

Harnessing EHRs for Diffusion-based Anomaly Detection on Chest X-rays

Harim Kim¹, Yuhan Wang², Minkyu Ahn¹, Heeyoul Choi¹, Yuyin Zhou², and Charmgil Hong¹

¹ Handong Global University, Pohang, South Korea
{hrkim, minkyuahn, hchoi, charmgil}@handong.ac.kr

² University of California, Santa Cruz, CA, USA
{ywan1332, yzhou284}@ucsc.edu

Abstract. Unsupervised anomaly detection (UAD) in medical imaging is crucial for identifying pathological abnormalities without requiring extensive labeled data. However, existing diffusion-based UAD models rely solely on imaging features, limiting their ability to distinguish between normal anatomical variations and pathological anomalies. To address this, we propose Diff3M, a multi-modal diffusion-based framework that integrates chest X-rays and structured Electronic Health Records (EHRs) for enhanced anomaly detection. Specifically, we introduce a novel Image-EHR Cross-Attention module to incorporate structured clinical context into the image generation process, improving the model’s ability to differentiate normal from abnormal features. Additionally, we develop a static masking strategy to enhance the reconstruction of normal-like images from anomalies. Extensive evaluations on CheXpert and MIMIC-CXR/IV demonstrate that Diff3M achieves state-of-the-art performance, outperforming existing UAD methods in medical imaging. Our implementation is available at <https://github.com/nth221/Diff3M>.

Keywords: Diffusion Models · Multi-modality · Anomaly Detection

1 Introduction

Recent AI advancements have remarkably improved radiographic image analysis, providing crucial diagnostic support to radiologists [7,9,19]. However, current clinical practices rely on AI tools designed for predefined disease predictions [21]. This reliance raises the risk of overlooking clinically significant but previously uncharacterized pathological features. To overcome this limitation, unsupervised anomaly detection (UAD) offers a promising direction, with the capability to identify undefined pathological features without prior annotation [2].

Generative model-based anomaly detection has gained interest in recent years [1,25], distinguishing anomalies by regenerating input data into anatomically plausible representations [26,25]. Among these approaches, diffusion-based models excel in industrial domains, due to their powerful generation capabilities [11,20]. These models extract semantic features characterizing normal data and

use them as a reference to reconstruct the input into a normal-like image. However, the variability in individual anatomical structures within radiographic images complicates the extraction of features that manifest normal data, making it difficult to enhance detection performance [3].

To address this challenge, we propose a Diffusion-based Multi-modal Medical Anomaly Detection (Diff3M) framework that leverages both chest X-rays and structured Electronic Health Records (EHRs) for enhanced anomaly detection. While previous studies show that integrating EHR data with medical images enhances clinical prediction [30,32], its potential in anomaly detection remains largely unexplored. To bridge this gap, we introduce a novel Image-EHR Cross-Attention (IECA) mechanism that allows EHR data to provide additional semantic context, enabling the model to differentiate between normal anatomical variations and pathological anomalies. This mitigates the impact of anatomical variability in the generation process, resulting in more reliable anomaly detection. Additionally, we propose a novel masking strategy for reconstructing anatomically normal-appearing images. Existing diffusion-based anomaly detection methods primarily employ random masking with predefined textures [13,31], which lacks generalizability. In contrast, we introduce a Pixel-level Checkerboard Masking (PCM) module, designed to capture a broader range of anomalies by incorporating static masking patterns that generalize beyond specific textures. Throughout the reverse diffusion process, Diff3M reconstructs masked regions into normal-like images by leveraging both the noised input and EHR-guided semantic embeddings, enhancing its ability to detect and characterize anomalies effectively.

We conduct comprehensive experiments using two major chest radiography datasets, CheXpert [14] and MIMIC-CXR/IV [16,15], both offering rich collections of chest X-rays paired with clinical information. Our evaluation, combining quantitative metrics and qualitative analysis, shows that the proposed method outperforms existing UAD approaches for medical images. The contributions of this paper can be summarized as follows:

1. The first diffusion-based medical anomaly detection framework that effectively integrates EHR data as conditioning.
2. A novel Image-EHR Cross-Attention module to generate EHR embeddings that incorporate structured clinical information into the detection pipeline.
3. A new static masking strategy to enhance the reconstruction of normal-like images from anomalies.
4. State-of-the-art performance in medical UAD, surpassing existing methods.

