mass prompts are used to inpaint images, facilitating the learning of distinctions between normal
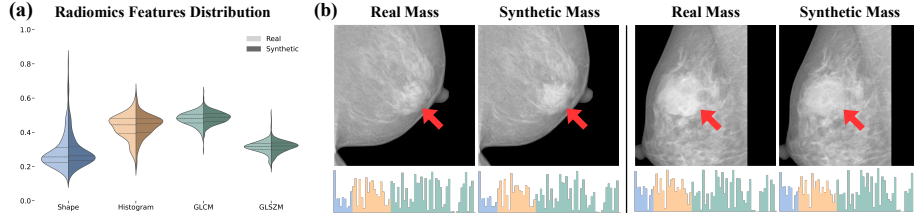
Fig. 4. Radiomics Feature Consistency Between Real and Synthetic Masses.
(a) Grouped violin plots for shape, histogram, and texture (GLCM and GLSZM) features, contrasting real (left) and synthetic (right) distributions. (b) Visual comparison of real and synthetic masses, with red arrows highlighting masses and histograms visualizing the 67 extracted radiomics features underneath.

tissue and tumors. For inference, we applied Classifier-Free Guidance (CFG) with tailored prompts to enhance image specificity and clinical relevance [4]. Positive prompts detailed tumor characteristics, while negative prompts ensured differentiation from normal cases. This CFG approach significantly improves the model's accuracy by focusing on tumor-specific attributes and avoiding features typical of normal cases.

3   Results

3.1   How Well Does It Generate Diverse and Realistic Tumors?

Radiomics Feature Consistency We assessed *RadiomicsFill*'s ability to generate tumors reflecting specified radiomics features by comparing the radiomics features of generated tumors against those of the input image in unseen test set conditions. We employed a pretrained nnUNet model [7], trained on the mass training set, to predict synthetic tumor contours from mass images and bounding box mask images. The high similarity observed in Fig. 4 between input and synthetic tumor radiomics features confirms *RadiomicsFill*'s effectiveness in accurately mirroring the given conditions in tumor generation.

Quality Assessment of Synthetic Tumor   All models, leveraging tailored prompts, consistently generated high-quality synthetic tumors. The performance was measured with metrics reflecting image fidelity: FID (Fréchet Inception Distance), PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), and MS-SSIM (Multi-Scale Structural Similarity Index). Lower FID scores indicate better image quality, while higher PSNR, SSIM, and MS-SSIM scores suggest greater similarity to the target images. The strong performance across all prompt types suggests that the stable diffusion backbone is effective as a tumor generator. The results in Table 1 highlight the consistent excellence in synthetic image quality.

Table 1. Synthetic Tumor Quality Assessment.

Prompt type	FID $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	MS-SSIM $\uparrow$
MassTextFill	0.1144	32.4783	0.8395	0.9461
ClinicalTextFill	0.0989	**33.1762**	0.8430	0.9495
RadiomicsFill-MLP	0.0874	32.7828	0.8322	0.9498
RadiomicsFill-MET	**0.0570**	32.8009	**0.8535**	**0.9507**

Table 2. Mass detection performance on the VinDr-Mammo test set. AP; average precision. mAP; mean AP.

Prompt type	L-B AP	L-M AP	H-B AP	H-M AP	mAP$_{50}$	mAP$_{50-95}$
Baseline	0.7500	1.0000	0.4907	0.5056	0.5438	0.3569
MassTextFill-1x	0.5833	0.7500	0.4352	0.3708	0.4378	0.2488
MassTextFill-2x	0.7778	0.8333	0.4815	0.4045	0.4900	0.3301
ClinicalTextFill-1x	0.5417	0.6667	0.4047	0.2921	0.3871	0.2245
ClinicalTextFill-2x	0.9444	1.0000	0.5648	0.5281	0.6006	0.3754
RadiomicsFill-MLP-1x	0.5833	0.8333	0.2963	0.3708	0.3733	0.2227
RadiomicsFill-MLP-2x	0.8333	0.9167	0.5648	0.5730	0.6037	0.3784
RadiomicsFill-MET-1x	0.5000	0.7500	0.3796	0.4157	0.4240	0.2550
RadiomicsFill-MET-2x	**1.0000**	**1.0000**	**0.5648**	**0.6292**	**0.6429**	**0.3935**

3.2   How Can *RadiomicsFill-Mammo* Be Utilized?

Enhancing Downstream Tasks   In evaluating *RadiomicsFill-Mammo*'s impact on mass detection, we generated synthetic data using normal training images with mass prompts. The generated data were used for training a YOLO v8 detection model [8]. We trained models with two regimes: '1x' using only synthetic data, and '2x' combining synthetic data with the actual mass training set. For comparison, the baseline model was trained solely on the real mass training set. To assess mass detection performance, we conducted a comprehensive comparison using commonly employed metrics, with results detailed in Table 2.