2 Background

Diffusion-based UAD Diffusion-based models reconstruct anomalies into normal-like images. AnoDDPM [29] improves normal-like image generation using Simplex noise [22] but underperforms compared to standard denoising autoencoders [6]. mDDPM [13] applies random masking but targets specific anomalies rather than undefined abnormalities. DDAD [20] and the more recent DiAD [11]

condition the reverse process on extracted image features, which may mix normal and anomalous traits, leading to generalization issues.

Feature-based UAD Feature-based UAD methods detect anomalies by analyzing features extracted from pre-trained foundation models [2]. PaDiM, CFA and PatchCore [4,18,24] use memory bank to compare features from normal training and testing data, while RD4AD [5] and MambaAD [10] employ knowledge distillation to detect anomalies by measuring discrepancies between teacher and student networks. However, the general-purpose foundation models may struggle to extract meaningful medical features.

Diffusion Models The Denoising Diffusion Probabilistic Model (DDPM) [12] consists of two main processes: a forward process $q(\cdot)$ and a reverse process $p_\theta(\cdot)$. The forward process gradually adds Gaussian noise to input data \mathbf{x} over T steps, mapping it to a standard normal distribution. Each step t ($t = 1, \dots, T$) introduces noise $\epsilon^{(t)} \sim N(0, 1)$, scaled by a variance parameter $\beta_t \in (0, 1)$ like $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$. The reverse process is a learnable procedure that reconstructs data from \mathbf{x}_T to generate sample \mathbf{x}_0 resembling the training data. It is trained by minimizing the KL divergence of $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ and $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$. It can be interpreted as approximating each step of the reverse process as the inverse operation of the corresponding forward process step. Consequently, the model learns to predict the noise $\epsilon^{(t)}$ using $\mathcal{L} := \|\epsilon_\theta^{(t)}(\mathbf{x}_t) - \epsilon^{(t)}\|_2^2$.

The Denoising Diffusion Implicit Model (DDIM) [27] enhances DDPM efficiency by reformulating the forward process with a reversed ordinary differential equation (ODE), making it deterministic and non-Markovian [28]:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \sqrt{\bar{\alpha}_{t+1}} \left[\left(\sqrt{\frac{1}{\bar{\alpha}_t}} - \sqrt{\frac{1}{\bar{\alpha}_{t+1}}} \right) \mathbf{x}_t + \left(\sqrt{\frac{1}{\bar{\alpha}_{t+1}}} - 1 - \sqrt{\frac{1}{\bar{\alpha}_t}} + 1 \right) \epsilon_\theta^{(t)} \right]. \quad (1)$$

where $\bar{\alpha}_t := \prod_{s=1}^t (1 - \beta_s)$. Since the processes are no longer Markovian, DDIM enables sample generation with fewer steps than T . The method proposed in this paper is built upon DDIM.

3 Proposed Framework

We present **Diff3M**, a diffusion-based UAD framework trained exclusively on normal data (Fig. 1). It comprises two key steps: (1) a *Noise Prediction* (NP) Network for estimating noise $\epsilon_\theta^{(t)}$ in the reverse diffusion process and (2) a *Masked Pixel Generation* (MPG) Network for reconstructing masked inputs into anatomically plausible images $\tilde{\mathbf{x}}_t$. To effectively integrate EHRs for enhanced anomaly detection, we propose two key designs. First, we introduce a novel *Image-EHR Cross Attention* (IECA) module, which extracts EHR embeddings from paired X-ray images \mathbf{x} and EHR data \mathbf{r} , conditioning the MPG Network and the NP Network (Sec. 3.1). This enables a more semantically informed diffusion process. Additionally, following the original DDPM framework [12], we incorporate

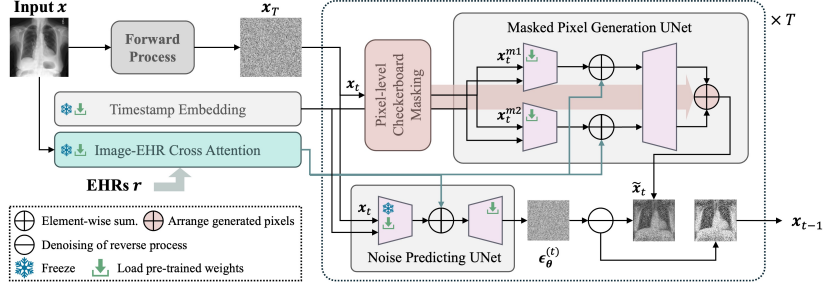


Fig. 1. Training process of Diff3M. Image-EHR Cross Attention (IECA) incorporates EHR embeddings as conditional inputs for the noise prediction (NP) network at each reverse process step. A novel Pixel-level Checkerboard Masking (PCM) strategy is employed to improve generating masked inputs into normal-like images from the Masked Pixel Generation (MPG) Network.

timestamp embeddings to capture temporal information. Second, we propose a *Pixel-level Checkerboard Masking* (PCM) strategy to enhance image reconstruction by generating masked inputs from \mathbf{x}_t , which the MPG Network then reconstructs into normal-appearing images (Sec. 3.2).

3.1 Image-EHR Cross Attention (IECA)

Existing diffusion-based models [11,20] extract normal semantic features solely from input images, potentially incorporating anatomical anomalies. To mitigate this, we propose an Image-EHR Cross Attention (IECA) module to derive an embedding that represents the X-ray image through EHR data, reducing the influence of anomalous features. As shown in Fig. 2-(a), we first use a feature tokenizer [8] to generate EHR embeddings, producing f tokens based on tabular features. The generated tokens $F \in \mathbb{R}^{f \times d}$ are then used to compute token-wise weights \mathbf{w}_r based on their similarity to the input image embedding \mathbf{e} . The final conditional embedding \mathbf{c}_r is then obtained as a weighted sum of F :

$$\begin{aligned} F &= \text{FeatureTokenizer}(\mathbf{r}), \quad \mathbf{r} \in \mathbb{R}^{1 \times (\text{col})}, \quad F \in \mathbb{R}^{f \times d}, \\ \mathbf{c}_r &= \mathbf{w}_r F, \quad \mathbf{w}_r = \sigma_{\text{softmax}}((F\mathbf{e}^T)/\sqrt{d}), \quad \mathbf{e} = \text{Encoder}(\mathbf{x}), \quad \mathbf{e} \in \mathbb{R}^{1 \times d} \end{aligned} \quad (2)$$

Since the generated conditional embedding \mathbf{c}_r is based on the similarity with \mathbf{e} , it implicitly encodes the EHRs information to align with the semantic features of the input image. \mathbf{c}_r is then used as a conditional input for MPG and NP to provide normal semantic features for the input image.

3.2 Pixel-level Checkerboard Masking (PCM)

Existing masking-based methods apply random masking with specific textures to simulate predefined target anomalies, training models to reconstruct the simulated patterns into anatomically acceptable shapes [13,31]. However, such approaches often struggle to detect real pathological abnormalities that deviate

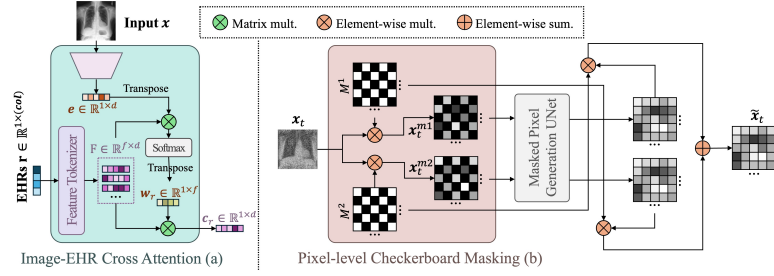


Fig. 2. Details of the proposed modules. (a) The Image-EHR Cross Attention module incorporate structured clinical context into the image generation process. (b) The Pixel-level Checkerboard Masking module facilitates the regeneration of anomalies into normal-like images using two types of static masks.

from predefined patterns. To avoid reliance on specific anomaly patterns and encourage the model to depend on auxiliary information—specifically, the EHR conditional embedding from IECA, we employ static masking that deliberately obscures image regions. This approach enables the detection of a broader range of anomalies beyond those seen in the training data.