Visualizing Condition-Specific Tumor Variations   Fig. 5 showcases synthetic tumors generated with progressively increasing the shape feature of size, while other features remain constant. This visualization demonstrates our model's ability to adapt to specific tumor characteristics, illustrating the potential for generating diverse, tailored synthetic tumors based on specific condition prompts.

Fine-Tuning with External Validation Dataset   This section underscores *RadiomicsFill-Mammo*'s capability to adapt to external datasets, using the INbreast dataset [11] from Portugal as a case study. To address data scarcity and enhance mass detection, we specifically fine-tuned the denoising U-Net of *RadiomicsFill-Mammo* by training only the added Low-Rank Adaption (LoRa) module [5]. This focused fine-tuning strategy demonstrates an effective adaptation even with limited data availability. To enrich the dataset, we generated
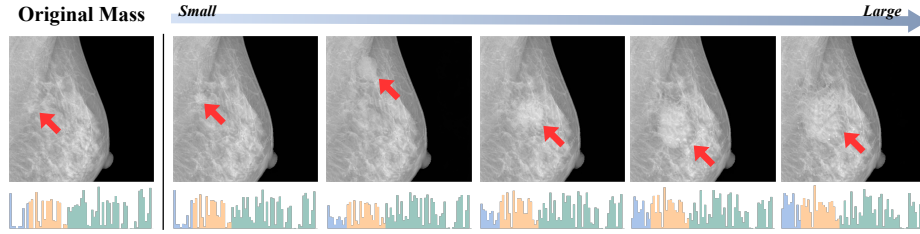
Original Mass   *Small*   *Large*

Fig. 5. Progressive Shape Variation in Synthetic Tumors. Starting with an image containing a mass, we use a normal prompt to replace the mass area with normal tissue. Then, at random positions, we generate synthetic tumors of increasing size, keeping histogram and texture features consistent. Red arrows in each image point to the mass. Below each image, histograms display the radiomics features extracted from the tumor areas.

Table 3. Mass detection performance on the INbreast dataset.

Prompt type	L-B AP	L-M AP	H-B AP	H-M AP	mAP$_{50}$	mAP$_{50-95}$
Baseline	0.3333	0.7778	0.3846	0.2500	0.5800	0.3978
MassTextFill	0.5000	0.7778	0.6923	0.2500	0.6800	0.4511
ClinicalTextFill	0.5000	0.7778	0.6154	0.5000	0.6800	0.4533
RadiomicsFill-MLP	**0.6667**	**0.8519**	0.4615	0.5000	0.7000	0.4689
RadiomicsFill-MET	0.5000	0.8148	**0.7692**	**0.5000**	**0.7400**	**0.4711**

synthetic masses on INbreast's normal images using mass prompts from the VinDr-Mammo trainset, generating an augmented dataset for training the mass detection model. Due to the limited size of the INbreast dataset, we proceeded directly to test set evaluation after 300 epochs of balanced training, without employing a separate validation set. The training and test sets included 7 L-B, 30 L-M, 4 H-B, and 3 H-M masses for training, and 9 L-B, 34 L-M, 18 H-B, and 6 H-M masses for testing, respectively using stratified sampling at the patient level. The improvements in diagnostic accuracy achieved through this refined fine-tuning method, as evidenced in Table 3, highlight the enhanced utility of *RadiomicsFill-Mammo* in real-world scenarios, even with limited available data.