As illustrated in Fig. 2(b), PCM consists of two complementary checkerboard masks (M^1, M^2) of input image size. A scale parameter s adjusts masking intensity by time step t to prevent later losses affecting earlier steps. At time step t , applying M^1 and M^2 to noised input x_t yields x_t^{m1} and x_t^{m2} . To obtain the final prediction, we apply the opposite mask to each reconstructed image. The formulation is as follows:

$$\begin{aligned} M_{i,j}^1 &= (i + j) \bmod 2, & M_{i,j}^2 &= 1 - (i + j) \bmod 2, \\ M^{1,2} &\leftarrow (M^{1,2} \times (1 - s) + s), & s &= t/T, \\ \tilde{x}_t &= \text{MPG}(x_t^{m1}) \odot M^2 + \text{MPG}(x_t^{m2}) \odot M^1, & x_t^{m1}, x_t^{m2} &= x_t \odot M^{1,2} \end{aligned} \quad (3)$$

The regenerated \tilde{x}_t is used for loss function during training and for generating the previous step image x_{t-1} during detection.

Training and Detection To train the noise prediction capability of the NP Network and the masked image reconstruction capability of the MPG Network, we compute the mean squared error (MSE) between output of each network ($\epsilon_\theta^{(t)}$, \tilde{x}_t) and its corresponding target ($\epsilon^{(t)}$, x_t) as follows:

$$\mathcal{L}_{Diff3M} = \mathbb{E}_{t, x, c_r, \epsilon} \left[\lambda \|\epsilon_\theta^{(t)} - \epsilon^{(t)}\|_2^2 + (1 - \lambda) \|\tilde{x}_t - x_t\|_2^2 \right] \quad (4)$$

where $\lambda \in (0, 1)$ is a weighting factor. For more efficient training, the base DDPM and IECA module can be pre-trained on X-ray images and EHR data (Fig. 1).

Algorithm 1 outlines the Diff3M detection process. Using the DDIM sampling strategy, the forward process introduces noise into the image up to T' ($T' < T$)

steps. During the reverse process, the model generates \mathbf{x}_{t-1} using produced noise $\epsilon_\theta^{(t)}$ and image $\tilde{\mathbf{x}}_t$ following [28]. The final anomaly score measures the difference between the reconstructed image \mathbf{x}_0 and the original input \mathbf{x} . Samples with high anomaly scores are flagged as potential anomalies.

Algorithm 1 Anomaly detection using Diff3M

```

1: Input: input image  $\mathbf{x}$ , EHRs  $\mathbf{r}$ , timestamp  $t$ , noise level  $T'$ 
2: Output: image-level anomaly score  $a$ 
3: for all  $t$  from 0 to  $T' - 1$  do
4:    $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t + \sqrt{\bar{\alpha}_{t+1}} \left[ \left( \sqrt{\frac{1}{\bar{\alpha}_t}} - \sqrt{\frac{1}{\bar{\alpha}_{t+1}}} \right) \mathbf{x}_t + \left( \sqrt{\frac{1}{\bar{\alpha}_{t+1}}} - 1 - \sqrt{\frac{1}{\bar{\alpha}_t}} - 1 \right) \epsilon_\theta^{(t)} \right]$ 
5: end for
6:  $\mathbf{c}_r \leftarrow IECA(\mathbf{r}, \mathbf{x})$ 
7: for all  $t$  from  $T'$  to 1 do
8:    $t_{emb} \leftarrow TimestampEmbedding(t)$ 
9:    $\tilde{\mathbf{x}}_t \leftarrow MPG(\mathbf{x}_t^{m1}, \mathbf{x}_t^{m2}, \mathbf{c}_r, t_{emb})$ ,  $(\mathbf{x}_t^{m1}, \mathbf{x}_t^{m2}) \leftarrow PCM(\mathbf{x}_t)$ 
10:   $\mathbf{x}_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}} \left( \frac{\tilde{\mathbf{x}}_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta^{(t)}}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon_\theta^{(t)}$ ,  $\epsilon_\theta^{(t)} \leftarrow NP(\mathbf{x}_t, \mathbf{c}_r, t_{emb})$ 
11: end for
12:  $a \leftarrow diff(\mathbf{x}, \mathbf{x}_0)$ ,  $diff(\mathbf{x}, \mathbf{x}_0) \in \{\|\mathbf{x} - \mathbf{x}_0\|_2^2, \max_{i,j} |\mathbf{x}^{i,j} - \mathbf{x}_0^{i,j}|\}$ 
13: return  $a$ 

```

4 Experiments and Results

Datasets and Evaluation Metrics CheXpert [14] and MIMIC-CXR [15] both contain X-ray images, the same types of demographic information, and 12 disease classes. We define normal samples as cases with *No Finding*, while all other cases are considered anomalies. For demographic features, we use sex, age, and AP/PA view. In MIMIC-CXR, we further incorporate BMI, blood pressure, height, and weight from MIMIC-IV [17] as additional EHR features, selecting records within three months of the X-ray imaging date. CheXpert has 16,969 normal samples for training, and (26/176) (normal/anomaly) samples for testing. MIMIC has 29,310 normal samples for training, and (487/1646) (normal/anomaly) samples for testing. We evaluate image-level anomaly detection using AUROC and AUPRC to compare performance across various detection thresholds.

Implementation Details Diff3M is trained using 8 Nvidia RTX A5000 GPUs. All images are resized to 256×256 . IECA, MPG, and NP adopt the UNet-based architecture from [28]. We use Adam optimizer with a maximum time step T of 1,000, a batch size of 3, a weighting factor λ of 0.5 and a learning rate of 10^{-4} . For Chexpert, the model is trained from scratch for 60,000 iterations. For MIMIC, the model is pre-trained on CheXpert, and fine-tuned with EHRs for 70,000 iterations. During inference, DDIM sampling is applied with 600/400 steps for CheXpert/MIMIC.

Table 1. Performance comparison on CheXpert and MIMIC datasets. *demo*: demographics information for IECA. *ehr*: additional EHRs from MIMIC-IV for IECA.

Dataset	Metric	Feature-based			Diffusion-based		
		PatchCore	RD4AD	MambaAD	DiAD	Ours _{demo}	Ours _{demo,ehr}
CheXpert	AUROC	0.584	0.529	0.594	0.664	0.664	-
	AUPRC	0.888	0.719	0.889	0.917	0.931	-
MIMIC	AUROC	0.573	0.555	0.570	0.597	0.610	0.617
	AUPRC	0.796	0.800	0.779	0.803	0.818	0.821

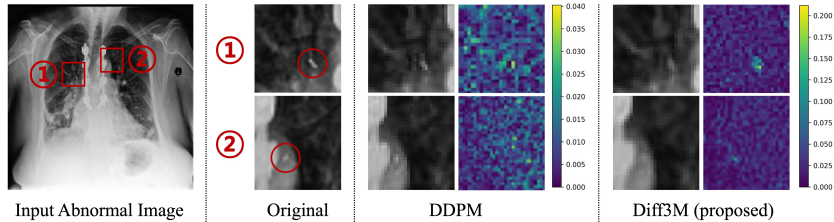


Fig. 3. Qualitative comparison. Red circles indicate anomalous artifacts.

4.1 Performance Comparison

We compare Diff3M with state-of-the-art (SOTA) feature-based models, *e.g.*, PatchCore [24], RD4AD [5], and MambaAD [10], along with the diffusion-based model DiAD [11]. For qualitative evaluation, we compare the anomaly maps, computed as the absolute pixel-level difference between the input and output.

Quantitative and Qualitative Results As shown in Table 1, Diff3M achieves SOTA performance in image-level anomaly detection for both CheXpert and MIMIC. For MIMIC, Diff3M improves AUROC by approximately 2.0–5.5% over previous approaches. This highlights the importance of effectively leveraging EHR data to enhance anomaly detection. For CheXpert, the performance of Diff3M using only demographics-based EHRs is comparable to DiAD. This suggests that demographic information alone may not fully capture anatomical variations. Overall, diffusion-based approaches outperform feature-based methods, likely because foundation models struggle to extract effective features in the medical domain, as discussed in Sec. 2.