4   Conclusion

Our study introduces the pioneering application of utilizing radiomics features for generating synthetic tumors. We have successfully validated the generator's precision in mirroring specified tumor conditions, markedly enhancing mass detection capabilities when used as a data augmentor. This improvement is especially significant for high-density masses, which are often difficult to detect due to their reduced visibility in dense breast tissue. The findings demonstrate that defining specific tumor conditions in conjunction with an appropriately configured prompt encoder enables the generator to produce a wide range of synthetic

tumors, preventing the generation of typical results. The innovative framework of *RadiomicsFill-Mammo*, which incorporates opposite breast images and clinical variables such as breast density and BI-RADS assessment as additional conditions for tumor generation, establishes a clinically meaningful approach. As the first endeavor in the *RadiomicsFill* series, *RadiomicsFill-Mammo* unveils substantial potential to advance the field of medical imaging research, achieving significant outcomes. The ability of *RadiomicsFill* to generate tumors with specific attributes highlights its utility for a range of clinical purposes, including simulating treatment responses. Furthermore, this work paves the way for applying *RadiomicsFill* techniques across different medical imaging fields, setting a promising direction for our future research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018)
2. Gao, L., Zhang, L., Liu, C., Wu, S.: Handling imbalanced medical image data: A deep-learning-based one-class classification approach. Artificial intelligence in medicine **108**, 101935 (2020)
3. Garcea, F., Serra, A., Lamberti, F., Morra, L.: Data augmentation for medical imaging: A systematic literature review. Computers in Biology and Medicine **152**, 106391 (2023)
4. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022)
5. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
6. Ibrahim, A., Primakov, S., Beuque, M., Woodruff, H., Halilaj, I., Wu, G., Refaee, T., Granzier, R., Widaatalla, Y., Hustinx, R., et al.: Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework. Methods **188**, 20–29 (2021)
7. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
8. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics yolov8 (2023), https://github.com/ultralytics/ultralytics

9. Kim, J., Park, H.: Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 7604–7613 (January 2024)

10. Majmundar, K., Goyal, S., Netrapalli, P., Jain, P.: Met: Masked encoding for tabular data. arXiv preprint arXiv:2206.08564 (2022)

11. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: Inbreast: toward a full-field digital mammographic database. Academic radiology **19**(2), 236–248 (2012)

12. Na, I., Kim, J., Park, H.: Synthetic tumor manipulation: With radiomics features. NeurIPS Workshop on Medical Imaging meets NeurIPS (2023)

13. Na, I., Noh, J.J., Kim, C.K., Lee, J.W., Park, H.: Combined radiomics-clinical model to predict platinum-sensitivity in advanced high-grade serous ovarian carcinoma using multimodal mri. Frontiers in Oncology **14** (2024)

14. Nguyen, H.T., Nguyen, H.Q., Pham, H.H., Lam, K., Le, L.T., Dao, M., Vu, V.: Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. Scientific Data **10**(1), 277 (2023), https://vindr.ai/datasets/mammo

15. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging (2023)

16. Parmar, C., Rios Velazquez, E., Leijenaar, R., Jermoumi, M., Carvalho, S., Mak, R.H., Mitra, S., Shankar, B.U., Kikinis, R., Haibe-Kains, B., et al.: Robust radiomics feature quantification using semiautomatic volumetric segmentation. PloS one **9**(7), e102107 (2014)

17. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10684–10695 (June 2022)

18. Rouzrokh, P., Khosravi, B., Faghani, S., Moassefi, M., Vahdati, S., Erickson, B.J.: Multitask brain tumor inpainting with diffusion models: A methodological report. arXiv preprint arXiv:2210.12113 (2022)

19. Sagers, L.W., Diao, J.A., Groh, M., Rajpurkar, P., Adamson, A.S., Manrai, A.K.: Improving dermatology classifiers across populations using images generated by large diffusion models. arXiv preprint arXiv:2211.13352 (2022)

20. Scutt, D., Lancaster, G.A., Manning, J.T.: Breast asymmetry and predisposition to breast cancer. Breast cancer research **8**(2), 1–7 (2006)

21. Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3. pp. 1–11. Springer (2018)

22. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of big data **6**(1), 1–48 (2019)

23. Traverso, A., Wee, L., Dekker, A., Gillies, R.: Repeatability and reproducibility of radiomic features: a systematic review. International Journal of Radiation Oncology* Biology* Physics **102**(4), 1143–1158 (2018)

24. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational

radiomics system to decode the radiographic phenotype. Cancer research **77**(21), e104–e107 (2017)