Fig. 3 compares the anomaly regeneration capability of DDPM and Diff3M. We examine an abnormal image containing statistically anomalous artifacts. Unlike DDPM, Diff3M successfully transforms anomalous artifacts into normal-like structures for each region, assigning significantly higher anomaly scores compared to surrounding areas. These results demonstrate that Diff3M effectively generates anomaly maps, which can serve as a valuable medical assistive tool.

Table 2. Ablation study on the modules of Diff3M (MSE/ max_{abs}).

Dataset	Metric	DDPM	DDPM+PCM	Ours _{demo}	Ours _{demo,ehr}
CheXpert	AUROC	0.500/0.500	0.553/0.640	0.547/ 0.664	-
	AUPRC	0.838/0.867	0.902/0.922	0.899/ 0.931	-
MIMIC	AUROC	0.562/0.551	0.592/0.562	0.610/0.567	0.617 /0.568
	AUPRC	0.800/0.791	0.807/0.793	0.818/0.799	0.821 /0.798

Table 3. Analysis of attention weights for each input feature (mean \pm std). *entire*: analysis across entire samples. *top_{10%}*: analysis for the top 10% of feature values.

input feature f		BMI	BP _{max}	BP _{min}	Height	Weight
mean	<i>entire</i>	0.262 \pm 0.117	0.183 \pm 0.043	0.116 \pm 0.037	0.210 \pm 0.040	0.230 \pm 0.038
$\mathbf{w}_r^{(f)}$	<i>top_{10%}</i>	0.475 \pm 0.033	0.124 \pm 0.022	0.081 \pm 0.024	0.145 \pm 0.021	0.174 \pm 0.013

4.2 Ablation Study

Table 2 shows that adding PCM and IECA sequentially improves performance on both datasets. This demonstrates that each module contributes to anomaly detection and that their combination further enhances performance. We also compare the performance of MSE and pixel-level maximum absolute difference (max_{abs}) as anomaly scoring methods. The results show that MSE is more effective for MIMIC, whereas max_{abs} performs better for CheXpert. To understand this discrepancy, we analyze anomaly detection performance for individual disease classes. In MIMIC, switching to max_{abs} leads to a 12.3–19.9% performance drop for *Fracture* and *Pleural Other* cases. Conversely, in CheXpert, the number of positive samples for these classes is extremely low (0–1), which explains why using max_{abs} does not negatively impact performance. These findings suggest that different scoring functions may be more suitable for different diseases. Therefore, when using the anomaly detection models as a medical assistive tool, it may be necessary to review anomalies based on different scoring functions.

Effects of EHRs Table 3 presents the mean of attention weights (\mathbf{w}_r) assigned to each IECA input feature f in the MIMIC test dataset. Among the EHR features, BMI has the highest average weight. In particular, for the top 10% of samples, BMI accounts for 47.5%, indicating a significantly dominant contribution. This suggests that BMI is the most representative feature for capturing the anatomical characteristics of input images. Unlike height and weight, which independently provide limited information about body shape, BMI (calculated as weight/height²) offers a more meaningful approximation of body shape. These findings demonstrate that IECA effectively combines EHR features, extracting conditional embeddings that reflect anatomical information. In the case of demographic features, most of the weights converged to zero. These results suggest that utilizing richer EHR features is crucial for robust anomaly detection.

5 Conclusion

This paper proposes Diff3M, a novel framework that leverages EHR data for enhancing medical anomaly detection. By integrating EHR data through Image-EHR Cross Attention (IECA) and enforcing structured feature learning with Pixel-level Checkerboard Masking (PCM), Diff3M consistently outperforms existing diffusion-based and feature-based methods on CheXpert and MIMIC datasets. In the future work, we will apply lightweight models such as latent diffusion [23] to analyze the trade-off between detection performance and efficiency resulting from structural shifts in the model. In addition, we will explore incorporating richer clinical data to further enhance its diagnostic support capabilities.

Acknowledgments. This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the Global Research Support Program in the Digital Field program (RS-2024-00431394) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

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

References

1. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14. pp. 622–637. Springer (2019)
2. Bao, J., Sun, H., Deng, H., He, Y., Zhang, Z., Li, X.: Bmad: Benchmarks for medical anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4042–4053 (2024)
3. Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A.: Diffusion models with implicit guidance for medical anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 211–220. Springer (2024)
4. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: International conference on pattern recognition. pp. 475–489. Springer (2021)
5. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9737–9746 (2022)
6. Fan, Y., Liao, H., Huang, S., Luo, Y., Fu, H., Qi, H.: A survey of emerging applications of diffusion probabilistic models in mri. *Meta-Radiology* p. 100082 (2024)
7. Feng, X., Huang, Q., Li, X.: Ultrasound image de-speckling by a hybrid deep network with transferred filtering and structural prior. *Neurocomputing* **414**, 346–355 (2020)
8. Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A.: Revisiting deep learning models for tabular data. *Advances in Neural Information Processing Systems* **34**, 18932–18943 (2021)

9. Han, Z., Wang, Y., Zhou, L., Wang, P., Yan, B., Zhou, J., Wang, Y., Shen, D.: Contrastive diffusion model with auxiliary guidance for coarse-to-fine pet reconstruction. In: International conference on medical image computing and computer-assisted intervention. pp. 239–249. Springer (2023)
10. He, H., Bai, Y., Zhang, J., He, Q., Chen, H., Gan, Z., Wang, C., Li, X., Tian, G., Xie, L.: Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. arXiv preprint arXiv:2404.06564 (2024)
11. He, H., Zhang, J., Chen, H., Chen, X., Li, Z., Chen, X., Wang, Y., Wang, C., Xie, L.: A diffusion-based framework for multi-class anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 8472–8480 (2024)
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020)
13. Iqbal, H., Khalid, U., Chen, C., Hua, J.: Unsupervised anomaly detection in medical images using masked diffusion model. In: International Workshop on Machine Learning in Medical Imaging. pp. 372–381. Springer (2023)
14. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghighi, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 590–597 (2019)
15. Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L., Mark, R.: Mimic-iv (version 3.1). *physionet* (2024)
16. Johnson, A., Pollard, T., Mark, R., Berkowitz, S., Horng, S.: Mimic-cxr database (version 2.0.0). *physionet* **2**, 5 (2019)
17. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L.A., Mark, R.: Mimic-iv. *PhysioNet*. Available online at: <https://physionet.org/content/mimiciv/1.0/> (accessed August 23, 2021) pp. 49–55 (2020)
18. Lee, S., Lee, S., Song, B.C.: Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access* **10**, 78446–78454 (2022)
19. Mao, J., Wang, Y., Tang, Y., Xu, D., Wang, K., Yang, Y., Zhou, Z., Zhou, Y.: Medsegfactory: Text-guided generation of medical image-mask pairs. arXiv preprint arXiv:2504.06897 (2025)
20. Mousakhan, A., Brox, T., Tayyub, J.: Anomaly detection with conditioned denoising diffusion models. arXiv preprint arXiv:2305.15956 (2023)
21. Obuchowicz, R., Lasek, J., Wodziński, M., Piórkowski, A., Strzelecki, M., Nurzynska, K.: Artificial intelligence-empowered radiology—current status and critical review. *Diagnostics* **15**(3), 282 (2025)
22. Perlin, K.: Improving noise. In: Proceedings of the 29th annual conference on Computer graphics and interactive techniques. pp. 681–682 (2002)
23. 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. pp. 10684–10695 (2022)
24. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2022)
25. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis* **54**, 30–44 (2019)
26. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International conference on information processing in medical imaging. pp. 146–157. Springer (2017)

27. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
28. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. In: International Conference on Medical image computing and computer-assisted intervention. pp. 35–45. Springer (2022)
29. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 650–656 (2022)
30. Yao, W., Yin, K., Cheung, W.K., Liu, J., Qin, J.: Drfuse: Learning disentangled representation for clinical multi-modal fusion with missing modality and modal inconsistency. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 16416–16424 (2024)
31. Zhang, X., Li, N., Li, J., Dai, T., Jiang, Y., Xia, S.T.: Unsupervised surface anomaly detection with diffusion probabilistic model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6782–6791 (2023)
32. Zhou, Y., Huang, S.C., Fries, J.A., Youssef, A., Amrhein, T.J., Chang, M., Banerjee, I., Rubin, D., Xing, L., Shah, N., et al.: Radfusion: Benchmarking performance and fairness for multimodal pulmonary embolism detection from ct and ehr. arXiv preprint arXiv:2111.11665 (2021)