25. Zwanenburg, A., Leger, S., Agolli, L., Pilz, K., Troost, E.G., Richter, C., Löck, S.: Assessing robustness of radiomic features by image perturbation. Scientific reports **9**(1), 614 (2019)
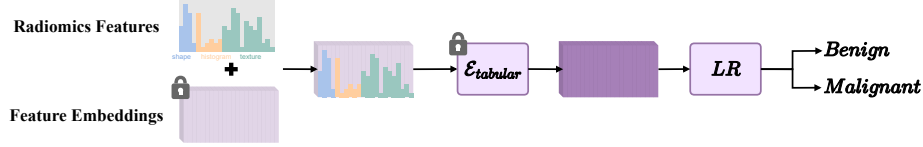
Supplementary Material

Fig. A. Evaluating Framework for the Efficiency of the Pretrained Tabular Encoder. During the training phase, clinical variables were included alongside radiomics features. For evaluation, all clinical variables were masked, and only the 67 radiomics features were fed into the encoder. The encoder's output, once flattened, served as input for a logistic regression model tasked with a binary classification of BI-RADS into benign and malignant categories. The predictive performance of this model highlights the efficacy of MET in capturing clinically relevant information from radiomics features alone.
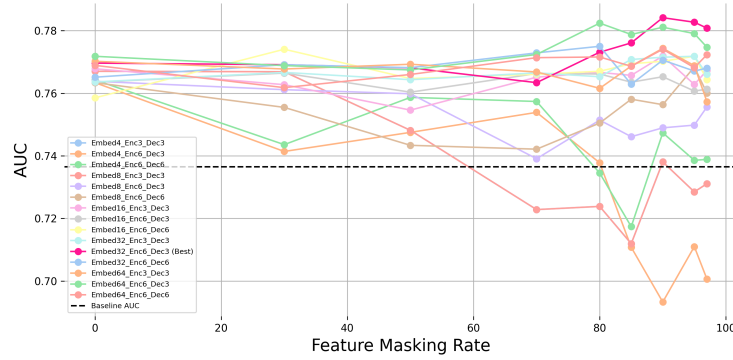

Fig. B. Optimizing MET Parameters: AUC Performance Across Masking Rates. This graph tracks AUC scores for benign/malignant predictions at various feature masking rates to determine MET's optimal parameters. Highlighted in hot pink, the best-performing setup—featuring a 32-dimensional feature embedding, 6 encoder layers, and 3 decoder layers at a 90% masking rate—was chosen as the prompt encoder configuration for RadiomicsFill-MET.
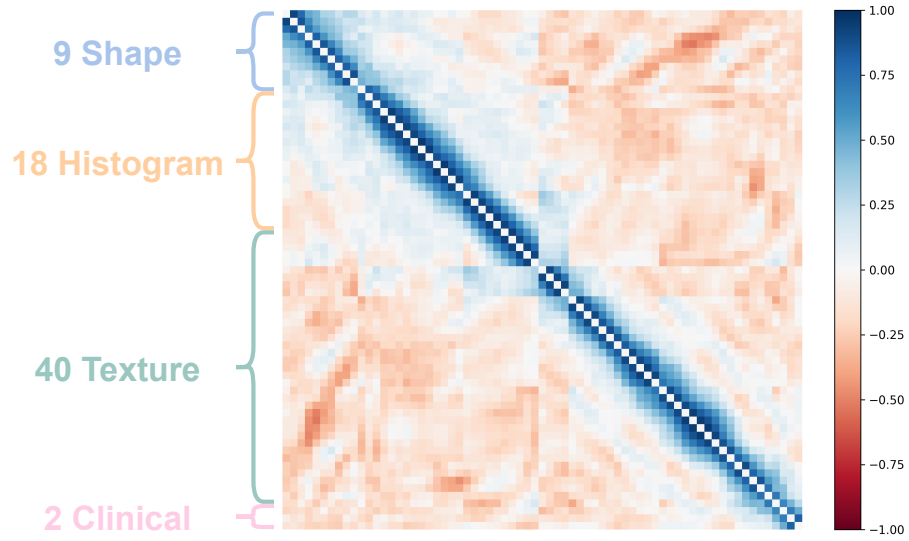
Fig. C. Cosine Similarity between MET Feature Embeddings. This heatmap visualizes the cosine similarity for MET's 69 feature-specific embeddings, highlighting strong intra-group correlations within shape, histogram, and texture clusters. The clear block diagonal structure observed signifies the encoder's proficiency in grouping and differentiating feature categories.